Comparative Quantification of Health Risks

Global and Regional Burden of Disease Attributable to Selected Major Risk Factors

Volume 2

Edited by

Majid Ezzati, Alan D. Lopez, Anthony Rodgers and Christopher J.L. Murray

World Health Organization
Geneva

WHO Library Cataloguing-in-Publication Data

Comparative quantification of health risks : global and regional burden of disease attributable to selected major risk factors / edited by Majid Ezzati . . . [et al.].

2 v. + v.3 in 1 CD-ROM.

Contents: vol. 1, Childhood and maternal undernutrition—Other nutrition-related risk factors and physical activity—Addictive substances—vol. 2, Sexual and reproductive health—Environmental and occupational risks—Other selected risks—Distribution of risks by poverty—Data analysis and results—Multi-risk assessment.—Annex tables CD-ROM, Population attributable fractions, mortality and disease burden for selected major risk factors.

1. Risk factors 2. Cost of illness 3. Risk assessment
4. Comparative study I. Ezzati, Majid. II. Title: Global and regional burden of disease attributable to selected major risk factors.

ISBN 92 4 158031 3 (NLM Classification: WA 105)

Typeset in Hong Kong
Printed in Switzerland

Contents

Volume 1

CD-ROM

Annex tables

Population attributable fractions, mortality and disease burden for selected major risk factors

Chapter 14

Unsafe sex

Emma Slaymaker, Neff Walker, Basia Zaba and Martine Collumbien

Summary

The risk factor "unsafe sex" has been defined here as sex between a susceptible person and a partner who has a sexually transmitted infection (STI), without taking measures to prevent infection. Unsafe sex cannot therefore be defined *a priori* (because sex is only unsafe with respect to the context in which it occurs), or measured directly from reported behaviours. A set of behaviours was defined as "risky sex" and the prevalence of various behaviours was estimated for 57 countries. The prevalence of risky sex as defined here is given by the proportion of the population who have had sex in the last year with a non-co-resident partner, and who did not use a condom on the last occasion with that partner. For the comparative risk assessment (CRA) estimates, the prevalence of risky sex between men and women was the primary focus.

The main outcome considered was infection with HIV, which is responsible for the majority of the burden of mortality and morbidity associated with STIs. Infections with *Chlamydia trachomatis* (chlamydia), *Neisseria gonorrhoeae* (gonorrhoea), human papillomavirus (HPV) and *Treponema pallidum* (the causative agent of syphilis; hereafter referred to as "syphilis") were considered in less detail because the information available for these infections is inadequate for detailed analysis.

Infection with HIV/AIDS is the fourth leading cause of mortality in the world. Currently, most (29.4 million) of the 42 million people globally who are infected with HIV are concentrated in Africa, but epidemics elsewhere in the world are growing rapidly. Prevalence is increasing most swiftly in eastern Europe and central Asia (UNAIDS/WHO 2002). Most of the infections prevalent in 2001 were acquired through heterosexual sexual intercourse. Most people infected with HIV do not know they are infected, making prevention and control difficult. The other STIs included in the burden estimates, *C. trachomatis*, *N. gonorrhoeae*, HPV

and *T. pallidum* (syphilis), cause morbidity in all regions of the world. Infection with some of these agents can lead to infertility (e.g. *C. trachomatis*) or cancer (HPV), and an acute STI may enhance the transmission of and susceptibility to HIV.

To estimate the prevalence of sexual risk behaviours, suitable studies were located and, where possible, the data produced by these studies were analysed to create a set of standard indicators for different aspects of sexual behaviour. The prevalence of different sexual behaviours and characteristics varies greatly between countries and between subregions.[1] The levels of risk behaviour did not vary in a predictable manner, and variations in reported behaviour at the aggregate level do not correspond to differences in HIV prevalences. A literature search was also carried out to identify reported risk factors for HIV infection and estimates of the risk associated with each factor. Since the outcomes are infections transmitted from person to person, the relative risk of infection changes with the prevalence of the infection, and changes in prevalence affect incidence.

Two different approaches were used to estimate the avoidable burden of HIV/AIDS attributable to unsafe sex. For countries in sub-Saharan Africa where the prevalence of HIV/AIDS in adults is high and the epidemic is largely driven by heterosexual sex, a mathematical projection model (the Epidemic Projection Package [EPP]) was used to estimate how many infections were attributable to unsafe sex, and how many were potentially avoidable. For countries where adult prevalence is lower and the spread of HIV/AIDS is confined to specific subgroups, a different approach was used whereby current estimates of HIV/AIDS and projections were based on estimates of sub-epidemics related to the mode of transmission (e.g. injecting drug use, men who have sex with men, heterosexual transmission). For these countries the risk associated with unsafe sex was the percentage of all infections that were sexually acquired. The other STIs were assumed to be entirely the result of unsafe sex and therefore 100% of the burden caused by these STIs is avoidable.

The modelling exercise suggests that there would not have been an HIV epidemic in Africa had there never been any sexual transmission since of the cases of HIV infection prevalent in 2001, >99% were associated with a sexually acquired infection at some point in the chain of transmission. In the rest of the world, the estimated percentage of the HIV infections prevalent in 2001 that were attributable to unsafe sex ranged from 25% in eastern Europe (EUR-C) to 95% in parts of Latin America (AMR-D). Using these estimates, the mortality attributable to unsafe sex ranged from 4000 deaths in EUR-C to 1 632 000 in AFR-E. Globally, 2 444 000 deaths and 75 783 000 disability-adjusted life years (DALYs) were attributable to this risk factor. If unsafe sex were to cease, most parts of the world would see a substantial drop in the number of new HIV infections.

1. Introduction

A variety of infectious agents can be transmitted through sexual contact, including HIV, chlamydia, gonorrhoea, HPV and syphilis). While having sex (which in this chapter refers to vaginal sexual intercourse, unless otherwise stated) may place a person at risk of being infected by one or more of these agents, it is difficult to assess the magnitude of this risk. Sex can only be defined as "safe" or "unsafe" if something is known about the context in which it takes place and with whom. Having sex does not place a person at risk of contracting a disease unless that person's partner has an infection, which they can transmit. Therefore, unlike many other risk factors, which are independent of the situation in the broader population, or with respect to other individuals, unsafe sex cannot be uniquely defined by the set of actions of an at-risk individual. Rather, a definition must be based on an analysis of the individual's actions in light of the background prevalence of disease. The principal health outcome considered in this chapter is the number of adults aged 15–49 years who become infected with HIV as a consequence of unsafe sex, and the number of these infections that is potentially avoidable. Infections in children resulting from vertical transmission were not included since these are not caused directly by unsafe sex, but by infection via the mother, together with a lack of pre- and postnatal treatment of mother and child. The other STIs (chlamydia, gonorrhoea, HPV and syphilis) were considered separately and in less detail because they contribute to a lesser degree to the burden of disease and mortality, and because of the limited amount of information available regarding the prevalence and current transmission dynamics of these infections in different subregions. A thorough review of the epidemiology and importance of these STIs was given in previous burden of disease work (Rowley and Berkley 1998).

The relationship between the risk factor unsafe sex and the disease outcomes, which contribute to the global burden of disease, cannot be described using the standard epidemiological tradition of constant, extrapolable hazards. This is owing to the fact that the outcomes considered in this chapter all relate to infections which are transmitted from person to person. The relative risk of being infected by any one of these diseases is therefore dependent on the prevalence of the disease.

1.1 Definitions of unsafe sex

In this chapter, STIs were the only negative outcomes of sexual contact considered. Other potentially deleterious outcomes, such as an unwanted pregnancy or the psychological consequences of sexual violence, are considered elsewhere in the CRA (see chapters 15 and 23). It is important to define the group of people who share a common risk factor for contracting an STI in order to be able to carry out a risk assessment. The risk factor has been called "unsafe sex", but this term does not

immediately suggest a clearly defined characteristic of either an individual or a population that can be used to determine how many people are affected by the risk factor. "Safe" sex has previously been defined as follows:

> Consensual sexual contact with a partner who is not infected with any sexually transmitted pathogens and involving the use of appropriate contraceptives to prevent pregnancy unless the couple is intentionally attempting to have a child. (Berkley 1998)

This definition is not useful for the purposes of this chapter, since many of the ways in which the above definition can be negated would not put an individual at risk of acquiring an STI. For example, sex with an uninfected partner without using contraception does not pose a risk of infection and nor would sex with an infected partner if a condom was used properly.

Therefore, before defining unsafe sex we must first consider what type of classification would be suitable to describe the degree of risk experienced by an individual or population. The risk of contracting an STI depends both on the individual and on the population. Individual behaviours determine whether or not it is possible for infection to occur. The prevalence of infection in the population determines whether or not the individual becomes exposed to an infectious agent. Therefore an ideal measurement of this risk would include both individual and population characteristics.

If it were possible to measure this risk at the individual level, a gradation across the population would be observed. Risk gradation suggests the possibility that a continuous index of risk could be constructed by combining several factors. However, many if not most behavioural factors do not retain a simple dose–response relationship when considered in combination with others. For example, consider a person who is not infected with HIV at a particular instant in time. The frequency of this person's sexual relations with a regular partner could show a dose–response relationship relative to the risk of acquiring infection, but only if the partner were infected. Similarly, the rate at which this person acquired new partners could also show a dose–response relationship, but only if each partner were infected. Past partner history would not be relevant, unless the individual had contracted another STI which could enhance the dose–response relationships between risk of infection and both coital frequency and partner acquisition rate. The conditionality of these interactions makes it practically impossible to quantify and construct a continuous measure of risk.

Individuals must therefore be categorized into static groups based on average levels of risk. This can be done so as to allow for the important effect of STI prevalence if the definitions of risk are based on probability of contact with cases, rather than on reportable behaviours. These

definitions will be valid at an instant in time, or for a very short time period; it is important to realize that such distinctions may be very short-lived since sexual networks are dynamic. The following definitions provide a way of thinking about the true distinctions.

Unsafe sex

Unsafe sex occurs if a susceptible person has sex with at least one partner who has an STI, without taking measures to prevent infection. Susceptible people are not yet infected, either because the infectious agent has not been successfully transmitted, or because the agent has been transmitted but infection has not yet been established. Such susceptible people form the group which is truly exposed to infection and they are at a very high risk of becoming infected. For intervention and prevention purposes, this group is not as important as that defined below, because it is too late to prevent the members of this group from being exposed to infection. However, this group is the most relevant in terms of predicting the number of new cases of STIs.

Hazardous sex

The group of people engaging in hazardous sex are those susceptible persons who either engage in unprotected sex but who have not yet encountered a partner who has an STI, or who have had sex with at least one partner who has an STI, but have taken measures to prevent transmission. These people have the potential to be exposed to infection, either by encountering an infected partner, or if the measures taken to prevent transmission are ineffective (e.g. condom failure). This group of people is important for prevention efforts; a change in the size of this group has the most potential to change the number of new infections occurring in the future.

These two definitions ("unsafe" and "hazardous" sex) would provide a way to allocate people to risk groups if membership of the two groups could be measured. However, there is currently no way by which this can be measured and so it is necessary to use a definition based on reportable behaviours, i.e. "risky sex", as a proxy.

Risky sex

The people who have risky sex share a certain set of behaviours; these can be different in different epidemic situations, but are likely to include having many sexual partners and not using condoms. This classification is based on individual reports and will include infected as well as susceptible people, because infection status is not known from such reports. Ideally, we seek to identify reportable behaviours so that the group identified as having risky sex would include those having unsafe sex and those having hazardous sex but exclude others (those having "safe" sex or no sex).

Figure 14.1(a) shows the relationship between groups of infected and susceptible people in terms of those who have unprotected sex, those who have sex with an infected partner, and those who report a "risk behaviour". The reported risk behaviour is the measurable component of an individual's sexual behaviour.

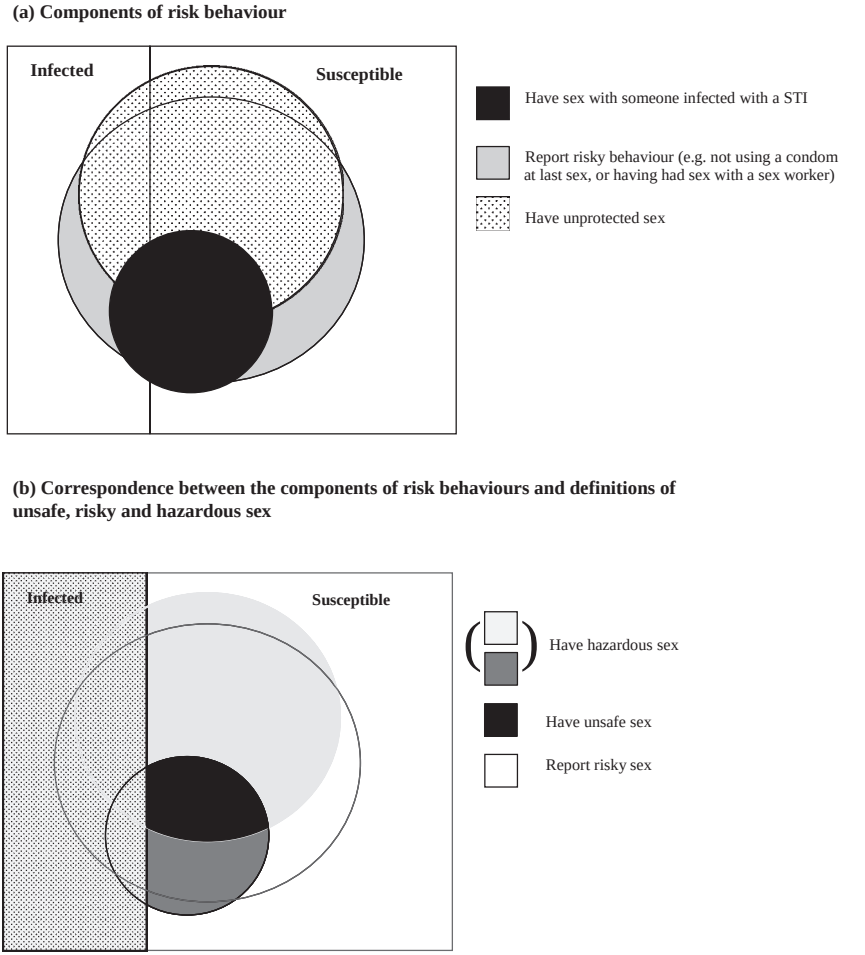
From this figure it is possible to see why classifying people into risk groups on the basis of reported behaviours is not necessarily a good measure of exposure to infection. Some people will be wrongly classified as at risk because they report risk behaviours, but actually they are already infected. Others will be wrongly classified because they report behaviours which have not exposed them to infection, as they did not behave in exactly the way they reported. Some people will be wrongly classified as not at risk because although they have had a sexual contact which could potentially have led to infection, they did not report this behaviour. This could be deliberate, because they do not wish to admit to "undesirable" behaviour, or unintentional, because the behaviour has been forgotten. People will also be misclassified if the risk behaviour they are asked to report is not the best predictor of the actual risk experienced. For example, in a population of married women this could happen if risk were classified on the basis of reporting sex with non-marital partners, but the main source of infection was in fact the women's husbands.

Figure 14.1(b) shows where the groups defined above (unsafe sex and hazardous sex) fall in this schematic. The black section shows the group having unsafe sex: the susceptible people who have unprotected sex with a partner who has an STI. By definition, this group falls wholly within the group of susceptibles and includes some of the people who report risk behaviours and some of the people who are classified as having hazardous sex. The group which has hazardous sex is shown in the light grey and dark grey sections and is composed of those who either have unprotected sex or who have protected sex with a partner who has an STI. This includes people who do report risk behaviours and some who do not. Again, by definition, this group includes only susceptible persons. Some members of the group who report risky behaviour, shown in white, are not included in either the group having hazardous sex or the group having unsafe sex.

1.2 Estimating levels of risky sex in a population

If there were no STIs, then there would be no unsafe and/or risky sex. In areas where there is a high prevalence of STIs, a larger number of sexual behaviour patterns will be dangerous than in places where very few people are infected with a sexually transmitted pathogen. A pragmatic definition of a specific behaviour, or group of behaviours, (e.g. sex without a condom) as "risky" can be useful, providing it is understood that the degree of risk associated with this behaviour will not be the same in different populations, or at different times in the same population.

Figure 14.1 Venn diagram illustrating the relationship between three ways of defining unsafe sexual behaviour



With this caveat in mind, the question arises as to how best to describe populations with different levels of risk.

Aggregate measures of sexual behaviour will inevitably be less informative than more local measures. However, even country-level indicators cannot capture the more subtle variations in sexual mixing patterns, such as partnership concurrency. The level of risk attached to a particular behaviour changes with the prevalence of the infection; if prevalence is high, there are more infected people in the population and so a susceptible person has a greater chance of choosing an infected person as their next sexual partner. In each subregion, the prevalence of STIs and

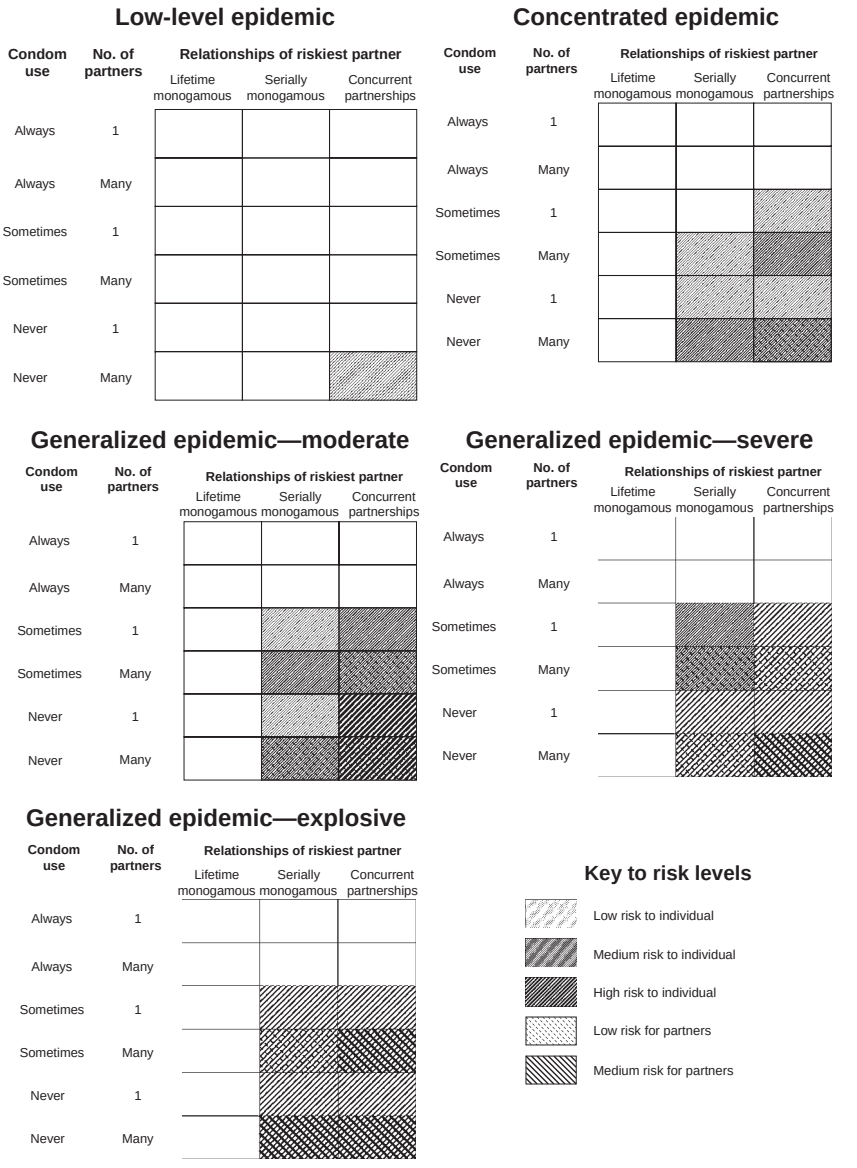
of certain behaviours varies between the countries. Within the different countries, STI prevalence and sexual behaviour can vary between urban and rural areas, age groups and sexes, socioeconomic classes, religious groups, between people of different sexual orientation and according to other factors such as proximity to transport links and health services. Personality and physiology play a significant role in determining a person's sexual behaviour and, in the case of the latter, susceptibility to infection. The impact of these determinants cannot be measured at the population level, but they are of great importance in determining how many people are exposed to STIs and how many people become infected.

The effect of heterogeneity in sexual behaviour on the ability to measure the level of dangerous exposure is more subtle. When sexual behaviour is measured in a survey, data are only collected regarding the respondents' behaviour. However, the behaviour of the sexual partner is as important a predictor of risk as the behaviour of the respondent. A respondent who has a large number of sexual partners is probably at a high risk of contracting an STI. However, if all of these partners have never had sex with anybody else, the respondent is perfectly safe. Therefore in a population where people vary greatly in the number and frequency of their sexual contacts, a one-sided measure of "sexual behaviour" is difficult to interpret. It is known that in most populations men and women have very different patterns of sexual behaviour. Most populations also have a subset of both men and women who are distinguished by high levels of sexual activity. Both of these imbalances make it difficult to quantify risk based on reported behavioural data from surveys.

The level of risk, to oneself and one's partner, is illustrated assuming different patterns of partnership and condom use in three different epidemic situations in Figure 14.2. The epidemic states correspond to those defined in WHO/UNAIDS (2000). A low-level epidemic is one in which, although HIV infection may have been present in the population for some time, it has not spread outside defined groups at a high risk of infection, and the prevalence among these groups has not exceeded 5%. A concentrated epidemic is one in which HIV infection has spread within defined groups and prevalence has exceeded 5% in at least one of these groups, but prevalence among pregnant women in urban areas remains below 1%. A generalized epidemic is one in which HIV infection has spread throughout the general population, as indicated by a prevalence of infection of greater than 1% among pregnant women.

A partnership is mutually monogamous if both partners only have sex with each other for the duration of the relationship. Lifetime mutual monogamy is always safe, regardless of the prevalence of STIs in the population. One-sided lifetime monogamy is safe for one of the partners in this type of relationship: individuals who have sex with a partner who has never had sex with anyone else do not place themselves at risk of infection from this partner, but may themselves present a risk to this partner if they have also had sex with other people. Serial monogamy is

Figure 14.2 Risk matrices: the level of risk to an individual and their partner is illustrated assuming different behavioural patterns in different epidemic situations



defined as a succession of monogamous relationships. These relationships are monogamous from the individual's standpoint, but no assumptions can be made about the behaviour of the partner. Partnerships of this sort may last for days or years. The frequency with which partnerships are dissolved and reformed will affect the risk of acquiring an STI and this will also be affected by the prevalence of STIs in the population. If an individual has sexual partnerships that overlap, such partnerships are said to be concurrent. STIs can be spread more easily if people have sex with several partners within a short space of time. Therefore although having concurrent partnerships is associated with the greatest risk of contracting an STI, serial monogamy with very short intervals between successive partners also places the partners at high risk.

2. Data sources

The data used to calculate levels of risky sexual behaviour came from general population surveys designed to be nationally representative. More than 300 surveys were identified that could potentially have been used in this analysis. Many of these had been carried out under the auspices of the Demographic and Health Surveys (DHS) programme conducted by Macro International. The focus of these surveys was family formation and fertility, and only more recently have questions on sexual behaviour been incorporated. Most DHS data are from African countries, but some surveys have been carried out in South America and Asia. South American countries are also covered by the Centers for Disease Control and Prevention (CDC) Reproductive Health Surveys (RHS) which asked questions about sexual behaviour. CDC has also carried out some surveys in Asian and eastern European countries. Other organizations, such as Population Services International (PSI), also carry out surveys which provide suitable information.

Most of the established market economy countries have carried out their own surveys of sexual behaviour, many of which date from the late 1980s and early 1990s, a time when policy-makers began to be concerned about the potential for the spread of HIV in these populations. For example, the data from the United Kingdom of Great Britain and Northern Ireland used in this analysis date from 1990; the survey was repeated in 2000 but the data were not yet available. The problem of standardization is greater for established market economy countries' surveys because they have been carried out by many different organizations, each of which sought different information to address different concerns.

2.1 Search strategy

The scientific literature was searched for information on the prevalence of different sexual behaviours and the relationship between risky sex and STIs. Information dating from after 1990 was used wherever possible.

Identifying survey data sets and or reports which incorporated information on sexual behaviour was not straightforward since these terms are not indexed in the major bibliographic databases. Therefore the use of a formal search strategy alone would not have been adequate. Suitable surveys were located in several ways:

Step 1: Web sites of surveying organizations

Organizations that carry out surveys that include information on sexual behaviours provide lists of these on their web sites; this was the first source of information for the majority of surveys. These organizations are:

- Demographic and Health Surveys (DHS), Macro International and Measure, USA (http://www.measuredhs.com) and (http://www.measureprogram.org)
- Reproductive Health Surveys (RHS) carried out by CDC, Atlanta, USA (http://www.cdc.gov/nccdphp/drh/gp_surveys.htm)
- Population Services International (PSI), USA (http://www.psi.org/)
- Family Health International (FHI), USA (http://www.fhi.org/)
- Global programme on AIDS (GPA) listed on http://www.unaids.org/publications/documents/epidemiology/determinants/Survey_Sexual_Behaviour.doc

Step 2: Search of published materials

Medline, Popline and Web of Science databases were searched for appropriate publications. Other databases providing qualitative information (such as Psychinfo) were not used because quantitative information was considered more important for this work.

Step 3: Contact with other researchers

This turned out to be the most efficient strategy because researchers involved with one survey frequently knew of other existing surveys.

Step 4: Internet search using Google

The Google Internet search engine was used, the principle search terms employed being the names of authors of surveys known to have been carried out and the names of institutes likely to have been involved in suitable surveys. It is not useful to carry out Internet searches using keywords related to sex.

Search terms

- Popline

Search terms used were “sex behaviour”, “condoms, male” “condoms, female” “population” “HIV infections”. This yielded 709 references, of which 75 were selected.

- Medline

Search terms employed were sexual behaviour, risk, ratio, odds, changes, sexual behaviour, incidence or prevalence, change*, reduction or lower or decline, HIV.

- Web of Science

Search terms used were sexu* and country name. If a search term returned a large number of hits, it was narrowed by adding "risk".

All the databases were searched for information from countries where there was no DHS, CDC, PSI or national (state) survey available.

2.2 Prevalence studies: HIV and other sexually transmitted infections

Data on the prevalence of HIV are generally from national surveillance systems. In most countries, women who attend antenatal care clinics (ANC) are tested for HIV anonymously and these data are taken to be representative of the general population in these countries. Other sources of surveillance data include blood donors, STI clinic patients and military recruits. Only ANC clinic prevalence data were used in this work. These data are collected by the United States Bureau of the Census and at the Joint United Nations Programme on HIV/AIDS (UNAIDS) in Geneva, from which the information is disseminated. The quality, coverage, history and competence of national surveillance schemes vary enormously (Walker et al. 2001). Consequently, prevalence data from some areas are more reliable than from others and more recent estimates are generally more reliable than older ones.

WHO collects the available STI prevalence data on a regular basis. However there is a lack of time-series data, which means that mathematical projection models cannot be used to make projections of future prevalence.

2.3 Prevalence studies: sexual behaviour

As described above, it is not clear which types of behaviours best define the group of people who are at risk of contracting an STI. Therefore information was collected on all behaviours which might be important in defining this group.

Measuring sexual behaviour

Target population

Many surveys of sexual behaviour have focused on high-risk groups within a population. In countries with concentrated epidemics, most STIs occur within these groups, which are often composed of people such as commercial sex workers or men who have sex with men. Unfortunately, the size of these groups relative to the total population is rarely known.

Information from surveys of groups at a high risk of infection cannot be extrapolated to the general population without an accurate estimate of the overall size of the group. General population surveys are unlikely to find a representative sample of members of groups at a high risk of infection and therefore underestimate the prevalence of risk-associated behaviour in a population. Data from groups at a high risk of infection have not been used directly in this work because sufficient information is rarely available to be able to use these data in the context of aggregate national estimates of risk behaviour. Therefore the estimates of the level of exposure could be too low in countries where STIs occur mainly within groups at a high risk of infection.

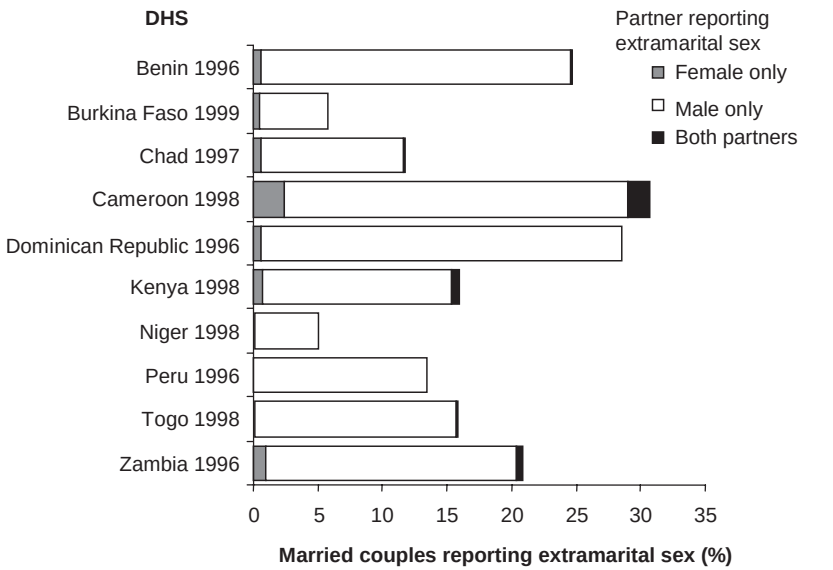
Methods of data collection

Sexual behaviour surveys are a fairly recent activity and the best methods for obtaining the required information have not been established. The most appropriate reference period for information on the number and characteristics of sexual partners is not known. There is no reliable method for comparing data collected for different periods of time. For example, somebody who reports having had one partner in the last month has not necessarily had twelve in the last year, but may well have had more than one partner in this time. Asking people for information from a longer time period will introduce a recall bias. This bias could be a problem because people might be more likely to recall partners of longer duration than those with whom contact is more short-term. This could lead to underestimates of the number of more risky sexual contacts.

If sexual behaviour patterns are changing over time, a cross-sectional survey will not give a good estimate of the cumulative lifetime exposure to risk, because the risk exposure of the youngest age groups at the time of the survey will not reflect the risk that the older age groups experienced when they were young. Ideally, current state measurements should therefore be supported by life course measures, even if we have to rely on recall data for the latter.

Surveys which only collect information on the respondent's own behaviour will misclassify some individuals with respect to their risk of acquiring an STI. They will systematically underestimate risk because people who are at a low risk because of their own behaviour could be put at risk by the behaviour of their partner. If people who are at a low risk always chose low-risk partners, then surveys could accurately estimate the proportion of those who are at risk. There is evidence from selected DHS with a couple subsamples that this is not the case, and that there is substantial misclassification of women as at a low risk based on their own behaviour, but who are in fact at risk through their husbands. This is illustrated in Figure 14.3 and supported by the results of other studies (e.g. Rwanda and Kenya, Chao et al. 1994; Hunter et al. 1994).

Figure 14.3 The proportion of married couples in which at least one partner reports having had extramarital sex during the last year, by country



Source: Data from selected DHS.

The accuracy of survey instruments in correctly evaluating people's behaviour is unknown. In studies where cross-sectional household surveys have been validated with in-depth interviews, it has been found that people tend to under-report "undesirable" behaviours (Konings et al. 1995). The age, sex and personal characteristics of the interviewer may also influence the reporting of sensitive information (Malamba et al. 1994), which is likely to include the behaviours of interest. Many surveys find that the number of partners reported by men greatly exceeds the number reported by women. Two factors contribute to this: general population surveys may fail to include the few women who have a large number of partners, and women may consistently under-report how many partners they have had (Glynn et al. 2001).

In choosing which data to use for the risk assessment, the first criterion was that the survey sample should be representative of the general population. Some surveys, mainly those with a demographic focus, only interviewed women, and some were concerned only with ever-married women (women who are currently married or who have been married at some stage in their lives). The latter samples were generally carried out in countries where it is not possible to discuss sexual behaviour

openly and consequently they provide only a limited amount of information. The age range of persons included in the surveys also varied. If a survey was limited to a narrow population in terms of sex, age or marital status, it was only analysed in the absence of a suitable alternative. This was the case for several countries. The surveys used, the populations covered and the data sources for each are listed in Appendix A. The type of information available and the number of countries and subregions covered are listed in Table 14.1.

The surveys referred to in Table 14.1 were carried out between 1989 and 2001. In general, the most recent survey available was used for each country. Individual-level data were required to calculate values for most of the sexual behaviour indicators because these were not usually given in a suitable format in published reports. In some cases, the data used to calculate the estimates presented in this chapter did not come from the most recent survey because such data were not available at the time of writing.

There were very few countries for which more than one survey was suitable for calculating sexual behaviour indicators. If more than one eligible survey for a country existed, the survey providing the most information was used first. Data from different surveys were not combined when calculating any one indicator for a particular country, but the full set of indicators for a country were not always derived from the same survey. The estimates for each indicator were rated according to how directly each indicator could be calculated from the information elicited by the survey questions and the number of assumptions which had to be made in the calculation. In cases where two estimates were available for one indicator, the estimate that was considered better was used.

The responses received in a survey may have been influenced by the manner in which the questions were phrased. The data presented here were derived from responses to several differently-phrased questions and this may have distorted the results. Within subregions this should not be of concern, beyond increasing the uncertainty of the measurement, as it seems unlikely that this error should vary with respect to exposure to STIs. However, a bias may well be introduced when making comparisons between subregions with different styles of questions because questionnaire styles are generally more similar within subregions.

Most general population data only cover heterosexual behaviour. Those surveys which discuss sex between men are generally carried out only among men who have sex with men and the number of these individuals in a population is rarely known. Therefore the behavioural measures collected for this analysis focussed entirely on heterosexual sex. For some subregions where sex between men plays a key role in the epidemic, this is an important omission. However, data are rarely available on behaviour in homosexual men in the subregions with the greatest burden of STIs, and the focus of this chapter has been largely dictated by the epidemic in these countries. As will be explained in more detail below,

Table 14.1 Indicators of sexual behaviour and the number of countries and subregions for which relevant data are available, by sex

			Information available (n)			
			Countries		*Subregions*[c]	
Indicator of sexual behaviour	*Denominator*	*Numerator*[a]	*Female*	*Male*	*Female*	*Male*
Ever had sex	Everyone	Number who say they have ever had sex	63	42	13	9
Sexually active in the last year	Everyone who has ever had sex	Number who had sex in the last year	59	40	12	6
Higher-risk sex in the last year	All who have had sex in the last year	Sex with non-co-resident partner in the last year	47	34	10	8
Condom use last time had higher-risk sex	All who have had higher-risk sex in the last year	People who used a condom last time had higher-risk sex	34	30	7	8
Men who had sex with a CSW in the last year	All men	Men who had sex with a CSW in the last year	NA	41	NA	10
Condom use last time had commercial sex	Men who report having had commercial sex in the last year	Men who used a condom last time they had commercial sex	NA	23	NA	6
Young people[b] having premarital sex in last year	All young people who have never had a co-resident partner (i.e. currently single)	Never had a co-resident partner and had sex in the last year	53	37	9	7
Condom use last time had premarital sex	All young, single and sexually active people	Young, single, sexually active and used a condom last time had sex	33	29	7	7

Young people having multiple partnerships in the last year	All young people	Young people who report more than one partner in the last year	31	31	7	7
Young people's condom use last time had higher-risk sex	All young people who had sex within the last year	Young people who used a condom the last higher-risk sex in the last year	45	28	12	7
Condom use first time had sex	All young people who have ever had sex	Young people who used a condom the first time they had sex	9	10	3	4
Had sex by age 15 years	Everyone	First had sex before the age of 15 years	57	39	11	8
Median age at first sex	Everyone	Lifetable median	65	51	12	11
Condom use last time had marital sex	Married people (including co-resident partnerships that are not legal marriages)	Married people who used a condom the last time they had sex with their spouse	20	22	6	6
Extramarital sex in the last year	All married people	Married people who had sex with a non-co-resident partner during the last year	25	25	7	7
≥2 non-marital partners in last year	All people who have had sex in the last year	Number who report ≥2 partners, with whom they do not live, during the last year	18	18	5	6
Number of partners	Everyone	Mean and median number	27	30	7	7

CSW Commercial sex worker.

NA Not applicable.

[a] Only people who can contribute to the denominator are included in the numerator.

[b] Young people are defined as people aged 15–24 years inclusive.

[c] Subregions for which at least one country-level estimate was available.

whilst homosexual men are not included in the exposure estimates, they are included in the estimates of attributable and avoidable infections as a result of the modelling approach taken.

Flow of data

Having identified a suitable survey, the questionnaire (if available) was assessed to ensure that the data would be suitable for inclusion in this analysis. If suitable, the data were obtained and converted (if necessary) for analysis using Stata version 7.0. Variables were created for as many of the standard indicators (those listed in Table 14.1) as was possible for each survey. These were then used to calculate the weighted numbers of people in each category, and the results were exported to a Microsoft Access database.

Survey design issues

There is no standard survey questionnaire. Even those carried out by the same organization, such as DHS, differ slightly from country to country and from year to year. DHS use a standard questionnaire for each survey round, but countries do not necessarily use all of, or only, the standard questions in their surveys. The standard questionnaire for the round four DHS has departed from the previous standard in the AIDS module and now asks about the previous three partners, in contrast to the prior rounds which asked about marital and non-marital partners. Other surveys have differently-worded questions and a different structure and order of questions. Therefore the data had to be standardized in some way.

Two major problems emerged while trying to compile the responses to different questionnaires to allow comparison. First, the reference period for questions on sexual behaviour varied. The majority of surveys asked about behaviour in the year prior to the survey but a few used different timescales. It is difficult to relate the responses to questions with one reference period to those with another reference period and therefore some of the data could not be used to calculate the standard indicators. Second, questions relating to condom use followed two styles. One style asked about condom use on the last occasion (with a particular partner). The other asks whether condoms were always, sometimes, or never used (with a particular partner). The latter question is impossible to compare between different surveys since it would be necessary either to quantify "sometimes" or to get an estimate of consistency of condom use with different partners. A significant amount of data on condom use could not be included here for this reason. Work has been done on methods for comparing responses to different types of questionnaire design; however, to do this effectively for this analysis would have required many assumptions to be made, and would thus have introduced another possible source of error.

Standardization of questionnaires

Given the differences in question wording and questionnaire structure, it was not possible to define a set of rules for this process. Table 14.2 shows some of the questions used in constructing the same indicator for different countries.

2.4 Outcome studies: sexual behaviour and HIV/AIDS

Estimating the relative risk of HIV infection in exposed vs non-exposed people

The relative risk or odds ratio for various indicators of sexual behaviour has been assessed in a number of general population studies listed in Table 14.3. The accuracy of these estimates is influenced by the following factors.

Methodological issues

The time at which a person became infected is an important piece of information because it is their behaviour at around that time which is the most relevant when estimating relative risk. People do not usually know that they are infected, let alone when this occurred, so behaviour is seldom measured for the relevant period of time. This could reduce the chances of detecting a real association. Studies which attempt to find risk factors for STIs, in particular HIV, face problems because of cultural unease about discussing STIs. Other problems include a lack of laboratory resources and expertise in geographical areas with high prevalence, as well as the ethical issues involved in serological testing.

The studies which estimate the risk associated with particular behaviours are mostly cross-sectional. If sexual behaviour patterns are changing over time then these surveys, which look for patterns of association between estimates of exposure and prevalence, could produce misleading results. The behaviour reported by HIV-positive people who have been infected for some time, and whose behaviour has changed between the time of infection and the time of the survey, will not reflect their behaviour at the time of infection. The degree to which people are misclassified in this way will depend on the stage of the epidemic (because in the early stages more infections are recently acquired) and on the reference period used in the survey.

This effect could be mitigated if life course measures were also considered. Comparison between life course and the more recent measures could show if behaviours have changed. Some indicators of behaviour are known to correlate with others. For example, age at which the individual first has sex has been shown to correlate with number of extramarital partners later in life (White et al. 2000) and so inconsistencies in this relationship, where this has been previously documented, could point to changing patterns of behaviour.

Table 14.2 Examples of questions whose responses were used to construct various indicators of sexual behaviour

Name of survey	*Question asked*	*Mode*
Number of people who have ever had sex		
NEM European Group	Have you ever had sexual intercourse?	FTF
NATSAL 1990 (United Kingdom)	How old were you when you first had sexual intercourse with someone of the opposite sex, or hasn't this happened?"	FTF
DHS Zambia 1996	Married: When was the last time you had sexual intercourse with (your husband/the man you are living with)? Not married: When was the last time you had sexual intercourse (if ever).	FTF
DHS Kazakhstan 1999	How old were you when you first had sexual intercourse (if ever)?	FTF
PSI Rwanda 2000	Avez-vous jamais fait l'amour avec une personne de sexe opposé?	FTF
Number of people who had sex in the year before the survey		
NEM European Group	With how many persons of the opposite sex have you had sex over the last 12 months, even only once?	FTF
NATSAL 1990 (United Kingdom)	When, if ever, was the last occasion you had vaginal sexual intercourse with a (man/woman)?	SAQ
DHS Zambia 1996	Married: When was the last time you had sexual intercourse with (your husband/the man you are living with)? Not married: When was the last time you had sexual intercourse (if ever)	FTF
DHS Kazakhstan 1999	When was the last time you had sexual intercourse?	FTF
PSI Rwanda 2000	Quand avez-vous fait l'amour la dernière fois?	FTF
Number of men who had sex with a commercial sex worker in the year before the survey		
NEM European Group	Have you ever had sex with a person you paid to have sex? If yes: When was it for the last time?	FTF
NATSAL 1990 (United Kingdom)	Have you ever paid money for sex with a woman? If yes: When was the last time you paid money for sex with a woman?	SAQ
DHS Zambia 1996	Have you given or received money, gifts or favours in return for sex at any time in the last 12 months?	FTF
DHS Kazakhstan 1999	Have you ever paid for sex? If yes: How long ago was the last time you paid for sex?	FTF
PSI Rwanda 2000	Au cours des douze derniers mois, avez-vous reçu de l'argent ou des cadeaux en échange des rapports sexuels ou bien avez-vous payé quelqu'un pour faire l'amour avec vous?	FTF

Key: FTF, face-to-face; SAQ, self-administered questionnaire.

3. Estimating levels of sexual risk behaviour

3.1 Factors which determine the incidence of a sexually transmitted infection

Worldwide, there is great variation in the prevalence of STIs and in patterns of sexual behaviour, but there is little concordance in the variation between the two. Figure 14.4 shows schematically some of the factors which theoretically determine the incidence of an STI, using the example of HIV. The first box shows societal factors which determine general patterns of sexual behaviour and sexual mixing. The second shows the characteristics which influence whether the sexual contact is potentially infectious, i.e. whether a person is exposed to infection. The third shows the mediating factors, which affect the potential for transmission of infection from the infected partner. The fourth box shows those factors which determine whether or not the contact results in a new infection.

Table 14.3 shows some of the factors which have been found to be associated with HIV infection in the general population in a variety of studies.[2] The Ugandan samples are from cohort studies, which were designed to elucidate some of these relationships. Table 14.4 shows some of the behavioural changes which have been reported at the same time as observed HIV prevalence has decreased, as has happened in some countries, most noticeably Uganda and Thailand. Changes in HIV prevalence can be attributed to changes in behaviour if incidence has also decreased, but it is difficult to establish if this is the case because prevalence can decline due to excess mortality among people already infected with HIV.

Distal determinants of behaviour

Age

Age is correlated with whether or not someone is sexually active and the likelihood that their sexual partner is their spouse. In countries where the HIV epidemic is of recent origin, older age groups may have a lower cumulative exposure to infection because most of their past sexual

Figure 14.4 Factors which can influence the incidence of HIV infection

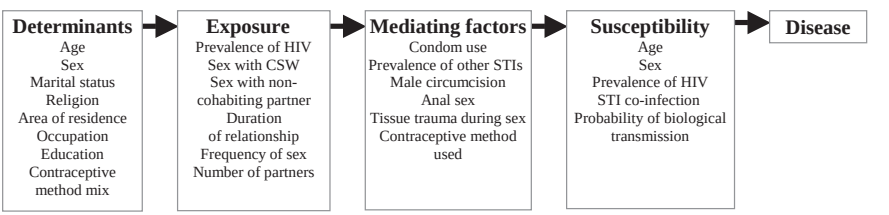
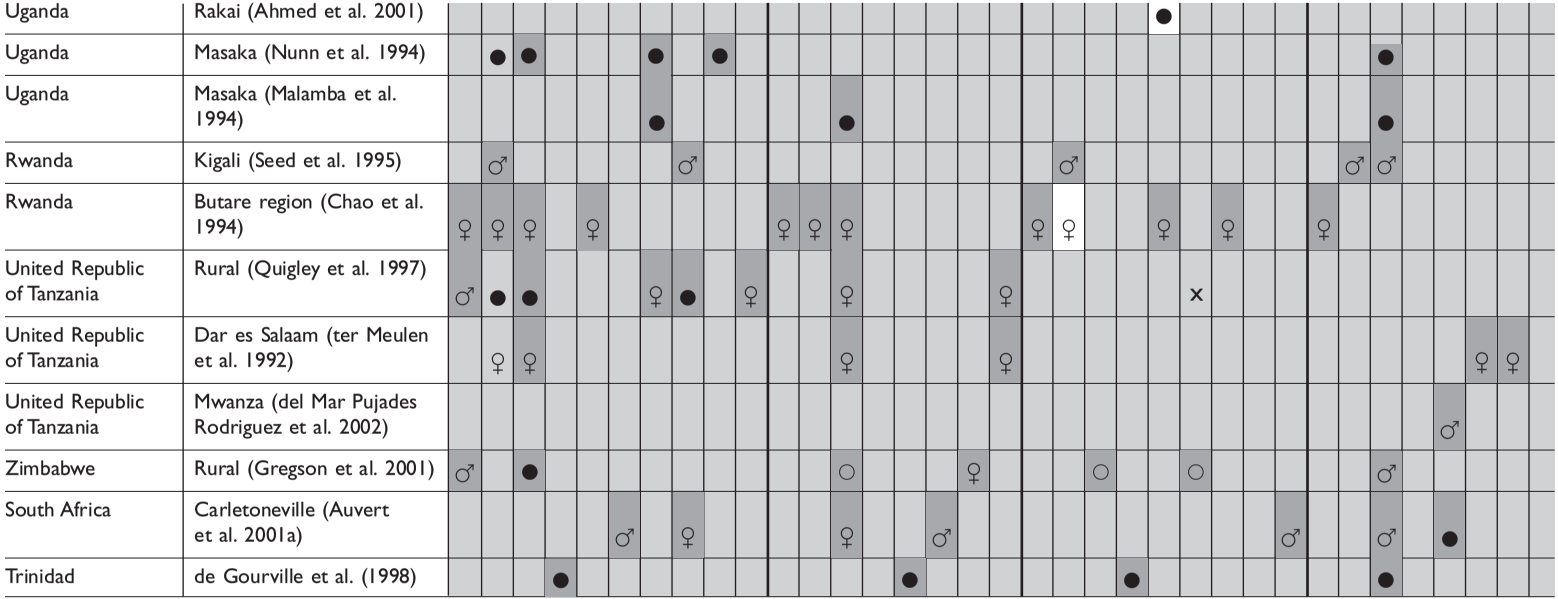


Table 14.3 Factors found to be associated with HIV infection among members of the general adult population.

		Sociodemographic factors											*Exposure to possible sources of infection*									*Factors affecting transmission*									*Other STIs*						
Country or continent	*Area*	*Socioeconomic status*	*Age*	*Not currently married*	*Relationship ≥1 year duration*	*More education*	*Out of school*	*Religion: non-Muslim*	*Place of residence*	*Working outside village*	*Ttravel*		*Younger age at first preg*	*Lower parity*	*Number of partners*	*Type of partner(s)*	*Contact with CSW*	*Forced to have sex*	*Partner visits beer halls*	*Blood transfusion*	*Had sex to support self*	*Lack of male circumcision*	*Drinking alcohol*	*Cocaine use*	*Smoking*	*Condom use*	*Using oral contraceptives*	*Anal sex*	*Lack of penile hygiene*	*History of STI*	*Current STI*	*Genital ulcer disease*	*Gonorrhoea infection*	*HSV-2 infection*	*Syphilis infection*	*Trichomoniasis infection*	*Urethritis*
India	Pune: STI Clinic patients (Rodrigues et al. 1995)																♂				+♀				●		●				●					
Thailand	North: Military conscripts (Nelson et al. 1996)															♂	♂												♂							
Africa	Four Cities (Auvert et al. 2001c)																					♂	♂							♂			♂			
Africa	Meta-analysis (Weiss et al. 2000)																					♂														
Africa	Four Cities (Auvert et al. 2001b)		♀	♀											♀	I+♀	♀				I+♀	♀	♀										●	♀		I♂
Kenya	Nairobi (Hunter et al. 1994)			+♀											+♀							+♀										+♀		+♀		
Uganda	Rakai (Gray et al. 2000)		♂					♂							♂	♂						♂			X								♂			



Key:
CSW Commercial sex worker.

●, both sexes consistently; ○, both sexes inconsistently; ♀, females consistently; ♀, females inconsistently; ♂, males consistently; ♂, males inconsistently; **X**, no association;
■, factor associated with increased risk of HIV infection; □, factor associated with decreased risk of HIV infection; ▣, association observed in both directions.

Table 14.4 Changes in behaviour which have been observed concomitantly with a decline in HIV prevalence or incidence

	Age first had sex	*Age at first marriage*	*Age at birth of first child*	*Number of partners*	*Higher risk sex in the last year*	*Commercial sex*	*Ever used condom*	*Condom used last high risk sex*	*Condom used last commercial sex*	*Ever had STI*	*Sex with a girlfriend*
Males											
Uganda (Asiimwe Okiror et al. 1997)	↑			→	↓		↑	⇧			
Uganda (Kamali et al. 2000)	↑			↑			↑	⇧			
Thailand (Kilmarx et al. 2000)						↓			↑		
Thailand (Nelson et al. 1996)						↓				↓	↑
Zambia (Fylkesnes et al. 2001)				↓			↑	↑			
Females											
Uganda (Asiimwe Okiror et al. 1997)	↑			→			↑	⇧			
Uganda (Kamali et al. 2000)		↑		↑			↑	⇧			
Thailand (Kilmarx et al. 2000)									↑		
Zambia (Fylkesnes et al. 2001)			↑	↓			↑	↑			

Key: ↑, increase in prevalence of this factor observed at the same time as decline in HIV prevalence or incidence; ↓, decrease in this factor observed at the same time as decline in HIV prevalence or incidence; ⇧, inconsistent or non-significant increase; ⇩, inconsistent or non-significant decrease; →, factor did not change significantly.

Note: In two of the countries in this table (Uganda and Thailand), the declines in prevalence are evidence of convincing long-term downward trends in incidence, but in the remainder the decrease in prevalence has not been observed over such an extended period, and may in fact be the result of rapidly increasing number of AIDS deaths masking a steady incidence of new infections, a phenomenon that was observed in the early stages of the prevalence decline in Uganda (Wawer et al. 1997).

exposure occurred at a time of low prevalence. The association of HIV infection with young age was not seen in all the studies listed in Table 14.3, and where an association was found it was not always in the same direction and was sometimes different for men and women.

Sex

In most cultures, men and women initiate sexual activity at different ages. The typical age difference between partners may be different for men and women. Societies may condone some sexual behaviours for men and not for women. In countries with generalized epidemics, prevalence is usually higher in women than men, especially in younger age groups.

Travel, place of residence, workplace

Factors such as travel, area of residence and occupation or place of work have been measured differently by the various studies. In those studies where these factors were associated with HIV infection (Auvert et al. 2001a; Nunn et al. 1994; Quigley et al. 1997; Seed et al. 1995) they could be acting as proxy measures for potential contact with infected sexual partners. These factors will all influence the number of sexual partners and proportion of available partners who are infected.

Religion

In the studies which found religion to be associated with HIV infection, the comparison was between Muslims and non-Muslims (Malamba et al. 1994; Nunn et al. 1994; Quigley et al. 1997). There are two characteristics of Muslims which may be relevant to HIV infection: the customary practice of male circumcision and the requirement to abstain from alcohol. The use of alcohol and other mood-altering substances is an independent risk factor in other studies (Auvert et al. 2001b, 2001c; de Gourville et al. 1998; Gregson et al. 2001). The social values incorporated in the Muslim faith may also cause people to have risky sex less frequently.

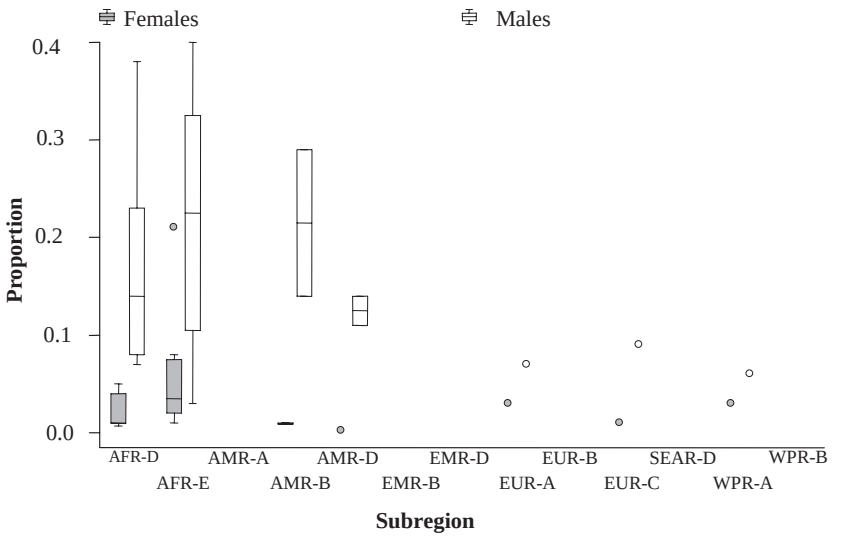
Marital status

Married people usually have sex more frequently than unmarried people. In most countries, being sexually active outside of a co-resident (cohabiting) relationship is associated with an increased incidence of HIV infection. Sex between co-resident partners usually carries a lower risk of infection than sex with other types of partner, so prevalence may be lower among married people. This may depend on how much extramarital sex is taking place: the proportion of people who have sex outside marriage varies between countries (see Figure 14.5). In countries where HIV prevalence is high, the surviving partners of people who died of AIDS will tend to have a higher prevalence of HIV infection than the group of people who are still married. However, in some places being currently married has been shown to increase the risk of HIV infection. This may be because married people gain an additional sexual partner at the time of marriage (Auvert et al. 2001b) compared to their never-married peers. The increased frequency of (unprotected) sex within marriage may also increase the risk of HIV transmission.

Contraceptive method mix

Condoms may be used to prevent unwanted pregnancies but many couples choose to use non-barrier methods. In many cultures, condoms are not seen as appropriate for use within marriage or in a long-standing relationship. A pattern commonly seen in developed countries is that initial condom use with a new partner is followed by a switch to oral contraceptives after a few months, e.g. France (Commissariat Général du

Figure 14.5 The proportion of married men and women who report having had sex with someone other than their spouse in the last year, by subregion



Note: The figure shows the range of values, within a subregion, for the national estimates of the proportion of married people who report extramarital sex. The line across the middle of the box represents the median value, the box itself spans the interquartile range and the lines extend to the adjacent values at either end of the interquartile range (only shown where the adjacent values fall outside of this range). Data points which fall outside this range are plotted separately. There are no suitable data for AMR-A, EMR-B and EMR-D, EUR-B and SEAR-D.

Plan: Observatoire régional de santé d'Ile-de-France and Agence Nationale de Recherches sur le SIDA 2001). In some populations, negative attitudes towards condoms may lead to lower levels of use.

Exposure

Prevalence of HIV

The proportion of people infected with HIV in the population is the main factor influencing the probability of having sex with somebody who is infectious for HIV.

Sexual mixing patterns

Partner selection would be described as completely assortative if people always chose partners who were similar to themselves in all the measured respects. However, the way in which people select their sexual partners is usually incompletely assortative, that is, people tend to choose partners who are similar to themselves in most, but not all respects. The

differences may be predictable and some mixing patterns can have significant influence over the spread of STIs. For example, age-mixing in sexual relationships (older men with young women) is thought to be an important factor in accelerating the spread of HIV (Anderson and May 1991).

Traditional STI epidemiology defines "core" and "non-core" groups. The incidence of infection is high in the core groups, and most transmission occurs within these groups. The core group is composed of people who have a large number of sexual contacts compared to the rest of the population. Core groups tend to be small, and as long as infection remains confined to these groups, the population prevalence will remain fairly low. Since mixing patterns are incompletely assortative with respect to frequency of partner change, there will be occasional contacts between members of the core group and the rest of the population. The people involved in these sorts of partnerships are known as "bridge" groups and provide the route by which an infection moves from the core group to the general population. An example of this would be married men who visit commercial sex workers: married men are mostly members of the non-core group, commercial sex workers are members of a core group and the subset of married men who visit the sex workers is the bridge group. Simple measures of partner change and proportions exposed in either group fail to capture variation in density of exposure which arises from non-random mixing (Anderson and Garnett 2000).

Number of partners

If condom use is not widespread in the population, then having a greater number of sexual partners means being exposed to a greater risk of infection. This is probably not a linear relationship because in many countries a disproportionate number of STIs occurs among the small group of people who have numerous partners. Most of the surveys listed in Table 14.3 found an increasing risk of infection (Auvert et al. 2001b; Chao et al. 1994; Hunter et al. 1994; Quigley et al. 1997; ter Meulen et al. 1992) and seroconversion (the detection of antibodies to HIV in a person who has not previously produced such antibodies, indicating recent infection with HIV) (Gray et al. 2000) associated with increasing numbers of partners. The reference periods were not the same in these surveys so it is not possible to compare the magnitude of the increased risk; this pattern was not clear in all studies. In the Masaka cohort in Uganda, the effect of the number of partners seemed to be modified by age. There was a greatly increased risk associated with having more partners for those aged <25 years, but no clear pattern among older people (Malamba et al. 1994). In the Four Cities study, women reporting a greater number of lifetime partners had a significantly increased risk of being infected with HIV in Kisumu (Kenya), Ndola (Zambia) and Yaoundé (Cameroon) but not in Cotonou (Benin) (Auvert et al. 2001b). Only in Ndola (Auvert et al. 2001b) was an increased chance of being HIV-positive observed among men reporting a higher number of lifetime partners.

In Uganda, a reduction in the number of partners does not appear to have been necessary for a decline in prevalence to occur (Asiimwe Okiror et al. 1997; Kamali et al. 2000). In Zambia, localized decreases in the prevalence of HIV among young women attending antenatal care clinics were observed between 1994 and 1998, and the proportion of people reporting large numbers of sexual partners in the same area in coincident general population behavioural surveys was also seen to decline (Fylkesnes et al. 2001).

Commercial sex

Contact with sex workers, a group that often has a high prevalence of HIV infection, seems mainly to be important outside of Africa. Commercial sex is difficult to define in a meaningful way across cultures because the exchange of money or gifts may generally accompany sex in some cultures, but this may not mean that the woman has a great many partners, or that she is demanding the payment in return for sex.

Duration of relationships

Sex within a relationship that has been established for a long time is thought to carry a lower risk of HIV infection than sex with a more recently acquired partner. Logically, this would only be the case if both the partners were mutually monogamous throughout the duration of the relationship. It may be that mutually monogamous partnerships last longer than others and that the observed association is a selection effect.

Mediating factors

Male circumcision

In Africa, male circumcision is associated with a lower probability of male HIV infection (Auvert et al. 2001c; Gray et al. 2000; Hunter et al. 1994; Seed et al. 1995; Weiss et al. 2000). There is a plausible biological mechanism for this (Glynn et al. 2001; Royce et al. 1997), although its importance outside of Africa remains to be measured. It is also unclear whether a circumcised, infected man is less likely to transmit the infectious agent to a female partner than an uncircumcised man.

Sexually transmitted infections

HIV infection is likely to be associated with a history of infection with another STI because these agents share the same mode of transmission. Being infected with an STI indicates that a person has had a sexual contact which could also have led to HIV infection, if their partner was infectious for HIV. However, it has been found that, in addition to providing a marker for this type of contact (Obasi et al. 1999), the presence of another active disease increases the risk of both HIV transmission and infection (Mbopi Keou et al. 2000). In the studies summarized in Table 14.3, relevant information was collected for different diseases in differ-

ent ways. This is because the locally important STIs vary and the setting of the studies imposes restrictions on the information collected. However, in all the studies, having ever had another STI clearly increased the chance of being infected with HIV.

Condom use

The efficacy of condoms in preventing the transmission of HIV and other STIs has been established (Weller and Davis 2002). However, only one of the studies listed in Table 14.3 (a study carried out among men attending an STI clinic in India (Rodrigues et al. 1995) found a protective association between reported condom use and HIV infection. The reason for this may be that in African countries condom use is actually a marker for risky sex. That is, condoms are only used by those who (rightly) perceive themselves to be at risk of infection. In this case, condom use would only be protective if condoms were properly used at every risky encounter. Condom use would only be revealed as protective in a statistical analysis if this could be confined to those who indulge in risky sex, or if the propensity to have risky sex could be controlled for. If members of groups at a high risk of HIV infection were initially more likely to use condoms, a protective effect would only become apparent as condom use became more widespread in the general population. The availability and acceptability of other methods of contraception might affect the chances of a couple using a condom.

Sexual practices

Anal sex, both in homosexual male and in heterosexual couples, carries a higher risk of transmission than other practices. It has been suggested that drying the vagina before sex, and having sex during menses also increase the risk of HIV infection in women. However, this has not been clearly demonstrated (Auvert et al. 2001b; Buve et al. 2001a; Malamba et al. 1994).

Susceptibility

There is a high incidence of HIV infection among young women who have become sexually active at an early age. A partial explanation for this observation may be that young women are particularly vulnerable to HIV infection because the immaturity of the genital tract renders them physiologically susceptible. This is a complex issue, as demonstrated by the results of the Four Cities study, which showed that the high prevalence of HIV infection among young women was not fully explained by behavioural factors (Glynn et al. 2001).

In Europe, transmission from males to females has been observed to be more efficient than vice versa (Anonymous 1992) but this was not confirmed in Rakai (Uganda) (Quinn et al. 2000). This pattern of differential transmission probabilities between the sexes is inconsistent in the rest of the world (Mastro and Kitayaporn 1998).

3.2 Choice of indicators of potentially hazardous sexual behaviours

Sexual behaviour can be summarized by a variety of different measures and, as described above, many of these measures have been found to be associated with HIV infection. However, it is also clear that these associations are not found in all populations, nor are they consistent in direction and magnitude across those populations in which an association has been observed. The most appropriate indicators of potentially hazardous behaviour were judged to be those which have been associated with HIV infection in different settings, and which:

- are available and relevant for all age groups, both sexes and all subregions;
- describe an important aspect of behaviour in all subregions; and
- are associated with the risk of acquiring HIV infection, or with being already infected with HIV.

First, an empirical approach was used to identify this subset. Population-level estimates are available for many of these behavioural indicators and estimates of HIV prevalence are also available for many countries. However, it is well known that there is no simple relationship between observed HIV prevalence and reported sexual behaviours at the population level. In many African countries with generalized epidemics, the national prevalence estimates are based on fitting a mathematical model of the HIV epidemic to observed HIV prevalence data acquired from among women attending antenatal care clinics. An estimate of the model parameter representing the fraction of the population that is at risk of contracting HIV infection was extracted from the model and a regression analysis was conducted to examine the association between this estimate and various indicators of sexual behaviour. The model, known as the Epidemic Projection Package (EPP), is described in detail in section 4.

Estimates for the behavioural indicators were calculated for all countries with data, as described in Table 14.1. Suitable model fits were available for 16 countries, and a complete set of indicators and model fits were available for nine countries. Each country contributed an urban and a rural estimate, bringing the sample size for the regression analysis to 18.

All analyses were carried out in Stata version 7.0. Correlation coefficients were calculated for each combination of model parameter and behavioural indicator. The results of these correlations governed which behavioural indicators were included in a linear regression model. This model also included another parameter, which describes the force of infection, as an instrumental variable. It was not possible to quantify a relationship between the behavioural data and the model parameter using this method. The analysis was hampered by the small sample size

and the associations that did emerge as statistically significant were not easy to interpret. Some indicators whose effects would be expected to be similar (such as age at first sex and the proportion of the population who had ever had sex), when included in the same regression model produced opposing coefficients. A robust analysis would require a much larger sample size than was available, given the large number of behavioural indicators and the high degree of correlation between these indicators.

The failure of our work, and that of other groups, to find a suitable quantification suggests that there may be no single relationship between any one measure of sexual behaviour at the aggregate level and the incidence of HIV infection in the general population. Rather, data at the level of individuals and their partners before infection may be required. The choice of which indicators to present was therefore governed by which indicators were commonly found to be associated with HIV infection in observational studies and the measures recommended by UNAIDS (2000), even if the nature of the association with HIV infection was not clear.

3.3 Prevalence of potentially hazardous sexual behaviours

Different sexual behaviour patterns are summarized here by three measures: lifetable median age at first sex; mean number of sexual partners in the last year; and the proportion of adults in the subregion who have had sex with a non-co-resident partner within the year preceding the survey, and who did not use a condom the last time they had sex with this partner. All the indicators were calculated for individual countries and the subregional estimates were created by weighting these estimates by the total population size of the country relative to the subregion. The number of countries and sample size used for each estimate are given in Table 14.5.

No subregions were completely described and there were no data at all for some subregions. Values had to be estimated for the missing categories by extrapolation of the results from other subregions; this was based primarily on the values of the available estimates. If no estimates were available for a subregion, the values were extrapolated from the subregion with the most similar proportion of people currently married (Figure 14.6). Throughout the results, extrapolated estimates are shown in the shaded cells as explained in the footnotes of the tables.

Median age at first sex

The median age at which people first had sex is presented in Table 14.6. This was calculated from the reported age at first sexual intercourse, or current age for people who have not yet had sex. Lifetable techniques were used to calculate this measure to allow for the inclusion of those people who had not yet had sex. The age of sexual debut is important because it affects the duration of exposure to STIs. There is evidence that

Table 14.5 Numbers of people and countries on which the estimates for the behavioural indicators considered were based, by subregion, sex and age

	Ever had sex						*Had sex in the last year*						*Sex with non-co-resident partner in last year ("higher-risk" sex)*						
	Females			*Males*			*Females*			*Males*			*Females*			*Males*			
Subregion[a]	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*N countries*[b]
AFR-D	52 570	32 091	6 679	13 912	9 512	5 351	36 747	28 471	6 212	8 286	8 317	4 627	20 020	14 323	2 664	6 388	6 426	3 431	26
	14	14	14	12	12	12	13	13	13	11	11	11	7	7	7	8	8	8	
AFR-E	52 502	28 165	6 220	11 645	6 613	3 150	36 076	26 360	5 773	8 304	6 372	3 036	20 807	13 954	2 713	6 374	5 035	2 291	20
	13	14	14	10	10	10	13	13	13	9	9	9	8	8	8	8	8	8	
AMR-B	22 461	16 133	3 589	4 900	2 889	1 632	14 256	15 501	3 475	4 384	2 878	1 622	6 424	6 998	1 363	3 452	2 747	1 487	26
	5	5	5	3	3	3	5	5	5	3	3	3	2	2	2	3	3	3	
AMR-D	35 205	23 372	5 290	3 789	2 557	1 402	19 544	21 937	5 065	2 758	2 511	1 368	7 058	9 003	1 865	1 461	1 283	647	6
	5	5	5	3	3	3	5	5	5	3	3	3	1	1	1	2	2	2	
EUR-A	6 119	6 975	881	6 359	7 090	922	5 364	6 922	876	5 603	6 992	914	2 811	3 148	—	2 817	3 162	—	26
	8	8	7	8	8	7	8	8	7	8	8	7	1	1	—	1	1	—	
EUR-B	4 417	3 222	616	—	—	—	4 362	3 197	613	—	—	—	—	—	—	—	—	—	16
	2	2	2	—	—	—	2	2	2	—	—	—	—	—	—	—	—	—	
EUR-C	4 475	2 047	508	582	566	292	1 219	2 004	497	—	—	—	1 203	1 946	422	385	543	269	9
	2	1	1	1	1	1	1	1	1	—	—	—	1	1	1	1	1	1	
SEAR-B	—	—	—	—	—	—	—	—	—	—	—	—	789	739	147	546	498	82	3
	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	1	1	1	
SEAR-D	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7
	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
WPR-A	682	791	365	675	639	321	655	780	348	647	632	315	611	682	264	616	565	261	5
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	
WPR-B	7 412	5 355	1 158	—	—	—	7 305	5 250	1 146	—	—	—	—	—	—	—	—	—	22
	1	1	1	—	—	—	1	1	1	—	—	—	—	—	—	—	—	—	

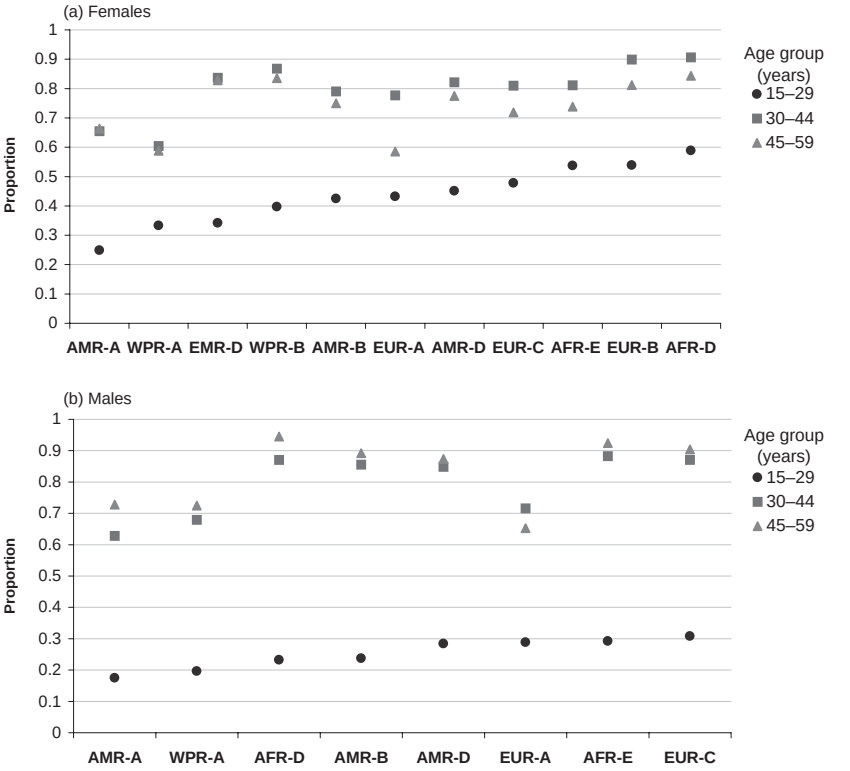
	Condom use last higher-risk sex						*Number of partners*						*Age at first sex*						
	Females			*Males*			*Females*			*Males*			*Females*			*Males*			
Subregion[a]	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*	*N countries*[b]
AFR-D	3 942	999	102	4 308	1 568	365	23 842	14 041	3 037	11 758	7 151	3 911	43 507	26 158	5 465	16 377	10 519	6 057	26
	7	7	5	8	8	8	6	6	6	9	9	9	11	11	11	11	11	11	
AFR-E	4 317	1 414	220	4 178	1 205	315	29 204	15 020	3 239	11 140	6 101	2 948	50 776	27 047	5 581	13 371	7 502	3 516	20
	7	7	6	7	7	7	8	8	8	8	8	8	10	10	10	7	7	7	
AMR-B	1 762	863	120	2 571	786	312	11 354	8 766	2 183	2 661	1 856	950	20 490	14 588	3 309	6 582	4 216	2 347	26
	2	2	2	3	3	3	3	3	3	2	2	2	4	4	4	2	2	2	
AMR-D	680	282	32	600	221	91	14 619	10 665	2 513	1 149	814	429	35 317	23 360	5 192	3 770	2 546	1 427	6
	1	1	1	1	1	1	1	1	1	1	1	1	5	5	5	2	2	2	
EUR-A	806	363	—	1 200	419	—	5 047	5 601	689	5 305	5 659	731	5 903	6 796	683	4 642	5 487	668	26
	1	1	—	1	1	—	10	10	8	10	10	8	6	6	5	6	6	5	
EUR-B	—	—	—	—	—	—	—	—	—	—	—	—	4 389	3 226	643	836	2 108	868	16
	—	—	—	—	—	—	—	—	—	—	—	—	2	2	2	1	1	1	
EUR-C	240	196	30	237	108	21	2 151	2 130	522	573	555	291	2 097	2 043	515	578	570	290	9
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
SEAR-B	—	—	—	—	—	—	789	739	147	546	498	82	11 558	15 300	3 603	546	498	82	3
	—	—	—	—	—	—	1	1	1	1	1	1	2	2	2	1	1	1	
SEAR-D	—	—	—	—	—	—	—	—	—	—	—	—	4 091	3 177	700	—	—	—	7
	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	—	—	—	
WPR-A	26	11	10	35	23	9	144	297	231	120	262	206	171	309	249	675	640	321	5
	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	
WPR-B	—	—	—	—	—	—	—	—	—	—	—	—	7 333	5 380	1 199	—	—	—	22
	—	—	—	—	—	—	—	—	—	—	—	—	1	1	1	—	—	—	

— No data.

[a] Upper number for each subregion refers to number of people, lower number refers to number of countries from which data were available.

[b] Total number of countries in the subregion.

Figure 14.6 The proportion of people who are currently married, by age and subregion



Note: The lowest proportion of older women who are currently married is found in EUR-A, despite the fact that EUR-A falls in the middle of the range for the two younger age groups. This could be due to a larger proportion of women who never marry, or a higher incidence of marital dissolution in this subregion compared to the others.

young women are more susceptible to HIV infection and that people who start to have sex at a younger age may have more risky behaviour over a lifetime than those who delay the first time they have sex. Values for AMR-A were extrapolated from Australia and New Zealand. These values were used instead of those for the WPR-A subregion as a whole because the latter subregion is very heterogeneous and AMR-A is very similar to Australia and New Zealand for the other indicators, where there are data. The values for the EMR-B and EMR-D were extrapolated from EUR-C.

Table 14.6 The median age at first sex: lifetable estimates

	Females			*Males*		
Subregion	*15–29*	*30–44*	*45–59*	*15–29*	*30–44*	*45–59*
AFR-D	17.3	16.5	17.1	19.7	19.4	20.3
AFR-E	17.5	16.2	15.9	18.9	18.2	19.3
AMR-A	17.5	17.5	19.5	17.5	17.5	18.5
AMR-B	18.6	19.5	20.2	16.5	16.5	16.5
AMR-D	19.4	18.4	18.4	17.5	18.0	18.5
EMR-B	20.5	20.5	20.5	18.5	19.5	19.5
EMR-D	20.5	20.5	20.5	18.5	19.5	19.5
EUR-A	18.5	18.6	20.5	17.8	17.8	18.3
EUR-B	19.5	19.7	20.3	20.3[c]	20.8[c]	21.3[c]
EUR-C	20.5	20.5	20.5	18.5	19.5	19.5
SEAR-B	19.16[a]	19.0[a]	18.2[a]	18.5	18.5	20.5
SEAR-D	16.5[b]	16.5[b]	15.5[b]	18.5	18.5	20.5
WPR-A	18.8	18.8	20.1	19.0	19.0	19.6
WPR-B	23.5	21.5	21.5	20.9	20.13	19.1

[a] Estimate includes the results of an Indonesian survey of ever-married women.

[b] Estimate based on DHS of ever-married Nepalese women.

[c] Estimate calculated from published medians reported for Polish men in different age groups, only that for men aged 30–44 years is complete.

Note: Extrapolated estimates are given in the shaded cells.

Sex with non-co-resident partners

The proportion of all people who report having had sex within the last year, with a partner with whom they do not live, and who did not use a condom the last time they had sex with that partner is perhaps the closest measure of unsafe sex that it is feasible to calculate and is our working definition of risky sex. Sex outside of a cohabiting (co-resident) partnership (within the last year) without using a condom is thought to carry a greater risk of HIV infection than marital sex. As shown in Table 14.7, there are striking variations in the levels of this indicator across the subregions, but they do not follow the pattern of HIV prevalence.

Mean number of partners during the last year

Table 14.8 shows the mean number of sexual partners in the preceding year in the adult population (aged 15–59 years), regardless of the relationship to any of the partners. Again, there are clear differences between the subregions.

Table 14.7 The proportion of the adult population (aged 15–59 years) who report having had sex with a non-co-resident partner in the last year, without using a condom on the last occasion

	Females			*Males*		
Subregion	*15–29*	*30–44*	*45–59*[a]	*15–29*	*30–44*	*45–59*
AFR-D	0.116	0.061	0.049	0.239	0.161	0.090
AFR-E	0.108	0.075	0.067	0.230	0.111	0.094
AMR-A	0.070	0.040	0.030	0.090	0.090	0.070
AMR-B	0.110	0.057	0.055	0.218	0.120	0.122
AMR-D	0.016	0.013	0.005	0.289	0.140	0.117
EMR-B	0.073	0.078	0.055	0.216	0.099	0.099
EMR-D	0.073	0.078	0.055	0.216	0.099	0.099
EUR-A	0.212	0.074	0.074	0.267	0.119	0.119
EUR-B	0.073	0.078	0.055	0.216	0.099	0.099
EUR-C	0.073	0.078	0.055	0.140	0.087	0.048
SEAR-B	0.116	0.061	0.049	0.239	0.161	0.090
SEAR-D	0.116	0.061	0.049	0.239	0.161	0.090
WPR-A	0.068	0.043	0.025	0.091	0.087	0.066
WPR-B	0.068	0.043	0.025	0.091	0.087	0.066

[a] It was assumed that survey data for women aged 15–49 years applied to women aged 15–59 years.

Note: Extrapolated estimates are given in the shaded cells.

4. Risk factor–disease relationship

4.1 HIV

HIV infection is known to be sexually transmitted. Some sexual practices with an HIV-positive partner carry a greater risk of infection than others. In some populations, there are groups of people who can be identified as having a greater likelihood of being infected with HIV. The factors that govern whether a susceptible person chooses one of these people at a higher risk of being infected as a sexual partner will influence their own risk of infection. Sexual behaviour and its determinants are not easy to measure, and can vary in several dimensions, all of which may be pertinent for HIV transmission.

It is hard to model the impact of changes in exposure for an infectious disease with person-to-person transmission because the risk associated with exposure will change with changes in the prevalence of the infection. A sexual contact is only an exposure if one partner is infected with HIV and the other is not, and the likelihood of this occurring will change as the prevalence of infection changes. The social perception of risk may feedback to behaviour and further contribute to change. There-

Table 14.8 The mean number of sexual partners in the last year reported by the adult population (aged 15–59 years)

	Females			*Males*		
Subregion	*15–29*	*30–44*	*45–59*[a]	*15–29*	*30–44*	*45–59*
AFR-D	0.679	0.764	0.738	1.106	1.336	1.097
AFR-E	0.729	0.928	0.809	0.923	1.132	1.040
AMR-A	1.433	1.104	0.834	1.797	1.421	1.116
AMR-B	0.643	0.915	0.826	1.276	1.316	1.154
AMR-D	0.492	0.849	0.742	1.413	1.629	1.235
EMR-B	0.576	0.915	0.774	1.125	1.153	1.010
EMR-D	0.576	0.915	0.774	1.125	1.153	1.010
EUR-A	1.248	0.987	0.912	1.378	1.134	1.130
EUR-B	0.576	0.915	0.774	1.125	1.153	1.010
EUR-C	0.576	0.915	0.774	1.125	1.153	1.010
SEAR-B	0.649	0.842	0.755	4.007	1.941	1.469
SEAR-D	0.649	0.842	0.755	4.007	1.941	1.469
WPR-A	1.236	1.077	0.900	1.792	1.229	1.039
WPR-B	1.236	1.077	0.900	1.792	1.229	1.039

[a] It was assumed that survey data for women aged 15–49 years applied to women aged 15–59 years.

Base: All respondents.

Note: Extrapolated estimates are given in the shaded cells.

fore a relative risk measured for a particular population at a particular point in time is meaningless for another place or point in time, unless the overall prevalence, the epidemic maturity and the degree to which infected and susceptible people mix are almost identical.

An alternative way to predict the future prevalence of an infection which is transmitted from person to person is to use a recursive mathematical projection model to account for the increase in incidence caused by the increase in the number of prevalent cases. Simpler approaches, based on a static risk of infection, will not adequately capture the dynamics of infection over a period of time if prevalence is high, because the risk of infection will change as the prevalence of infection changes.

If prevalence is low, a simpler approach can be justified because the error introduced in the estimates of the number of new infections by ignoring changes in prevalence is much smaller. The size of the error that results from using an approach based on a static level of risk in a high-prevalence situation will depend on the prevalence of the infection, the speed at which prevalence changes and the period of time considered. To illustrate the scale of errors introduced by ignoring the non-linearities of epidemic dynamics, we note that in a population with an

HIV prevalence of 20%, with a concurrent infection rate among HIV-negative people of approximately 3.5% per year, over a five-year period the prevalence could increase by 2% or fall by 3% without any changes occurring in risk behaviour, but depending on the maturity of the epidemic at the time when the HIV prevalence of 20% was reached. Currently, UNAIDS estimates that the prevalence of HIV in nine African countries is in the order of ≥20% among women attending antenatal clinics in urban areas (and in four countries the prevalence of HIV is >20% among women attending clinics in rural areas) (UNAIDS/WHO 2002). The estimates of avoidable infections presented here are for a five-year period. To determine the range of probable outcomes, it is essential that a suitable mathematical model be used to derive estimates of new infections for these subregions, both for the "business-as-usual" scenario, and to estimate what may happen under the different counterfactual scenarios.

Methods for estimating HIV prevalence over a period of time

UNAIDS/WHO make country-specific estimates and projections of HIV infection worldwide and the UNAIDS Epidemiology Reference Group has developed a model to make projections of HIV prevalence. The model has been implemented in a program known as the Epidemic Projection Package (EPP) (The UNAIDS Reference Group on Estimates Modelling and Projections 2002). EPP is designed to represent the evolution of generalized epidemics and so its use for prediction is confined to countries in which generalized epidemics have developed. In this chapter, EPP was used to calculate estimates for the two African subregions (AFR-D and AFR-E) (Table 14.9).

Reasons for using the EPP model

There are a number of models which could have been used for the CRA, but the EPP model, currently used by UNAIDS, was deemed to be the most appropriate. The other available models include deterministic models, such as AVERT (Rehle et al. 1998), but most of these make no allowance for behaviour change. The GOALS model (http://www.futuresgroup.com), developed by WHO and the Futures Group models the impact of interventions concerning behavioural change, primarily from a programme manager's or policy-maker's perspective, with the focus on the cost-effectiveness of different interventions. This model requires a much larger amount of input data than EPP and is not appropriate for longer-term projections. Most of the models designed to explore the effects of different interventions are more complex than EPP. Additional assumptions (such as profiles of commercial sex work) would have been needed for such models to be used, as sufficient data are not always available.

Table 14.9 Countries for which an EPP model fit is available

Country	*EPP fit available*	*Country*	*EPP fit available*
AFR-D		*AFR-E*	
Algeria	—	Botswana	✓
Angola	✓	Burundi	✓
Benin	✓	Central African Republic	✓
Burkina Faso	✓	Congo	✓
Cameroon	✓	Côte d'Ivoire	✓
Cape Verde	—	Democratic Republic of the Congo	✓
Chad	✓	Eritrea	—
Comoros	—	Ethiopia	✓
Equatorial Guinea	✓	Kenya	✓
Gabon	✓	Lesotho	✓
Gambia	✓	Malawi	✓
Ghana	—	Mozambique	✓
Guinea	✓	Namibia	✓
Guinea-Bissau	✓	Rwanda	✓
Liberia	—	South Africa	✓
Madagascar	—	Swaziland	✓
Mali	✓	Uganda	✓
Mauritania	—	United Republic of Tanzania	✓
Mauritius	—	Zambia	✓
Niger	✓	Zimbabwe	✓
Nigeria	✓		
Sao Tome and Principe	—		
Senegal	✓		
Seychelles	—		
Sierra Leone	✓		
Togo	✓		

✓ Available.

— Not available (insufficient data points to fit the model; no generalized epidemic in the smaller countries).

Stochastic models are also available, the prime example being STDSIM (Korenromp et al. 2000; van der Ploeg et al. 1998), which is a complex model requiring detailed specification of a range of demographic, biological and behavioural inputs. STDSIM is designed to closely model the HIV epidemic in small communities and would not have been suitable for use at the international level, despite the fact that it does explicitly model changes in behaviour. A limitation of all sto-

chastic models is the need for repeated runs to ensure reasonably stable results. To run a stochastic model for the countries with sufficient data in all regions of the world would have taken a prohibitive amount of time.

Structure of the EPP model

The mathematical model used is fully described elsewhere (The UNAIDS Reference Group on Estimates Modelling and Projections 2002; UNAIDS Epidemiology Reference Group 2001), but is summarized below using a slightly simplified notation. EPP models both epidemiology, with a feedback loop from prevalence to incidence, and demography, with competing mortality risks and population renewal. This is important because AIDS mortality is a significant factor in the course of the epidemic. The model was deliberately kept simple to allow projections to be based on real data. The model is not subdivided by either age or sex.

Figure 14.7 shows how the model divides a population into three groups (infected, susceptible and at-risk, and susceptible and not-at-risk), and how people can move between these groups over time.

People enter either the at-risk or not-at-risk group on their 15th birthday. Exit from the not-at-risk group is by death from a non-AIDS-related cause. Exit from the at-risk group is either through a non-AIDS-related death or by becoming infected with HIV and moving to the infected group.

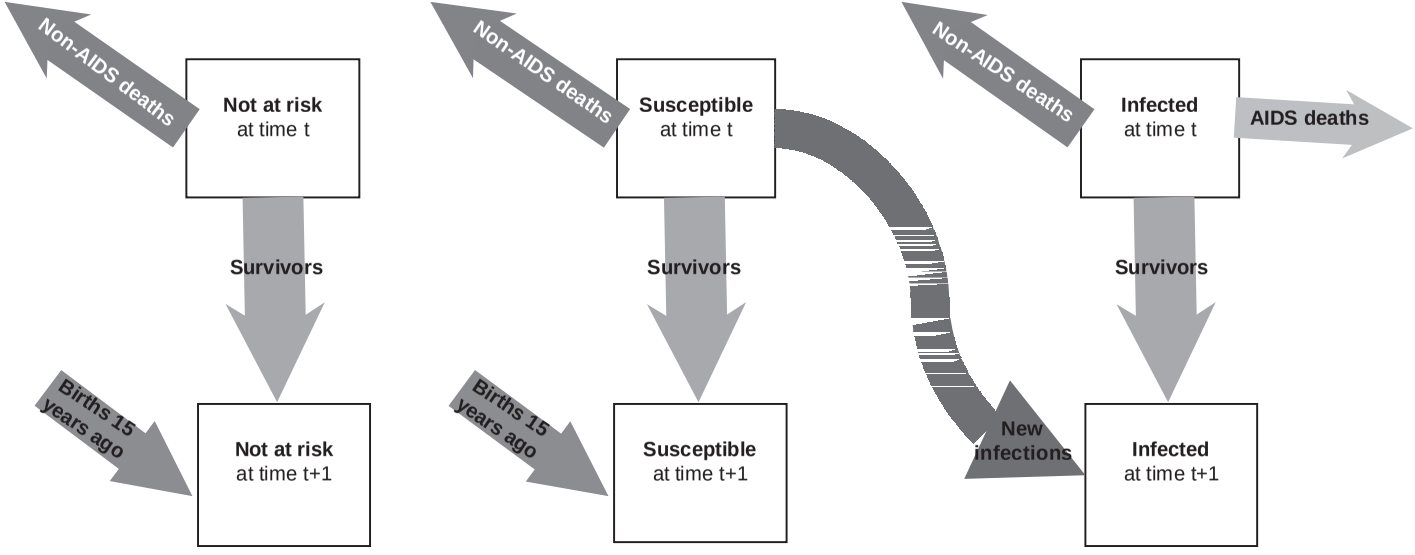
Entry to the population at age 15 years occurs at a constant rate, based on birth rates and rates of survival to age 15 years observed in the population being modelled. Adjustment is made for the impaired fertility of women infected by HIV and for the vertical transmission of HIV. HIV-infected children are assumed not to survive to age 15 years. Death rates from causes unrelated to HIV infection are assumed to be constant. Deaths resulting from AIDS are governed by a mortality function based on a Weibull distribution, which gives survival times after HIV infection. The Weibull survival function is based on data from observational studies in Uganda and the median survival time is compatible with data from Haiti, Thailand and Uganda.

The EPP model is controlled by four main epidemiological parameters:

t_0		The start year for the epidemic
s_0		The initial proportion susceptible
r		Proportionality constant of the force of infection
ϕ	(phi)	The relative recruitment rate into the susceptible category

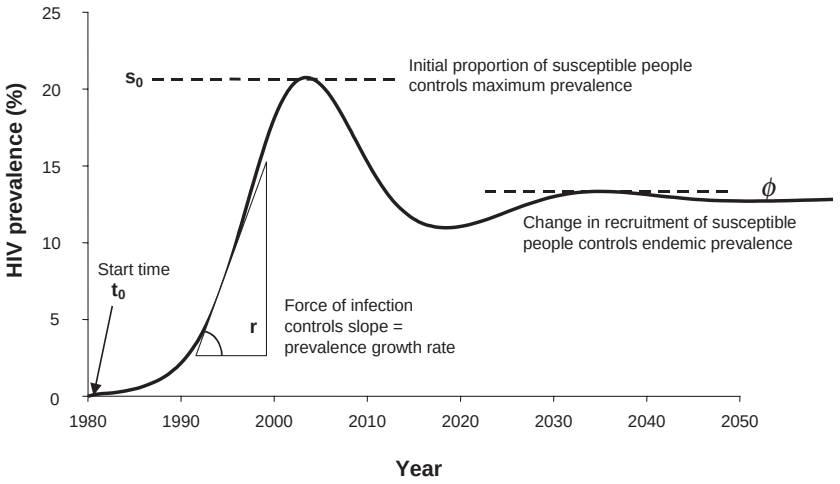
These parameters interact, but their main influence is exercised on the shape and location of different parts of the epidemic curve. These effects are shown in Figure 14.8.

Figure 14.7 Flow of people through the EPP model



Note: Birth and death rates are constant, determined by demographic and epidemiological measurement.
New infection rate varies; it = constant transmissibility factor x probability that partner is HIV+.
Probability that partner is HIV+ = infected / (not-at-risk + susceptible + infected) = prevalence.

Figure 14.8 The effects on HIV prevalence of changes in the main epidemiological parameters in the EPP model



The main demographic parameters governing the model are:

m		Modal survival age after HIV infection (Weibull level parameter)
k		Shape parameter for Weibull survival function
μ	(mu)	Adult mortality rate from non-HIV related causes
λ	(lamda)	Proportion of non-infected children surviving to age 15 years
ν	(nu)	Vertical transmission proportion
β	(beta)	Birth rate for the adult population
δ	(delta)	Low fertility adjustment for HIV-positive adults

In the mathematical exposition below, the following variables are also used, but as they are either derived from the formal parameters listed above, or treated as constants in the normal use of the model, they are not regarded as formal parameters. These are defined below.

Auxiliary constants:

Δ	(Delta)	Time increment for differential equations
ε	(epsilon)	Initial exogenous force of infection

Endogenous variables, dependent on formal parameters:

$\theta(t)$	(theta)	Force of infection between susceptible and infected, at contact time t
$\sigma(x)$	(sigma)	Proportion of infecteds surviving x years after infection

Finally, the numbers and proportions of not-at-risk, susceptible and infected persons at time t are denoted as shown below:

Number		Proportion	
$N(t)$	Not-at-risk population	n_t	Not-at-risk proportion
$S(t)$	Susceptible population	s_t	Susceptible proportion
$I(t)$	Infected population	i_t	Infected proportion = prevalence
$P(t)$	Total population		
$F(t)$	15-year olds entering population	f_t	15-year-old proportion susceptible
		$1 - f_t$	15-year-old proportion not-at-risk

The dynamics of the system are given by the following equations. The number of people aged 15 years entering the adult population at time t is the number of uninfected children born 15 years ago multiplied by the probability of surviving to age 15.

$$F(t) = \lambda\beta[N(t-15) + S(t-15) + (1-v)\delta I(t-15)]$$

The proportion of susceptible people entering the population at time t is a function of the overall current proportion of the adult population that is not at risk, governed by the formal parameters f and s_0, the initial proportion of susceptible people.

$$f_t = f(n_t) = \frac{\exp(\phi[n_t - 1 + s_0])}{\exp(\phi[n_t - 1 + s_0]) - 1 + 1/s_0}$$

Note that since at time zero there are no infected persons, $1 - n_0 = s_0$, so for any value of ϕ, $f_0 = s_0$. Similarly, if $\phi = 0$, then the proportion of susceptible 15-year olds is the same as the initial proportion of those who are susceptible at all times, $f_t = s_0$. If $\phi < 0$, recruitment to the susceptible group declines over time; if $\phi > 0$, recruitment increases.

The Weibull function gives the proportion of those infected surviving x years after infection.

$$\sigma(x) = \exp\left(-\mu x - [x/m]^k\right)$$

The force of infection at the t is given by:

$$\theta(t) = \varepsilon \quad \text{for } t = 0$$
$$\theta(t) = r \cdot i_t \quad \text{for } t > 0$$

Having defined these variable components, it is now possible to formulate the change-of-state equations governing transitions between the population classes.

$$\frac{\Delta N(t)}{\Delta t} = (1 - f_t)F(t) - \mu N(t)$$

$$\frac{\Delta S(t)}{\Delta t} = f_t F(t) - [\theta(t) + \mu]S(t)$$

$$I(t) = \int_{x=0}^{t} \theta(x)S(x)\sigma(t - x)dx$$

The last of these equations is presented as an integral equation rather than a differential, because this is the easiest way to express the fact that the infected population consists of survivors who were infected at a range of times in the past.

Fitting the EPP model to surveillance data

The four epidemiological parameters (t_0, s_0, r and ϕ) were fitted to prevalence data from antenatal clinic surveillance using maximum likelihood fitting. The model was fitted twice for each country, once for the clinics in urban areas and once for those in rural areas.

Alternative implementation of EPP model

The EPP package is designed for use by national AIDS programme managers, to help validate the UNAIDS estimates and projections. Not all the underlying calculations and parameter estimates that are needed for this chapter are the outputs of the standard EPP package, which makes the epidemic scenarios defined by the counterfactual assumptions difficult to create. Therefore, an alternative implementation of the same mathematical model was created as a spreadsheet using the Microsoft Excel program.

Current levels and projections of HIV prevalence

Subregions with a high prevalence of HIV

Estimates of the current levels of HIV infection and projections of future levels are necessary to be able to calculate the proportion of these infections that is attributable to unsafe sex and thus the proportion that is potentially avoidable.

The current estimates and projections of HIV prevalence in the African subregions (under the baseline scenario of no behaviour change) were based on fits of the EPP model to antenatal clinic surveillance data. These projections were prepared by UNAIDS/WHO. The parameter estimates from these model fits were used in the spreadsheet version of the model to calculate the future prevalence, incidence and number of infections for each of the countries concerned. Subregional estimates were created by combining these estimates, weighted by the total population of each country. Weighted estimates were used because the EPP model could not be fitted for those countries with insufficient data on prevalence (11

countries). It is important to note the time scales used in making the model-based estimates. The last available prevalence estimate generally exerts more leverage on the fitted curve than do other points, and a more robust fit is generally obtained when more data points are used. Therefore prevalence estimates for 2001 were included where available and 2001 was taken as the base year for all projections. The projections of avoidable infections extend until 2006 because the model is designed to give reasonably accurate short-term predictions over a five-year period.

Other subregions

For the 146 countries in which the prevalence of HIV/AIDS is low, a different approach was used to model the epidemic. For countries with epidemics that are concentrated in groups with higher-risk behaviour (e.g. men who have sex with men; injecting drug users, sex workers and their clients), a three-step process was followed to produce the current estimates (for the end of 2001) of HIV/AIDS prevalence. First, for each country, groups at the highest risk of acquiring HIV/AIDS were identified and estimates of the size of these groups were made. Next, estimates of point prevalence were made by applying the most recent prevalence rates for these groups to the populations. Finally, prevalence in populations at a lower risk of infection was estimated by allowing for transmission from high to low groups via sexual mixing. This estimate was made in one of the following ways. For countries with data from pregnant women, an adjusted prevalence rate from this group was applied to the number of women of reproductive age (aged 15–49 years) to produce an estimate of the number of women infected via sex with a partner from a group at an increased risk of being infected with HIV. Alternatively, for some countries where the epidemic was more recent and there were no data for populations at a lower risk, assumptions were about the number of infected people at a higher risk who had sexual partners with no other risk of infection. A transmission probability was then applied to produce an estimate of the number of women infected via sex with a partner from a group at a higher risk of being infected with HIV.

Projections of the extent of these epidemics up until 2006 were based on assumptions about saturation levels for each of the groups at a higher risk of infection, the time to saturation, and the spread over time from populations with a high risk to populations with a low risk of being infected with HIV.

For these same 146 countries (excluding countries with a generalized epidemic where EPP was used), trends in prevalence of HIV among groups at a high risk of infection were compiled and compared. Saturation levels for each risk group and time to reach saturation were determined by reviewing available data from countries in the subregion. The particular level of, and time to, saturation were applied to the risk groups

in each country based on current level of prevalence and rate of increase in the groups, and by comparison with saturation levels and rates in neighbouring countries.

Using this approach, we projected low growth for countries with long-running and relatively stable epidemics (e.g. Brazil, Myanmar, the United Kingdom). For countries with recent epidemics, but rapid rates of growth, the projections show much higher rates of increase (e.g. China, Estonia). For all of these countries, we assumed that there was no general heterosexual transmission except from individuals in groups at a higher risk of infection to their lower-risk sexual partners. These procedures, which have been previously described, gave us projections of adult HIV prevalence over time (Stover et al. 2002).

Estimates of incidence were made by using assumptions about survival (median adult survival for those without highly active antiretroviral therapy—HAART—was nine years), growth of populations and levels of accessibility to treatment with HAART. The specific assumptions and procedures used to translate prevalence into estimates of incidence and mortality have been described in detail elsewhere (Stover et al. 2002; The UNAIDS Reference Group on Estimates Modelling and Projections 2002; Walker et al. 2003).

Attributable infections and disease burden

In most subregions, some data were available on the probable mode of transmission for at least a sample of prevalent infections. These data have been used to estimate how many infections were sexually acquired in each subregion. The estimated burden due to unsafe medical injections and blood transfusions was taken from chapter 22 and from a WHO/UNAIDS review of blood safety. To calculate the proportion of infections that results from unsafe sex, the numbers of all people who, according to the model, were infected via unsafe blood transfusions, unsafe medical injections (based on the subregional level estimates) or due to injected drug use (based on country-level estimates) were combined to form the group infected via non-sexual transmission. The number of infections remaining, i.e. those acquired via sexual contact (either heterosexual or homosexual), was divided by the total number of infections to give the percentage of infections due to unsafe sex.

However, to estimate how many of the HIV infections prevalent in 2001 were truly attributable to unsafe sex, it is not enough to simply calculate how many infections arose from unsafe sex at a particular point in time. The burden of infections which result from unsafe sex is determined by the total number of cases of sexually transmitted HIV infection that have arisen since the beginning of the epidemic. In countries with low-level epidemics, estimates of attributable infections based on the mode of transmission of prevalent cases and estimates which account for the effects of past sexual transmission will be broadly similar. In

countries where prevalence is high, there will be a greater discrepancy between the two estimates. We calculated additional estimates for countries with a high prevalence by re-running the EPP model using the fitted value of the s_0 parameter (the initial proportion at risk) reduced to 5% of its original value. This value was used because it is estimated that 5% of HIV transmission is due to unsafe injections and blood transfusion in these countries (all in the WHO African Region). This estimate is based on the probable mode of transmission for existing infections. Estimates of HIV prevalence based on this reduced value of s_0 demonstrate what might have happened had there never been any sexual transmission in these populations. The difference in the number of infected persons estimated in 2001 and the number predicted by the model for 2001, under the altered circumstances, was taken to be the number of infections which were attributable to sexual transmission (see Table 14.12). The results shown for the two African subregions correspond to attributable burden, as defined by the CRA methodology. The results presented in Table 14.12 for the other subregions are an approximation of attributable burden, based only on the exposure of prevalent cases. To obtain better estimates of attributable burden in these subregions, we would need information on the patterns of sexual mixing between the groups at a high risk of infection and the general population for the duration of the epidemic.

The fraction of infections attributable to unsafe sex was applied to the mortality and disease burden (Table 14.13). Prevalent HIV infections are the result of episodes of HIV transmission which occurred over the 15 or so years before measurement. Prevalent AIDS cases and recent deaths will be, on average, the product of transmission patterns from approximately 10 years before measurement (in populations where there is no treatment for AIDS). The estimates for the non-African subregions are based on the assumption that the ratio of sexual to non-sexual transmission has not changed significantly over that time. The model-based estimate for Africa accounts for this possibility. If the ratio of sexual to non-sexual transmission has changed significantly over time, the estimates of attributable disease burden based on the current ratio may be inaccurate.

Avoidable infections

The counterfactual exposure scenarios

As described earlier, it was not possible to measure relationships between specific behaviours and the risk of acquiring HIV infection in a way that could be generalized to all populations. It may be that consistent relationships of this sort do not exist. Therefore counterfactual exposure scenarios cannot be defined in terms relating to measurable changes in behaviour. Predicting changes based on hypothetical scenarios, which are

not linked to a particular group of behaviours but to corresponding model parameters, is the best possible method for estimating how many future infections are potentially avoidable.

The counterfactuals were defined in a way that could be applied in subregions with both low and high prevalence. The counterfactual scenarios selected relate to decreases in the number of people having unsafe sex as represented by model parameters. The scenarios were chosen to provide a range of estimates based on proportional changes in the size of the at-risk group. The counterfactuals were operationalized differently for countries with low and high prevalence because the methods used to project future HIV prevalence in the two situations require different inputs. Three levels of reduction in unsafe sex were used in the calculation of the avoidable proportion of future infections: 100%, 50% and 10%. These levels were achieved by estimating what would happen if all, 50% or 10% of the people who were having unsafe sex immediately ceased doing so. In theory, the intermediate counterfactual scenarios (50% and 10% reductions) could have been engineered to describe a situation in which those who were having unsafe sex reduced the amount of unsafe sex that they were having. The net effect would be the same because the approach is based on person-time at risk, and assumes that length of exposure is proportional to risk of infection.

Reversibility

Infection with any of these STIs need only occur once to produce disease. Therefore removing exposure to the STI will automatically reduce the risk of infection with immediate effect and this is demonstrated by the results of the HIV prevalence projections under the different counterfactual scenarios. However, in reality it is unlikely that all exposure could be removed and the spread of infection reversed at a particular point in time. Infectious people will remain in the population even if all risky behaviour ceases. Unless every person with an infection (married and unmarried alike) stopped having sex without a condom (i.e. if there was no unsafe sex) they would continue to infect new people. This is the reason for considering counterfactual scenarios that include partial reduction in unsafe sex, as described above.

Countries with a high prevalence of HIV infection

It is possible to define counterfactuals in terms of changes in the size of the EPP model's at-risk group for the countries in the two African subregions. Reductions were made to the size of this group at the start of 2001, first by moving a specified fraction of the at-risk group to the not-at-risk group, and second, by slowing recruitment to the at-risk group by the same amount. Three reductions in the original size of the at-risk group were made: 10%, 50% and 95%. The greatest reduction (resulting from total cessation of unsafe sex) thus resulted in only 5% of the

original at-risk group remaining at risk after 2001 and recruitment to this group was cut to 5% of its former level. The size of this group was not reduced to zero because a certain fraction of HIV-infected people will continue to contract their infection through a non-sexual mode of transmission in a non-generalized epidemic. This proportion is estimated to be 5% of infections in sub-Saharan Africa. While some people will contract their infection in one way, and transmit it in another, the degree to which this happens cannot be estimated for this work.

The ratio of sexual vs non-sexual transmission among those already infected is known, but the ratio of sexual vs non-sexual exposure among the uninfected is not. Implicit in the use of a 95% reduction in the at-risk group as the theoretical minimum level of unsafe sex is the assumption that these ratios are the same. This may be incorrect because if a mode of transmission is very efficient (e.g. infected blood transfusion) then the incidence of infection among susceptible people who are exposed in this way may be higher than that among people who are otherwise exposed to HIV infection. If different modes of transmission have different rates of infection, the modes most likely to produce an infection will be over-represented among cases of infected people in comparison with the distribution of the different exposures among uninfected people. If the non-sexual modes of transmission in Africa are significantly more efficient than sexual transmission, then the fraction of the at-risk group which is exposed to HIV infection via sexual transmission may be >95%. However, the opposite could also be true if unsafe medical injections were the most common form of non-sexual transmission; such injections may be associated with a lower infection rate because the reused syringe does not always come into contact with the body fluids that could potentially transmit the infection. There is no means to assess the relative exposure to the different modes of infection and we must instead rely on the data from HIV-infected people, therefore 95% is an uncertain assumption.

The changes to the model were made through the s_0 parameter, and not the r parameter because the latter represents the transmission of infection, and the former describes the fraction of the population that is at risk of infection. Conceptually, transmission can be affected by changes in the level of unsafe sex (e.g. the proportion of sexual acts protected by condoms) but this could not be used satisfactorily to describe a counterfactual scenario. To model a total cessation of unsafe sex, we could not reduce r to zero because this would correspond to a total cessation of all HIV transmission. It is not possible to calculate a value of r which is related to the cessation of sexual transmission only. Two implicit assumptions in this approach are worthy of comment:

- a proportionate relationship between hazardous and unsafe sex: we have assumed that, when we reduce the size of the group of people

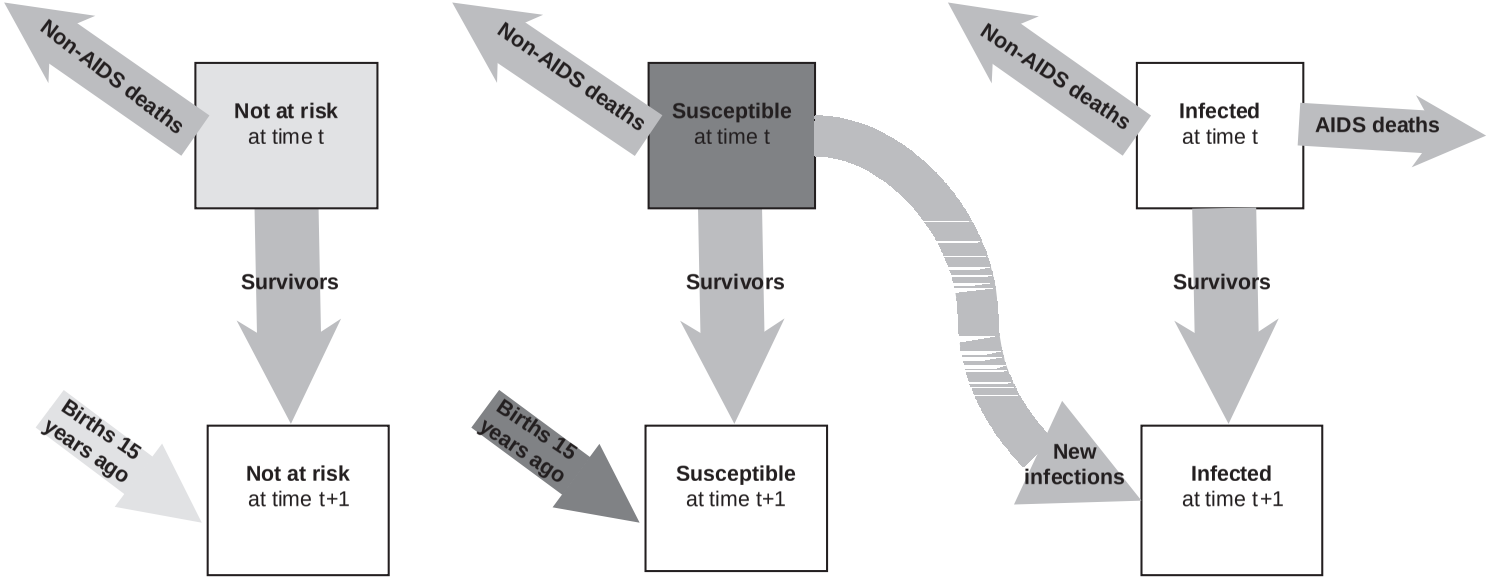
who have hazardous sex, the size of the group having unsafe sex will decrease by the same amount.

- random mixing in the at-risk population: the EPP model assumes random mixing, i.e. each person in the population has an equal chance of contacting another member. This assumption gives a good representation of the natural dynamics of a generalized epidemic of an STI. The question arises, in relation to the counterfactual scenarios, of whether random mixing is still a reasonable assumption in relation to the "hard core" of those remaining at risk after the sudden decrease in hazardous sexual behaviour. We would argue that in the case of sub-Saharan Africa, where the alternative modes of transmission are predominantly unsafe medical injections and unsafe blood transfusions, random mixing is still a close approximation. In the case of injecting drug users, one might want to model far more intensive contacts within the group of people at risk than outside of it, but use of injected drugs is not as important in these subregions as it is elsewhere in the world.

The spreadsheet (Excel) implementation of the EPP model was used, after the modifications described below were made in order to include the changes described by the different counterfactual scenarios. Recruitment of new members into the at-risk group was slowed by a specified amount, as defined in the counterfactual, starting in 2001 and continuing until the end of the projection in 2006. The slowing of the recruitment to the at-risk group was achieved by reducing the value of the s_0 parameter by the specified fraction, with effect from 2001. All the other model parameters remained unchanged. At the start of 2001, the size of the at-risk group was reduced to the fraction of its former size defined in the counterfactual, and the people removed from this group were added to the not-at-risk group. These modifications had the effect of reducing the pool of people who could potentially become infected with HIV, and therefore lowered the number of new cases occurring. Figure 14.9 shows how these modifications affected the projected infections. The dynamic relationship between the at-risk and infected groups remains the same, but the relative sizes of the two groups of susceptible people (at-risk and not-at-risk) are drastically altered and the rates of recruitment to both groups are changed.

Although the number of future infections would be small in the absence of unsafe sex, it was necessary to use a model to estimate the avoidable infections for the African subregions for two reasons. First, one of the counterfactual scenarios involves a reduction of only 10% in unsafe sex, which means that prevalence and the number of new infections remain high. Second, the relationship between current prevalence and the number of new infections in the future is not linear, even over a five-year period.

Figure 14.9 Effect of the modifications made to the EPP model which were used to calculate the number of new HIV infections occurring under different counterfactual scenarios



Note: Birth and death rates are constant, determined by demographic and epidemiological measurement.

New infection rate varies; it = constant transmissibility factor x probability that partner is HIV+.

Probability that partner is HIV+ = infected / (not-at-risk + susceptible + infected) = prevalence.

The model components affected are shaded: those components which are reduced in size by the modifications are shaded dark grey, those which are increased in size are shaded light grey.

Countries with a low prevalence of HIV infection

The counterfactuals for other subregions were again engineered to correspond to situations in which unsafe sex was reduced by 10%, 50% and 100% (no unsafe sex). Existing data on the distribution of prevalent HIV infections by mode of transmission was applied to the projections for the countries with a low prevalence of HIV infection. Reductions in unsafe sex were assumed to result in a decreased number of new STIs that were equal in proportion to the reduction in unsafe sex.

4.2 Other sexually transmitted infections

Estimation of the relationship between unsafe sex and other STIs (chlamydia, gonorrhoea, syphilis and HPV) is subject to the same constraints as that between unsafe sex and HIV infection. Relative risks of infection with chlamydia, gonorrhoea and syphilis following certain behaviours have been estimated. However, like HIV infection, these relative risks will change as the prevalence of infection changes. This problem is compounded by an even greater lack of information for any of these STIs than for HIV. As a result, we have not attempted to quantify this relationship, and assume that for all these STIs, by definition, all current prevalent infections are attributable to unsafe sex. Therefore, the total burden of disease attributed to these STIs can be considered to arise from unsafe sex. This includes cervical cancer attributable to infection with HPV; recent work suggests that all cases of cervical cancer are attributable to infection with sexually transmitted HPV (Walboomers et al. 1999).

In order to make a reasonable estimate of the future prevalence of these STIs, it is necessary to use a mathematical projection model. In common with that for HIV, such a model would need to be fitted to existing time-series prevalence data to create a projection of the future levels of infection. Since there is no appreciable mortality as a consequence of most of these other STIs, a suitable model would be much simpler than those used for HIV. Cervical cancer due to HPV infection would necessitate a model which accounts for mortality. However, the necessary time-series prevalence data are not available for a sufficient number of countries to make this a viable approach. The methods used to calculate the number of new HIV infections that are potentially avoidable cannot therefore be used for these STIs.

STIs have been virtually eliminated from some populations in the recent past. In the early 1950s, the Chinese government initiated a programme to eradicate sexually transmitted diseases that was successful in the short term. The campaign relied on mass screening to identify and treat people with an STI and also involved the abolition of commercial sex work. The methods used might not be transferable to other cultures, but demonstrate that the problem of STIs can be confronted. It has been

suggested that the incidence of STIs only began to increase after China resumed more open relations with the rest of the world in the early 1980s (Cohen et al. 1996).

With this in mind it seems reasonable to assume that all STIs are avoidable, given appropriate changes in sexual and treatment-seeking behaviour, if these changes are accompanied by the provision of suitable services.

5. Results

5.1 Prevalence of disease outcomes in 2001

Estimates of the current prevalence of HIV and other STIs were based on reported estimates from HIV surveillance and published studies. These were compiled and used to create subregional prevalence estimates (Tables 14.10 and 14.11). The estimates for the two African subregions were based on EPP model fits to antenatal clinic surveillance data. The prevalence in the other subregions was directly based on reported prevalence according to a variety of empirical sources (U.S. Census Bureau 2001).

Table 14.10 The prevalence of HIV infection in the adult population (aged 15–49 years) by subregion, in 2001

Subregion	*HIV prevalence (%)*
AFR-D	5.05
AFR-E	11.97
AMR-A	0.60
AMR-B	0.55
AMR-D	1.93
EMR-B	0.04
EMR-D	0.35
EUR-A	0.28
EUR-B	0.03
EUR-C	0.73
SEAR-B	0.45
SEAR-D	0.63
WPR-A	0.04
WPR-B	0.15
World	1.20

Table 14.11 The prevalence of chlamydia, gonorrhoea and syphilis in the adult population (all age groups) by subregion, in 2000

	Females			*Males*		
Subregion	*Chlamydia (%)*	*Gonorrhoea (%)*	*Syphilis (%)*	*Chlamydia (%)*	*Gonorrhoea (%)*	*Syphilis (%)*
AFR-D	0.50	0.50	0.09	0.47	0.47	0.07
AFR-E	0.27	0.29	0.07	0.25	0.27	0.06
AMR-A	1.05	0.41	0.03	0.89	0.36	0.03
AMR-B	0.44	0.36	0.14	0.37	0.30	0.11
AMR-D	0.42	0.32	0.14	0.34	0.26	0.11
EMR-B	0.67	0.22	0.02	0.48	0.16	0.02
EMR-D	0.45	0.15	0.02	0.37	0.13	0.01
EUR-A	0.16	0.03	0.00	0.14	0.03	0.00
EUR-B	0.20	0.10	0.01	0.20	0.10	0.01
EUR-C	0.64	0.36	0.01	0.60	0.33	0.01
SEAR-B	1.53	0.55	0.10	1.15	0.42	0.08
SEAR-D	1.98	1.49	0.17	1.51	1.16	0.14
WPR-A	0.48	0.37	0.02	0.61	0.49	0.02
WPR-B	0.24	0.14	0.01	0.20	0.11	0.01
World	0.62	0.41	0.06	0.76	0.50	0.07

5.2 Attributable infections and disease burden

The subregional estimates of the fractions of all HIV infections that are attributable to unsafe sex are given in Table 14.12. These comprise the percentage of infections prevalent in 2001 that were reportedly acquired through sexual contact. Therefore this fraction is directly attributable to unsafe sex. The feedback between prevalence and incidence has not been taken into account in the estimates for subregions outside Africa: in many of these subregions, the attributable fraction could be considerably higher if it included all infections for which sexual transmission had occurred at any point along the chain of transmission. As described above, the fractions were by definition 100% for other STIs.

5.3 Avoidable infections

The estimates of the fraction of infections which is potentially avoidable are given in the following tables and figures. Figure 14.10 shows the proportion of new infections that may be prevented by different reductions (100%, 50%, 10%) in the level of unsafe sex relative to the number of infections which would be expected to occur if there were no change in sexual behaviour. The height of the bar shows the total proportion that could be avoided if there was no unsafe sex. The proportions within the bar show the reductions that would be seen if unsafe sex was reduced

Table 14.12 The proportion of prevalent HIV infections in adults (aged 15–49 years) that is attributable to unsafe sex, by subregion, in 2001

Subregion	*% of HIV prevalence attributable to unsafe sex*
AFR-D	>99
AFR-E	>99
AMR-A	72
AMR-B	85
AMR-D	95
EMR-B	42
EMR-D	85
EUR-A	59
EUR-B	64
EUR-C	25
SEAR-B	73
SEAR-D	78
WPR-A	94
WPR-B	52
World	90

Figure 14.10 The proportion of new HIV infections currently predicted to occur during 2002–2006 that could be prevented by different reductions in the practice of unsafe sex, by subregion

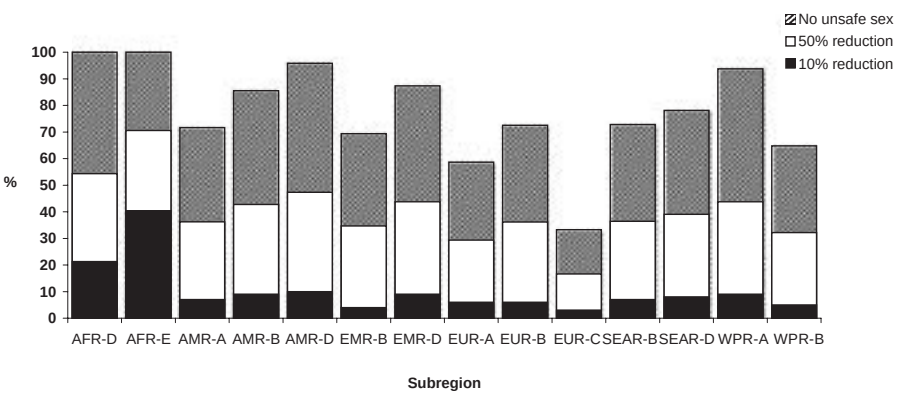


Table 14.13 The mortality and burden of disease attributable to sexually transmitted infections, cervical cancer and HIV, by subregion, in 2001

	STIs		*Cervical cancer*		*HIV*	
Subregion	*Mortality (000s)*	*DALYs (000s)*	*Mortality (000s)*	*DALYs (000s)*	*Mortality (000s)*	*DALYs (000s)*
AFR-D	43	2 224	21	283	367	11 451
AFR-E	58	2 828	37	508	1 632	50 386
AMR-A	0	73	6	93	11	350
AMR-B	1	484	19	293	29	978
AMR-D	1	73	5	74	23	684
EMR-B	0	135	3	53	0	4
EMR-D	19	1 146	8	121	45	1 366
EUR-A	0	80	8	107	4	128
EUR-B	1	150	7	112	1	28
EUR-C	0	130	12	163	4	136
SEAR-B	2	465	14	248	39	1 222
SEAR-D	57	3 891	82	1 323	268	8 204
WPR-A	0	34	3	35	0	7
WPR-B	5	582	29	377	21	839
World	188	12 296	254	3 790	2 444	75 783

by just 10% and if it was lowered by a half. These results are also given in Table 14.14.

Figure 14.11 shows how many new infections are predicted to occur in 2002–2006 in each subregion under the different counterfactual scenarios. These results are given in Table 14.15. The greatest changes would be seen in the African subregions, where sexual transmission dominates the epidemic. However in subregions such as WPR-B, which includes China, where a large number of new cases is predicted to occur, the proportion of infections that could be avoided is smaller, because use of injected drugs is a more important mode of transmission in this subregion.

It is important to consider the plausibility of the finding that almost all new HIV infections in Africa could be avoided if unsafe sex were to cease immediately despite the continuation of non-sexual transmissions. Intuitively, it seems unlikely that there would be almost no new HIV infections in the five years following the onset of behaviour change: transmission of the virus via other routes would continue, and it has been estimated that 5% of the newly-diagnosed infections in Africa in 2000 were acquired through a non-sexual mode of transmission. As discussed above, sexual and non-sexual transmission dynamics

Table 14.14 The predicted cumulative proportion of new HIV infections in adults during 2002–2006 that could be prevented by different reductions in unsafe sex, by subregion

	Reduction in unsafe sex		
Subregion	*10%*	*50%*	*100% (No unsafe sex)*
AFR-D	21	54	>99
AFR-E	40	71	>99
AMR-A	7	36	72
AMR-B	9	43	86
AMR-D	10	47	96
EMR-B	4	35	69
EMR-D	9	44	87
EUR-A	6	29	59
EUR-B	6	36	73
EUR-C	3	17	33
SEAR-B	7	36	73
SEAR-D	8	39	78
WPR-A	9	44	94
WPR-B	5	32	65

cannot be considered in isolation. Even without considering the extent to which these transmission networks are interlinked, the sheer scale of the change to the susceptible population serves to illustrate why it is not implausible that HIV transmission would cease if unsafe sex stopped altogether in the African subregions. Consider, for example, the urban areas of an east African country with a population of eight million where the estimated prevalence of HIV infection among women attending antenatal clinics in 2001 is 11%. This gives a total of 898 000 prevalent cases, of which 45 000 are thought to be non-sexually acquired. The EPP model fit to the observed prevalence data produces an estimate for the susceptible fraction of the total population of 20%. Therefore, there are 1 633 000 people who could acquire HIV infection at the start of 2001.

Using the same example, to simulate the immediate and total cessation of unsafe sex, the at-risk group was reduced by 95%, such that only 1% of the total population would be able to acquire HIV infections (5% of the original 20%), or 16 000 people. The 898 000 cases are still prevalent but not all prevalent cases are potential sources of a new infection. Some HIV-infected people will not exhibit risky behaviours and so will not have the opportunity to transmit infection. For a new case to arise, the HIV-infected people must have an effective contact (i.e. give a blood

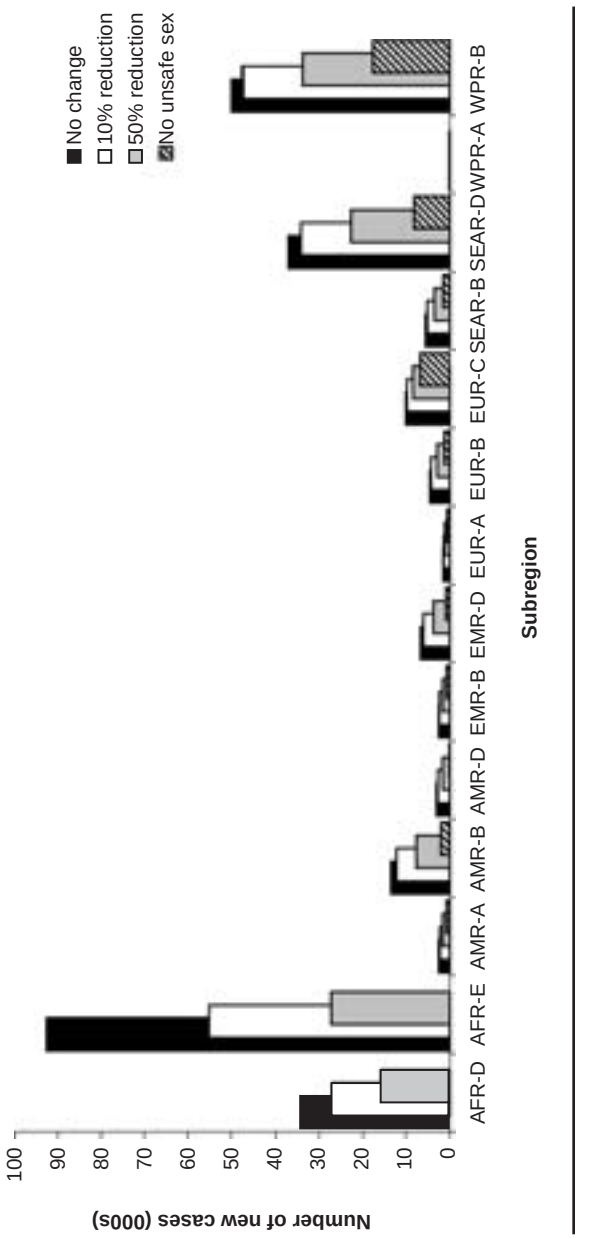


Figure 14.11 The total number of new HIV infections in adults predicted to occur during 2002–2006 assuming different reductions in unsafe sex, by subregion

Table 14.15 The total number of new HIV infections in adults predicted to occur during 2002–2006 assuming different reductions in unsafe sex, by subregion

	Reduction in unsafe sex			
Subregion	*No change*	*10%*	*50%*	*100% (No unsafe sex)*
AFR-D	3 420 598	2 691 787	1 562 368	Approx.0
AFR-E	9 250 954	5 512 072	2 724 441	Approx.0
AMR-A	240 000	223 200	153 000	68 000
AMR-B	1 350 000	1 228 500	773 000	195 750
AMR-D	300 000	270 000	158 000	12 500
EMR-B	245 000	235 200	160 000	75 000
EMR-D	665 000	605 150	374 000	84 000
EUR-A	150 000	141 000	106 000	62 000
EUR-B	451 000	423 940	288 000	124 000
EUR-C	1 008 000	977 760	840 000	673 000
SEAR-B	552 000	513 360	351 000	150 000
SEAR-D	3 720 000	3 422 400	2 268 000	815 000
WPR-A	8 000	7 280	4 500	500
WPR-B	5 000 000	4 750 000	3 390 000	1 760 000
World	26 360 552	21 001 649	13 152 309	4 019 750

transfusion or unsafe injection) with one of the 16 000 members of the at-risk group. In a population of eight million people, the probability of this happening is now much reduced, thus the number of new infections resulting is very small. The non-linearity in the relationship between changes in unsafe sex and the number of infections avoided results in a large fraction of infections averted by a 10% reduction in unsafe sex in the African subregions.

6. Uncertainty

6.1 Exposure

Data quality

Most of the behavioural surveys included in this analysis were large probability samples, which were weighted to be representative of the general population by age and sex. There may have been a selection or participation bias in these surveys. Reporting bias is probably inevitable in at least some surveys; people may have under-reported behaviours that are seen as undesirable, especially in the light of education and information campaigns aimed at promoting behavioural change. We have

limited means to assess the existence of such biases, and the assumption implicit in this work is that such biases can be ignored.

It is unclear how well quantitative household surveys measure sensitive information such as sexual behaviour and some surveys will have been designed and implemented better than others. There is little to indicate how good a survey is, apart from the quality of the data and an assessment of the questionnaire. One survey (Sri Lanka 1991 GPA survey) was excluded from the analysis because of poor quality data.

Methodological issues

In creating the set of standard behavioural indicators, different questions were used as though they were synonymous. If these questions or their translations are not in fact equivalent, the calculated indicators will not measure the same thing in all places. This is quite likely, at least with respect to the questions and indicators which depend on a classification of partner type in different countries.

The aggregation of the country-level data to subregional level is perhaps a cause for concern. For some indicators, the values estimated for countries within a subregion varied by as wide a range as was observed between the countries in different subregions. Once combined at a subregional level, this variation was no longer apparent. In addition, countries for which no data were available did not contribute to the subregional estimate; it is unlikely that the subregional estimate would not change if we did have data for the missing countries. It is plausible that, within a subregion, the countries for which no data are available are systematically different from those countries in which sexual behaviour surveys have been carried out. These differences could be related to behaviour.

Extrapolation of the estimates of the prevalence of sexual risk behaviour to subregions where there were no data was based on comparison of the proportions of the population who were currently married in each subregion. Values from the most similar subregion were substituted for the missing data. In subregions where some data were available, the missing values were taken from the subregion which was most similar according to the available estimates. However, it was clear that the subregions did not vary in a predictable manner and this method of extrapolation introduced some unquantifiable error.

Error

Confidence intervals can be calculated around the point estimates of behavioural indicators for individual countries. Almost all of these intervals are very narrow, mainly because most of the exposure data come from very large DHS. The error that was introduced by aggregating these estimates to the subregional level cannot readily be quantified: error is introduced because countries with no data are assumed to have average values for the subregion.

Omissions

Having concentrated solely on heterosexual sex, the behavioural review has clearly underestimated the amount of risk in populations where the main mode of HIV transmission is sex between men. The data on the prevalence of sex between men are too scanty to be used in an analysis of this type, and to include only the available data would introduce more uncertainty into these estimates. Infections that result from sex between men are included in the burden estimates, and in the estimates of attributable and avoidable infection, for both low- and high-prevalence subregions.

6.2 Outcomes

Model-based approach

The accuracy of the estimates of the avoidable burden of HIV infection due to unsafe sex depends initially on the precision of the five-year projections of the HIV/AIDS epidemics. These projections represent only one possible future course of the epidemic. The projections for both countries with a high prevalence of HIV infection (using the EPP model) and countries with a low prevalence (using the saturation approach) must be considered as representing a likely course, not the certain future course, of the epidemic.

Beyond the accuracy of the projections of HIV prevalence under the business-as-usual scenario, there are other sources of potential inaccuracy in the estimates of avoidable infections. The estimated proportion of all infections that are not sexually acquired is central to the calculation of the proportion of avoidable infections in all subregions. The figure of 5% employed for Africa, though widely used, should be viewed as very uncertain. Information on mode of transmission is derived from reports on the way in which people who have been diagnosed with HIV infection are thought to have acquired the infection. There are limitations to this data. In many places, a diagnosis of HIV infection will not be made before the onset of symptoms. If diagnosis is delayed it may be more difficult to identify the source of infection, especially for those people who have had more than one type of exposure. Late diagnoses or failure to diagnose may introduce another bias because the people who receive a timely diagnosis may have acquired their infection in a different manner from those whose infections are not promptly diagnosed. Subregional data are based on national data that have been aggregated to the subregional level. The different national data may be subject to different biases. Countries for which no data are available have been assumed to have the average proportion of HIV infections for the subregion. This may have distorted the picture still further. The direction of this error may be influenced by the scale and stage of the epidemic, the health care system and the equity of access to health care.

Africa

The EPP model is based on the assumption that sexual mixing patterns are homogeneous in a population. Therefore, the assumption implicit in the estimates of the numbers of avoidable HIV infections is that the reduction in prevalence of hazardous sexual behaviour is evenly distributed among the population. If declines in the prevalence of hazardous sexual behaviour are concentrated in certain groups, and the remaining risk behaviours (unsafe medical injections, unsafe blood transfusions and injected drug use) are also clustered, then the number of avoidable infections might be lower. If the remaining risk behaviours are evenly distributed throughout the population, the reductions in unsafe sex will have an effect on non-sexual modes of transmission. Infections acquired in one way are not necessarily transmitted in the same way (if they are passed on at all). Therefore sexually acquired cases of HIV infection may act as the source of infection for non-sexually-acquired cases. A reduction in unsafe sex that leads to fewer prevalent cases of HIV infection will therefore also lower the number of new non-sexually-acquired cases.

The assumption of random mixing must be tenable for the EPP model to perform well. This model is intended to give accurate projections of future HIV prevalence in a population with a generalized epidemic. If the modes of transmission that remain after unsafe sex is reduced were to be concentrated among certain groups, the subsequent number of new infections would be higher than that forecast using EPP. Therefore the estimate of the proportion of infections which is avoidable may be too high. However, given the current epidemic situation in the two African subregions, the assumption of random mixing, even in the absence of unsafe sex, may hold true because use of injected drugs is uncommon and unsafe medical injections and blood transfusions are less likely to be concentrated among specific groups.

6.3 Limitations

The departures from the standard relative risk methodology and the reasons for this have, for the most part, been fully discussed in the text. However, two further differences remain to be explained. The CRA framework requires that all estimates be presented separately for all age groups and for each sex. The estimates of avoidable infections under the different counterfactual scenarios should be made from 2000 until 2030. Neither has been done for unsafe sex because these extensions would greatly add to the uncertainty of the estimates.

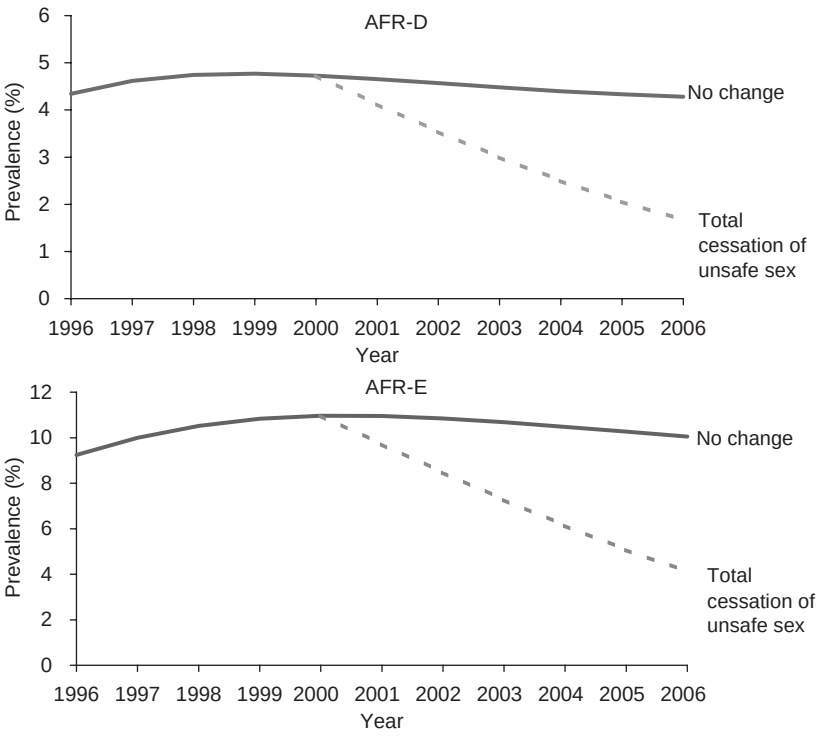
The models used for the prevalence projections are only valid in the short term. To extend them beyond 2006 would require additional assumptions about changes in the availability of treatment and prevention efforts. The effects of treatment on transmission are particularly

hard to predict since treatment will tend to increase the prevalence of infection (by prolonging the survival of infected people), but may also reduce the contagiousness of infected persons. Some of the counterfactuals considered include that there will be a massive reduction in the amount of unprotected sex after 2001. This would inevitably have an impact on fertility, which should in turn lower recruitment to the sexually active population. The methods used to predict HIV prevalence do not account for such changes. In the EPP model, as described in section 4, the recruitment of sexually active adults into the two groups of susceptible people is based on the number of births 15 years earlier and on survival rates to the age of 15 years. Therefore changes in fertility initiated in 2001 would not affect the projections until 2016. The EPP model assumes a constant birth rate that does not change over time and that is the same in the two groups of susceptible people. Although it would be possible to alter the process for implementing the counterfactual scenarios to allow for large future changes in fertility, and the emergence of a dramatic fertility differential, there is no information on which to base these estimates.

Because there is no way to estimate the size of the decline in fertility under the counterfactual scenarios in any subregion, there is no way to accurately model these scenarios beyond the short term. Similar issues are encountered in estimating HIV prevalence by age group and sex. In all subregions, prevalence is different for men and women and varies by age group. The distribution of infections by age and sex varies by epidemic duration and is not necessarily the same in all countries in a subregion, which makes it complicated to establish a subregional breakdown. Estimates of the prevalence of HIV infection among men are not available in most countries but must be inferred from prevalence in pregnant women. Changes in the number of people who are at risk of infection can be expected to change the distribution of new infections by age and sex, but the direction of these changes cannot be anticipated. Therefore, to make estimates of avoidable infections by age group and sex would be to add more uncertainty to the existing estimates.

Finally, the most extreme counterfactual scenario presented above produces some dramatic results for the African subregions. Modelling the complete cessation of unsafe sex implies that, even within marriage, discordant couples would no longer have procreative sex. Such a scenario is artificial and unprecedented: there are no historical examples of a total and sudden cessation of exposure to an infectious disease at the macro level. The reason for including this scenario is that if STIs are eliminated from a population there would be no unsafe sex: in this chapter the counterfactual has been defined in terms of the level of unsafe sex. Under the counterfactual scenario of no unsafe sex, the extraordinarily rapid decline in new infections also produces a discontinuity in prevalence, as the average duration of infection rapidly rises among those

Figure 14.12 HIV prevalence projections for subregions in Africa under two scenarios: no change in current levels of unsafe sex and the total cessation of unsafe sex

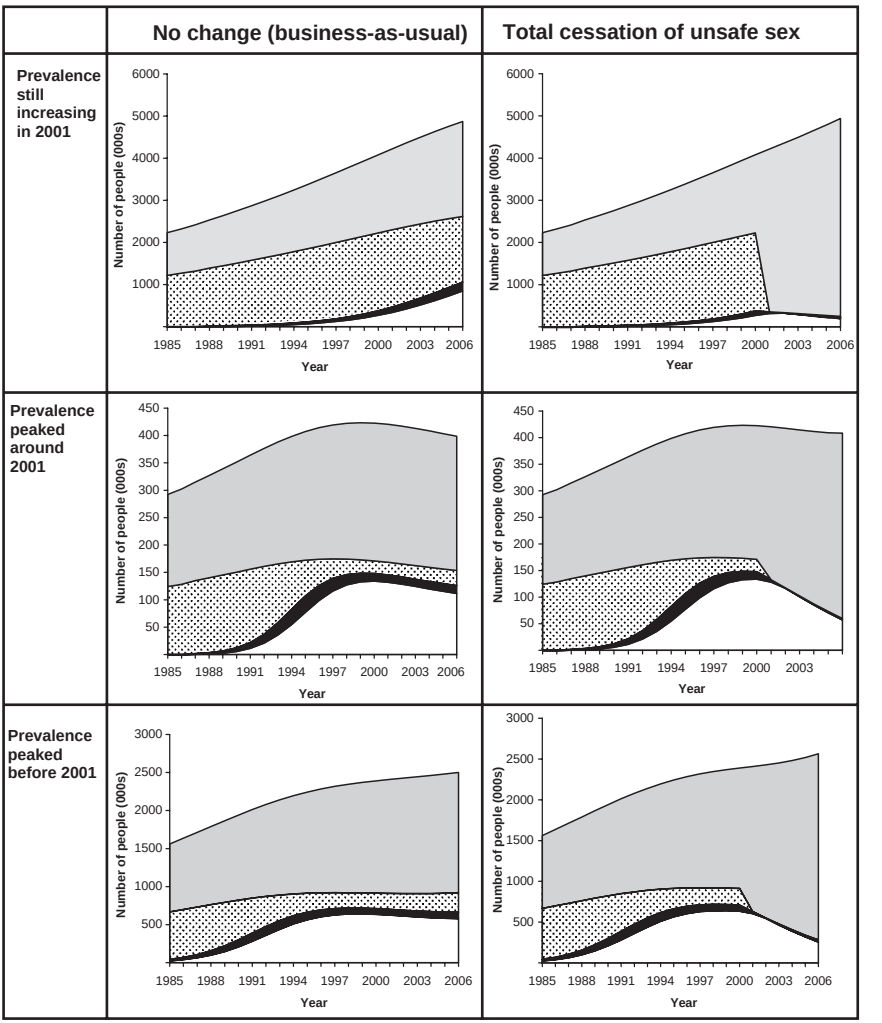


Note: The line representing the HIV prevalence following the total cessation of unsafe sex has been interpolated between 2000 and 2001 for the prevalence under the counterfactual scenario.

who are already infected, which in turn implies a rapid increase in mortality, since mortality of the people who are HIV-positive increases with the time since infection. This mortality increase exacerbates the decline in prevalence, with the results shown in Figure 14.12.

In any single EPP counterfactual scenario for a subnational population, the apparent effect of the rapid transfer of a large fraction of the susceptible population from the at-risk group to the not-at-risk group would depend on the timing of the decline in risk relative to the "natural" epidemic peak and the level of saturation (the proportion of infected people among infected and susceptible persons). Three different situations are shown in Figure 14.13, which illustrates the model fits for a population in which HIV prevalence is still rising rapidly in 2001, a second population in which growth in prevalence has stabilized in 2001;

Figure 14.13 EPP projections of the size of the infected, at-risk and not-at-risk groups for three subnational populations under two scenarios: no change in current levels of unsafe sex and total cessation of unsafe sex



Key: ▭, not at risk; ▨, at risk of infection; ■, newly infected; □, already infected.

and a third population in which HIV prevalence has begun to decline by 2001. The figure shows the relative contributions of the infected, at-risk and not-at-risk groups for two scenarios: no behaviour change (business-as-usual) and total cessation of unsafe sex.

When the data from populations at these three different epidemic stages are amalgamated at the subregional level, the business-as-usual scenario gives the impression of an epidemic with a much broader prevalence peak than that seen in any one national population. However, for the no-more-unsafe-sex scenario, because the decrease in unsafe sex is assumed to occur in the same calendar year in all places, it produces the artificial-looking declines in the number of new infections, shown in Figure 14.13.

7. Discussion and conclusions

Unsafe sex is a difficult exposure to address within the standard epidemiological framework of simple exposure measures and constant relative risks. The problem of relating behaviour patterns to risk of HIV infection is hardly a new one. Other researchers have tried to tackle this in many different ways. The Four Cities study (Buve et al. 2001b; Carael and Holmes 2001; Ferry et al. 2001) compared sexual behaviour in two African cities with a high prevalence of HIV infection and two cities with a relatively low prevalence in order to look for determinants of this heterogeneity. Individual and ecological analyses were carried out. Some behavioural factors were found to be more common in the cities with a high prevalence compared to the cities with a low prevalence of HIV infection. These were: young age at having sex for the first time (for women), young age at first marriage and the existence of a large age difference between spouses. Factors which affect transmission and which were more common in the cities with a high prevalence were herpes simplex virus (HSV-2, genital herpes) infection, trichomoniasis (for women) and lack of male circumcision. Factors that were not more common in the high-prevalence cities were: a high rate of partner change, sex with sex workers, concurrent partnerships, a large age difference between non-spousal partners, gonorrhoea, chlamydial infection, syphilis, dry sex and lack of condom use. The factors found more commonly in the cities with a high prevalence of HIV infection do not seem sufficient to explain the differences in prevalence (Buve et al. 2001a). A comparison of rural populations in Zimbabwe and the United Republic of Tanzania has also failed to find differences in sexual behaviour which could explain the higher HIV prevalence observed in the Zimbabwean population (Boerma et al. 2002). In this light, it is perhaps unsurprising that we have not been able to elucidate a relationship.

Using alternative methods for estimating the attributable disease burden, we found that most of the current burden of disease due to HIV infection is attributable to unsafe sex. If all sexual transmission were to cease, there would be just over 4 million new HIV infections between 2001 and 2006, compared to more than 26 million which are forecast to occur if there is no change in the pattern of transmission. Most of the avoidable infections are concentrated in the African subregions, which

is as expected given the current prevalence of HIV infection in these subregions. The other subregions where sexual transmission is expected to be important in the future are SEAR-D and WPR-B. These two subregions contain some countries which already report broad sexual spread of HIV infection, primarily through sex work (Cambodia, Myanmar) and even in countries where the current epidemic is now driven by injected drug use (Indonesia, China), HIV will spread more broadly from the injecting drug users to their sexual partners. In these countries, the fraction of future infections which would be averted by reductions in unsafe sex is higher than the fraction of current infections which is attributable to unsafe sex.

These findings do not come as a surprise. More important for intervention design and programme evaluation would be to identify which aspects of sexual behaviour contribute most to the spread of HIV in different settings. If this were known, the design and implementation of measures to prevent the spread of HIV infection could be improved. However, even in the absence of this information some measures are known to be effective in preventing HIV infection at the individual level. For example, increasing the levels of condom use can only help to slow the spread of infection.

Acknowledgements

The following kindly provided data:

Measure DHS+, Macro International Inc. provided the DHS for 63 countries.

Australia: Pilot data were obtained from the Australian Study of Health and Relationships. The data were made available by Anthony Smith and provided by Richard de Visser.

India: Data were obtained from the International Institute for Population Studies (IIPS), Mumbai, courtesy of Ravi K. Verma.

Europe: Data for eight countries (England, France, Germany, Greece, Italy, Norway, Portugal and Switzerland) from the European New Encounter Module (NEM) project were provided by Michel Hubert, on behalf of the NEM group.

France: Data from the 2001 Knowledge, Attitudes, Beliefs and Practices (KABP) survey (Grémy et al. 2001) were made available by Ruth Ferry and provided by Julien Chauveau.

Honduras: Data from the Centers for Disease Control and Prevention Encuesta Nacional de Salud Masculina 1996 survey were provided by Leo Morris.

Rwanda and Madagascar: Data were provided by Population Services International (PSI), courtesy of Dominique Meekers.

Models

We made use of the Epidemic Projection Package (EPP) and Spectrum. These are both available courtesy of the Futures Group.

We are greatly indebted to Tim Brown at the East West Center, not only for providing us with the source code for EPP but also for a modified version of the model, which made it possible to calculate the initial estimates of the avoidable infections in countries with generalized epidemics.

We are grateful to the UNAIDS/WHO Working Group on Global Surveillance of HIV/AIDS and STIs for making their country-specific models of HIV/AIDS available for use in this exercise.

Notes

1 See preface for an explanation of this term.

2 The exception is the Indian study which was among people attending a clinic for sexually transmitted infections. The study was included to provide some information on Asia.

References

Ahmed S, Lutalo T, Wawer M et al. (2001) HIV incidence and sexually transmitted disease prevalence associated with condom use: a population study in Rakai, Uganda. *AIDS*, **15**:2171–2179.

Anderson RM, Garnett GP (2000) Mathematical models of the transmission and control of sexually transmitted diseases. *Sexually Transmitted Diseases*, **27**:636–643.

Anderson RM, May RM (1991) *Infectious diseases of humans. Dynamics and control*. Oxford University Press, Oxford.

Anonymous (1992) Comparison of female to male and male to female transmission of HIV in 563 stable couples. European study group on heterosexual transmission of HIV. *British Medical Journal*, **304**:809–813.

Asiimwe Okiror G, Opio AA, Musinguzi J et al. (1997) Change in sexual behaviour and decline in HIV infection among young pregnant women in urban Uganda. *AIDS*, **11**:1757–1763.

Auvert B, Ballard R, Campbell C et al. (2001a) HIV infection among youth in a South African mining town is associated with herpes simplex virus-2 seropositivity and sexual behaviour. *AIDS*, **15**:885–898.

Auvert B, Buve A, Ferry B et al. (2001b) Ecological and individual level analysis of risk factors for HIV infection in four urban populations in sub-Saharan Africa with different levels of HIV infection. *AIDS*, **15**:S15–30.

Auvert B, Buve A, Lagarde E et al. (2001c) Male circumcision and HIV infection in four cities in sub-Saharan Africa. *AIDS*, **15**:S31–40.

Berkley S (1998) Unsafe sex as a risk factor. In: *Health dimensions of sex and reproduction: the global burden of sexually transmitted diseases, HIV, mater-*

nal conditions, perinatal disorders, and congenital anomalies. Global Burden of Disease and Injury, Volume 3. Murray CJL, Lopez AD, eds. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Boerma T, Nyamukapa C, Urassa M, Gregson S (2002) *The uneven spread of HIV within Africa: a comparative study of biological, behavioural and contextual factors in rural populations in Tanzania and Zimbabwe*. Population Association of America 2002 Annual Meeting. Atlanta, GA.

Buve A, Carael M, Hayes R et al. (2001a) The multicentre study on factors determining the differential spread of HIV in four African cities: summary and conclusions. *AIDS*, **15**:S127–131.

Buve A, Carael M, Hayes R et al. (2001b) Multicentre study on factors determining differences in rate of spread of HIV in sub-Saharan Africa: methods and prevalence of HIV infection. *AIDS*, **15**:S5–14.

Carael M, Holmes KK (2001) Dynamics of HIV epidemics in sub-Saharan Africa: introduction. *AIDS*, **15**:S1–4.

Chao A, Bulterys M, Musanganire F et al. (1994) Risk factors associated with prevalent HIV-1 infection among pregnant women in Rwanda. National University of Rwanda-Johns Hopkins University Aids Research Team. *International Journal of Epidemiology*, **23**:371–380.

Cohen MS, Henderson GE, Aiello P, Zheng H (1996) Successful eradication of sexually transmitted diseases in the People's Republic of China: implications for the 21st century. *Journal of Infectious Diseases*, **174**:S223–229.

Commissariat Général du Plan: Observatoire régional de santé d'Ile-de-France, Agence Nationale de Recherches sur le SIDA (2001) *Les connaissances, attitudes, croyances et comportements face au vih/sida en France. Evolutions 1992, 1994, 1998, 2001*. Paris.

de Gourville EM, Mabey D, Quigley M, Jack N, Mahabir B (1998) Risk factors for concordant HIV infection in heterosexual couples in Trinidad. *International Journal of STD and AIDS*, **9**:151–157.

del Mar Pujades Rodriguez M, Obasi A, Mosha F et al. (2002) Herpes simplex virus type 2 infection increases HIV incidence: a prospective study in rural Tanzania. *AIDS*, **16**:451–462.

Ferry B, Carael M, Buve A et al. (2001) Comparison of key parameters of sexual behaviour in four African urban populations with different levels of HIV infection. *AIDS*, **15**:S41–50.

Fylkesnes K, Musonda RM, Sichone M, Ndhlovu Z, Tembo F, Monze M (2001) Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys. *AIDS*, **15**:907–916.

Glynn JR, Carael M, Auvert B et al. (2001) Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia. *AIDS*, **15**:S51–60.

Gray RH, Kiwanuka N, Quinn TC et al. (2000) Male circumcision and HIV acquisition and transmission: cohort studies in Rakai, Uganda. Rakai Project Team. *AIDS*, **14**:2371–2381.

Gregson S, Mason PR, Garnett GP et al. (2001) A rural HIV epidemic in Zimbabwe? Findings from a population-based survey. *International Journal of STD and AIDS*, **12**:189–196.

Grémy I, al. e (2001) *Les connaissances, attitudes, croyances et comportements face au VIH-SIDA en France ; évolutions 1992, 1994, 1998 et 2001*. ORS Ile-de-France, Paris.

Hunter DJ, Maggwa BN, Mati JK, Tukei PM, Mbugua S (1994) Sexual behavior, sexually transmitted diseases, male circumcision and risk of HIV infection among women in Nairobi, Kenya. *AIDS*, **8**:93–99.

Kamali A, Carpenter LM, Whitworth JAG, Pool R, Ruberantwari A, Ojwiya A (2000) Seven-year trends in HIV-1 infection rates, and changes in sexual behaviour, among adults in rural Uganda. *AIDS*, **14**:0427–0434.

Kilmarx PH, Supawitkul S, Wankrairoj M et al. (2000) Explosive spread and effective control of human immunodeficiency virus in northernmost Thailand: the epidemic in Chiang Rai province, 1988–99. *AIDS*, **14**:2731–2740.

Konings E, Bantebya G, Carael M, Bagenda D, Mertens T (1995) Validating population surveys for the measurement of HIV/STD prevention indicators. *AIDS*, **9**:375–382.

Korenromp EL, van Vliet C, Bakker R, de Vlas SJ, Habbema JDF (2000) HIV spread and partnership reduction for different patterns of sexual behaviour—a study with the microsimulation model STDSIM. *Mathematical Population Studies,* **8**:135–173.

Malamba SS, Wagner HU, Maude G et al. (1994) Risk factors for HIV-1 infection in adults in a rural Ugandan community: a case-control study. *AIDS*, **8**:253–257.

Mastro TD, Kitayaporn D (1998) HIV type 1 transmission probabilities: estimates from epidemiological studies. *AIDS Research and Human Retroviruses*, **14**:S223–227.

Mbopi Keou FX, Gresenguet G, Mayaud P et al. (2000) Interactions between herpes simplex virus type 2 and human immunodeficiency virus type 1 infection in African women: Opportunities for intervention. *Journal of Infectious Diseases*, **182**:1090–1096.

Nelson KE, Celentano DD, Eiumtrakol S et al. (1996) Changes in sexual behavior and a decline in HIV infection among young men in Thailand. *New England Journal of Medicine*, **335**:297–303.

Nunn AJ, Kengeya Kayondo JF, Malamba SS, Seeley JA, Mulder DW (1994) Risk factors for HIV-1 infection in adults in a rural Ugandan community: a population study. *AIDS*, **8**:81–86.

Obasi A, Mosha F, Quigley M et al. (1999) Antibody to herpes simplex virus type 2 as a marker of sexual risk behavior in rural Tanzania. *Journal of Infectious Diseases*, **179**:16–24.

Quigley M, Munguti K, Grosskurth H et al. (1997) Sexual behaviour patterns and other risk factors for HIV infection in rural Tanzania: a case-control study. *AIDS*, **11**:237–248.

Quinn TC, Wawer MJ, Sewankambo NK et al. (2000) Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine*, **342**:921–929.

Rehle TM, Saidel TJ, Hassig SE, Boney PD, Gaillard EM, Sokal DC (1998) AVERT: a user-friendly model to estimate the impact of HIV/sexually transmitted disease prevention interventions on HIV transmission. *AIDS*, **12**:S27–35.

Rodrigues JJ, Mehendale SM, Shepherd ME, et al. (1995) Risk factors for HIV infection in people attending clinics for sexually transmitted diseases in India. *British Medical Journal*, **311**:283–286.

Rowley J, Berkley S (1998) Sexually transmitted diseases. In: *Health dimensions of sex and reproduction*. Murray CJL, Lopez AD, eds. Harvard School of Public Health, Boston, MA.

Royce RA, Sena A, Cates W Jr., Cohen MS (1997) Sexual transmission of HIV. *New England Journal of Medicine,* **336**:1072–1078.

Seed J, Allen S, Mertens T et al. (1995) Male circumcision, sexually transmitted disease, and risk of HIV. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **8**:83–90.

Stover J, Walker N, Garnett G et al. (2002) Can we reverse the HIV/AIDS pandemic with an expanded response? *Lancet*, **360**:73–77.

ter Meulen J, Mgaya HN, Chang Claude J et al. (1992) Risk factors for HIV infection in gynaecological inpatients in Dar es Salaam, Tanzania, 1988–1990. *East African Medical Journal*, **69**:688–692.

UNAIDS (2000) *National AIDS programmes: A guide to monitoring and evaluation*. http://www.unaids.org/publications/documents/epidemiology/surveillance/JC427-Mon&Ev-Full-E.pdf (accessed 2002).

The UNAIDS Reference Group on Estimates Modelling and Projections (2002) Improved methods and assumptions for estimation of the HIV/AIDS epidemic and its impact: recommendations of the UNAIDS reference group on estimates, modelling and projections. *AIDS*, **16**:W1–W14.

The UNAIDS Reference Group on Estimates, Models and Projections (2002) *Estimating and projecting national HIV/AIDS epidemics. The models and methodology of the UNAIDS approach to estimating and projecting national HIV/AIDS epidemics*. http://www.tfgi.com/epp_man.pdf.

UNAIDS Epidemiology Reference Group (2001) *Recommended methodology for the estimation and projection of HIV prevalence and AIDS mortality in the short-term*. http://www.epidem.org/Publications/Meeting%20summary%20and%20recommendations%20final.pdf.

UNAIDS/WHO (2001) *AIDS epidemic update*. 92-9173-132-3. Geneva. http://www.thebody.com/unaids/update1201/contents.html.

UNAIDS/WHO (2002) *AIDS epidemic update,* Geneva.

U.S. Census Bureau (2001) *HIV/AIDS surveillance data base* (accessed June 2001). http://www.census.gov/ipc/www/hivaidsd.html.

van der Ploeg CPB, van Vliet C, de Vlas SJ et al. (1998) STDSIM: A microsimulation model for decision support in STD control. *Interfaces*, **28**:84–100.

Walboomers JM, Jacobs MV, Manos MM et al. (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology*, **189**:12–19.

Walker N, Garcia Calleja JM, Heaton L et al. (2001) Epidemiological analysis of the quality of HIV sero-surveillance in the world: how well do we track the epidemic? *AIDS*, **15**:1545–1554.

Walker N, Stanecki KA, Brown T et al. (2003) Methods and procedures for estimating HIV/AIDS and its impact. The UNAIDS/WHO methods for end of 2001. *AIDS*, 17:2215–2225.

Wawer MJ, Serwadda D, Gray RN et al. (1997) Trends in HIV-1 prevalence may not reflect trends in incidence in mature epidemics: data from the Rakai population-based cohort, Uganda. *AIDS*, **11**:1023–1030.

Weiss HA, Quigley MA, Hayes RJ (2000) Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis. *AIDS*, **14**:2361–2370.

Weller S, Davis K (2002) Condom effectiveness in reducing heterosexual HIV transmission. In: *The Cochrane library issue 3*. Update Software, Oxford, http://cochrane.athens.ac.uk.

White R, Cleland J, Carael M (2000) Links between premarital sexual behaviour and extramarital intercourse: a multi-site analysis. *AIDS*, **14**:2323–2331.

WHO/UNAIDS (2000) *Guidelines for second generation HIV surveillance*. World Health Organization and the Joint United Nations Programme on HIV/AIDS, Geneva.

Appendix A

Surveys used to estimate exposure

Country	Year of survey	Age (years)		Marital status		Sample size		Type of survey	Survey organization	Source of information
		Female	Male	Female	Male	Female	Male			
AFR-D										
Benin	1996	15–49	20–64	All females	All males	5 491	1 535	DHS	DHS	Macro International
Burkina Faso	1999	15–49	15–59	All females	All males	6 445	2 641	DHS	DHS	Macro International
Cameroon	1998	15–49	15–59	All females	All males	5 501	2 562	DHS	DHS	Macro International
Chad	1997	15–49	15–59	All females	All males	7 454	2 320	DHS	DHS	Macro International
Comoros	1996	15–49	15–64	All females	All males	3 050	795	DHS	DHS	Macro International
Ghana	1998	15–49	15–59	All females	All males	4 843	1 546	DHS	DHS	Macro International
Guinea	1999	15–49	15–59	All females	All males	6 753	1 980	DHS	DHS	Macro International
Liberia	1986	15–49	—	All females	—	5 239	—	DHS	DHS	Macro International
Mali	1996	15–49	15–59	All females	All males	9 704	2 474	DHS	DHS	Macro International
Niger	1998	15–49	15–59	All females	All males	7 577	3 542	DHS	DHS	Macro International
Nigeria	1999	10–49	15–64	All females	All males	7 647	680	DHS	DHS	Macro International
Senegal	1997	15–49	≥20	All females	All males	8 593	4 306	DHS	DHS	Macro International
Togo	1998	15–49	12–59	All females	All males	8 569	3 819	DHS	DHS	Macro International
AFR-E										
Burundi	1987	15–49	≥20	All females	Husbands	3 970	542	DHS	DHS	Macro International
Central African Republic	1994	15–49	15–59	All females	All males	5 884	1 729	DHS	DHS	Macro International
Congo	1999	15–50	15–50	All females	All males	1 181	930	STIs		Supplied by PSI

continued

SURVEYS USED TO ESTIMATE EXPOSURE (*continued*)

Country	Year of survey	Age (years)		Marital status		Sample size		Type of survey	Survey organization	Source of information
		Female	Male	Female	Male	Female	Male			
Côte d'Ivoire	1994	15–49	15–59	All females	All males	8 099	2 552	DHS	DHS	Macro International
Ethiopia	2000	15–49	15–59	All females	All males	15 367	2 607	DHS	DHS	Macro International
Kenya	1998	15–49	15–54	All females	All males	7 881	3 407	DHS	DHS	Macro International
Lesotho	1989	15–55	15–56	All	All	1 033	549	KABP/PR	GPA	Supplied by ICP
Mozambique	1997	15–49	15–59	All females	All males	8 779	2 335	DHS	DHS	Macro International
Namibia	1992	15–49	—	All females	—	5 421	—	DHS	DHS	Macro International
Uganda	1995	20–44	15–59	All females	All males	1 750	1 356	In depth	DHS	Macro International
United Republic of Tanzania	1999	15–49	15–59	All females	All Men	4 029	3 542	Interim	DHS	Macro International
Zambia	1996	15–49	15–59	All females	All males	8 021	1 849	DHS	DHS	Macro International
Zimbabwe	1999	15–49	15–54	All females	All males	5 907	2 609	DHS	DHS	Macro International
AMR-A										
USA	1997	14–20	14–20	All females	All males	4 039	4 170	NLSY	NLS	NLS
USA	2000	≥15	≥15	All females	All males	—	—	Current population survey		Fields and Casper (2001)
USA	1988	18–59	18–59	All females	All males	—	—	Sexual behaviour		Laumann et al. (1995)
AMR-B										
Brazil	1996	15–49	15–59	All females	All males	12 612	2 949	DHS	DHS	Macro International

	Chile	1998	18–69	18–39	All	All	3 163	2 244	National Sexual Behaviour Survey	La Comision Nacional del SIDA (CONASIDA)	Published report
	Colombia	2000	15–49	—	All females	—	11 585	—	DHS	DHS	Macro International
	Dominican Republic	1996	15–49	15–64	All females	All males	8 422	2 279	DHS	DHS	Macro International
	El Salvador	1985	15–49	—	All females	—	5 207	—	DHS	DHS	Macro International Macro International
	Honduras	1996		15–59		All	—	2 925	RHS	CDC	Leo Morris at CDC Atlanta
	Mexico	1987	15–49	—	All females	—	9 310	—	DHS	DHS	Macro International
	Paraguay	1990	15–49	—	All females	—	5 827	—	DHS	DHS	Macro International
	Trinidad and Tobago	1987	15–49	—	All females	—	3 806	—	DHS	DHS	Macro International
AMR-D											
	Bolivia	1998	15–49	15–64	All females	All males	11 187	3 780	DHS	DHS	Macro International
	Ecuador	1987	15–49	—	All females	—	4 713	—	DHS	DHS	Macro International
	Guatemala	1999	15–49	—	All females	—	6 021	—	Interim	DHS	Macro International
	Haiti	1994	15–49	15–59	All females	All males	5 356	1 610	DHS	DHS	Macro International
	Nicaragua	1997	15–49	—	All females	—	13 634	—	DHS	DHS	Macro International
	Peru	2000	15–49	—	All females	—	32 000	—	DHS	DHS	Macro International
	Peru	1996	15–49	15–59	All females	All males	28 951	2 487	DHS	DHS	Macro International
EUR-A											
	France	1998	18–49	18–49	All	All	819	795	New Encounter Module		NEM European Group
	France	2001	18–59	18–59	All	All	1 892	1 429	KABP	ORS	ORS
	Germany	1998	15–49	15–49	All	All	1 422	1 161	New Encounter Module		NEM European Group
	Greece	1998	15–49	15–49	All	All	1 038	962	New Encounter Module		NEM European Group
	Italy	1998	15–49	15–49	All	All	1 384	1 219	New Encounter Module		NEM European Group

continued

Surveys used to estimate exposure (*continued*)

Country	Year of survey	Age (years)		Marital status		Sample size		Type of survey	Survey organization	Source of information
		Female	Male	Female	Male	Female	Male			
Norway	1997	15–49	15–49	All	All	2 122	1 582	New Encounter Module		NEM European Group
Portugal	1999	15–49	15–49	All	All	360	320	New Encounter Module		NEM European Group
Spain	1996	≥15	≥15	All females	All males	4 258	35 730	National Household Survey. sexual behaviour and condom use re HIV	Aids care—psychological and socio-medical aspects of AIDS/HIV	Castilla et al. (1998)
Switzerland	1997			All	All	1 418	1 359	New Encounter Module		NEM European Group
United Kingdom	1990	16–59	16–59	All females	All males	10 758	8 115	Sexual attitudes and lifestyles	NATSAL survey	Johnson et al. (1994)
EUR-B										
Kyrgyzstan	1997	15–49	—	All females	—	3 848	—	DHS	DHS	Macro International
Poland	1991	20–49	20–49	All females	All males	3 902	3 783	FFS	PAU	United Nations
Uzbekistan	1996	15–49	—	All females	—	4 415	—	DHS	DHS	Macro International
EUR-C										
Kazakhstan	1999	15–49	15–59	All females	All males	4 800	1 440	DHS	DHS	Macro International
Ukraine	1999	15–44	—	All females	—	7 128	—	RHS	CDC	Leo Morris at CDC
SEAR-B										
Thailand	1990	15–49	15–49	All	All	1 675	1 126	PR	GPA	

SEAR-D										
India	1993	15–60	15–60	All females	All males	969	836	Sexual behaviour in		Kumar et al. (1997)
India	1999		18–35	—	All males	—	2 087	Delhi Orissa	IIPS	IIPS, courtesy of Ravi K. Verma
WPR-A										
Australia	1999/2001	19–59	19–59	All females	All males	782	684	Sexual behaviour- data from pilot for forthcoming national study	Australian Study of Health and Relationships	Data made available by Anthony Smith, Australian Research Centre in Sex Health & Society at La Trobe University. Data kindly provided by Richard de Visser
New Zealand	1995	18–54	18–54	All females	All males	2 361 both sexes	—	National telephone survey		Paul et al. (1995)
Singapore	1989	15–49	15–49	All	All	1 109	1 006	PR	GPA	Supplied by ICP (2001) Family Health International
WPR-B										
Cambodia	2000		15–49	—	All males	—	3 166	Household BSS	FHI	
Philippines	1998	15–49	—	All females	—	13 983		DHS	DHS	Macro International

continued

Surveys used to estimate exposure (*continued*)

Key:

—	No data.
BSS	Behavioural surveillance study.
CDC	Centers for Disease Control and Prevention.
DHS	Demographic and Health Surveys.
FFS	Fertility and Family Surveys in countries of the United Nations Economic Commission for Europe (UNECE) region.
FHI	Family Health International.
GPA	Surveys carried out under the auspices of the WHO Global Programme on AIDS.
ICP	Mr Jean-Claude Deheneffe, Information Communication Partners, Brussels.
IIPS	Survey data from the International Institute for Population Studies, India.
KABP	Knowledge, attitudes, behaviours and practices (a sexual behaviour survey format).
NATSAL	National survey of Sexual Attitudes and Lifestyles, United Kingdom.
NEM group	Michel Hubert, Centre d'études sociologiques, Facultés universitaires Saint-Louis, Bruxelles, on behalf of the NEM group.
NLS	National Longitudinal Surveys (NLS) Program, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics.
NLSY	National Longitudinal Survey of Youth 1997, USA.
ORS	Observatoire Régional de Sanitaire d'Ile de France.
PAU	Population Activities Unit, UNECE.
PR	Partner relations (a sexual behaviour survey format).
PSI	Population Services International.
RHS	Reproductive Health Surveys.

References:

Castilla J, Barrio G, de la Fuente L, Belza MJ (1998) Sexual behaviour and condom use in the general population of Spain. *Aids Care*, 10:667–676.

Fields J, Casper LM (2001) *America's families and living arrangements: March 2000. Current Population Reports* (pp. 20–537), U.S. Census Bureau, Washington, DC.

Johnson A, Wadsworth J, Wellings K, Field J (1994) *The national survey of sexual attitudes and lifestyles*. Blackwell Scientific Publications, Oxford.

Kumar A, Mehra M, Badhan SK, Gulati N (1997) Heterosexual behaviour and condom usage in an urban population of Delhi, India, *Aids Care*, 9:311–318.

Laumann E, Gagnon J, Michael R, Michaels S (1995) *The social organization of sexuality. Sexual practices in the United States*. University of Chicago Press, Chicago, IL.

Paul C, Dickson N, Davis PB, Yee RL, Chetwynd J, McMillan N (1995) Heterosexual behavior and HIV risk in New Zealand. Data from a national survey. *Australian Journal of Public Health*, 19: 13–18.

Chapter 15

Non-use and use of ineffective methods of contraception

Martine Collumbien, Makeda Gerressu and John Cleland

Summary

This chapter estimates the burden of disease attributable to non-use of contraception and use of ineffective methods. The health outcomes include obstetric complications and abortion-related morbidity and mortality associated with unintended pregnancies (unwanted and mistimed). We have presented a model for linking data on contraceptive use and fertility preferences to unwanted births and unsafe abortions as intermediate outcomes, which were then related to the maternal disease burden.

The health outcomes considered were the conditions associated with unsafe abortion and unwanted births. The abortion-related conditions are a separate subcategory and the risk of abortion-related consequences is directly proportional to the risk of an unsafe abortion. The obstetric conditions linked to unwanted births are maternal haemorrhage, maternal sepsis, hypertensive disorders of pregnancy, obstructed labour and other maternal conditions. The burden of these obstetric complications attributable to non-use of contraception was assumed to be proportional to the percentage of unwanted births among all births.

Contraceptive use reduces the risk of unintended conception but does not altogether eliminate it, and failure rates are higher for traditional methods than for modern methods. The categorical variable "contraceptive use" has three levels of exposure: non-use, use of traditional methods and use of modern methods. Non-users experience the highest conception rates. The modern method category was used as the reference category for calculating the relative risk of having an abortion and an unwanted birth.

Not all conceptions lead to an avoidable burden, since many pregnancies are desired. We calculated how many unintended pregnancies are expected in one year by first estimating the proportion of women who would become pregnant and combining this with the probability that the

pregnancy would be unwanted or mistimed, based on current reproductive intentions. The proportion of women becoming pregnant was derived from contraceptive failure rates among modern and traditional method users and biological expectations of the number of conceptions among non-users. Within the non-users, conception rates were applied to the fecund women only, excluding those who would not be exposed to pregnancy for biological or behavioural reasons. Abortion probabilities were applied to determine how many of the mistimed and unwanted pregnancies would end as abortions and unwanted births. Unwanted pregnancies would contribute to both abortion-related burden and the obstetric burden of maternal complications. Mistimed pregnancies only contribute to the abortion-related burden since preventing mistimed births by use of more effective contraception does not avert—only delay—any potential associated obstetric burden.

As theoretical minimum exposure we have simulated the contraceptive distribution which would prevail if all women with a desire to either stop childbearing or postpone the next birth for at least another two years, adopt an effective modern method of contraception. All traditional method users and fecund non-users consist of women who want a birth in the next two years. At this theoretical minimum level, the relative risk of an unwanted birth and abortion becomes zero because only the reference category, modern method users, is at risk of unintended pregnancy. Counterfactual levels of relative risk were calculated to take account of the changing distributions of fertility desires within each exposure category.

Subregional[1] levels of distribution of contraceptive use and the relative risk levels of abortions and unwanted births were derived by aggregating country estimates based on data from 58 Demographic and Health Surveys (DHS). This source includes data on childbearing intentions and contraceptive use at the time of survey. Average method-specific and duration-specific failure rates were calculated from 18 countries with DHS calendar data on contraceptive use. In each country, the method–duration-specific failure rates were combined with the method mix and data on duration of use of current methods. Abortion probabilities were derived from the World Health Organization (WHO) estimates of incidence ratios (unsafe abortions per 100 live births).

It was estimated that globally 89% of the disease burden due to abortion complications is attributable to unprotected sex or use of less effective traditional methods. This amounted to 51 000 deaths and 4.4 million disability-adjusted life years (DALYs), with 82% of the burden falling on women aged <30 years. The highest absolute burden is experienced in South Asia (35% of the total abortion burden) while in relative terms women in the two African subregions are the worst effected. The burden of disease attributable to maternal conditions arising from unwanted births was 98 000 deaths and 4.5 million DALYs. In contrast to abortion, the largest part of the burden befalls women over 30 (74%) since

a higher proportion of all births is unwanted among older women. For women aged <30 years, about 7% of all births could be averted if all women who wished to stop childbearing used a modern method. This proportion is as high as 40% for the older age group.

1. Introduction

Sexual intercourse contributes positively to health and general well-being in both men and women; it leads to increased intimacy in relationships. Sexual intercourse is also an important risk factor for disease and disability. The most important negative consequence of sex is the risk of contracting a sexually transmitted infection, including HIV, through unprotected intercourse. HIV and sexually transmitted infections (STIs) are discussed in chapter 14. In this chapter we have concentrated on the reproductive consequences of sexual intercourse. Notwithstanding recent developments in assisted reproduction, sexual intercourse is a requirement for reproduction for the overwhelming majority of couples. Motherhood is highly valued in most societies, but each pregnancy and childbirth carries a health risk for the woman, and where obstetric services are poor, maternal mortality is still very high. The most recent estimates show that, of the global 515 000 maternal deaths in 1995, more than 99% occurred in Africa, Asia, Latin America and the Caribbean (Hill et al. 2001).

Reduction of maternal mortality and morbidity can be achieved by more effective treatment of pregnancy-related complications. The disease burden can also be reduced by avoiding pregnancies through adoption of effective contraception. Not all pregnancies and births are intended: many are either mistimed or unwanted at any time. Worldwide it is estimated that about 210 million recognizable pregnancies occur every year (The Alan Guttmacher Institute 1999)—of which about 15% end in spontaneous miscarriage or stillbirth. Another 22% are terminated by induced abortion and thus can be classified unambiguously as unintended. The remainder—some 133 million—result in the birth of a baby. Evidence from DHS and similar surveys has suggested that, globally, some 20% of all births are unintended (The Alan Guttmacher Institute 1999). Adding together unintended births and induced abortions, it may be concluded that about 40% of all pregnancies are unintended.

This chapter is concerned with estimating the burden of maternal complications and abortions that could be avoided if couples increased their use of effective contraception. We have presented a model for linking data on contraceptive use and fertility preferences to unwanted births and unsafe abortions. DHS data for 58 countries were used to calculate attributable fractions: what proportion of these unwanted births and unsafe abortions could be averted by perfect implementation of fertility preference through increased use of effective contraception. These

intermediate outcomes were then linked to estimates of the burden of maternal complications in pregnancy.

2. Risk factor definition and health outcomes

2.1 Health outcomes

The health outcomes considered for assessing disease burden due to lack of use of effective contraception are the conditions associated with unsafe abortion and unwanted births. The abortion-related conditions are a subcategory under the Global Burden of Disease (GBD) study's causes of maternal conditions. The main causes of mortality and morbidity associated with unsafe abortion are sepsis, following incomplete removal of the fetus, and perforation of the uterus. Based on the International Statistical Classification of Diseases and Related Health Problems (ICD-10), the obstetric conditions other than the abortion-related include maternal haemorrhage, maternal sepsis and obstructed labour. Other complications include hypertensive disorders of pregnancy, and the category of "other maternal conditions". These maternal complications were the causes considered in attributing the burden of disease to unwanted births.

Some other conditions are exacerbated by pregnancy. Indirect obstetric complications result from existing disease (malaria, anaemia, hepatitis, cardiovascular disease, tuberculosis and hypertension) but are aggravated by the physiological effects of pregnancies (AbouZahr and Vaughan 2000). Suicide and violence may be pregnancy related, and other forms of psychological morbidity are associated with childbirth and unintended pregnancies. Since the magnitude and the strength of these relationships are largely unknown, none of these conditions have been included in the burden attributable to non-use of contraception.

The morbidity related to use of contraception has been excluded for this exercise. Those conditions include allergic reactions to barrier methods, intrauterine device (IUD)-associated bleeding, and wounds from surgical procedures. Morbidity associated with systemic contraceptive such as the oral contraceptive pill, includes the impact on cardiovascular and hormonal systems and carcinogenicity (AbouZahr and Vaughan 2000).

The burden of obstetric complications attributable to non-use of (or use of less effective) contraceptive methods is proportional to the percentage of all births that are unwanted. Intergenerational effects of contraceptive use on the health of offspring have not been considered, nor has the burden of perinatal outcomes associated with the delivery of unwanted births been taken into account. It is clear that by averting unwanted pregnancies, a proportion of perinatal deaths can be avoided. However, by averting unwanted births the disease burden throughout

infancy and beyond can be reduced. The potential contribution of contraception to infant survival through better birth spacing is also well known. Babies born within 24 months of an elder sibling are at elevated risk of dying in infancy (Trussell and Pebley 1984). Mistimed births may therefore be associated with higher disease burden in childhood. Because of the conceptual problems of considering health impact in the next generation, the outcomes have been restricted to maternal ones.

2.2 Intermediate outcomes: unwanted births and unsafe abortions

In the "perfect contracepting society", all women, or couples, who do not wish to have a baby within the next year or so would use effective contraception. Under these circumstances, a small residue of unintended pregnancies would remain because of contraceptive failure but the overwhelming majority of births would be intended. In the real world, however, large discrepancies exist between reproductive wishes and contraception protection. These discrepancies arise for myriad reasons. In developing countries the main direct cause is lack of any contraceptive precautions despite the desire to delay the next child or have no more children. In the demographic literature, non-use of contraception among women desiring to space or limit childbearing is termed "unmet need" for contraception. Estimates of the prevalence of such unmet need in 55 developing countries in the 1990s ranged from 6% to 40% of all currently married women (Westoff 2000). The main underlying causes of unmet need in developing countries include a perception that risk of pregnancy is low, opposition to the use of contraception, stemming from the husband's attitude or religious considerations, and concerns about the safety or side-effects of methods. Lack of knowledge about contraceptive methods, or how to access them, are also important contributory causes in some countries.

In industrialized countries, contraceptive practice tends to be higher than in most developing countries. Nevertheless appreciable discrepancies between reproductive motivation and behaviour are also apparent. Contraceptive failure and irregular use of methods are more important direct causes of unintended pregnancies than in developing countries. For instance, in the United States of America about half of all unintended pregnancies are the result of failure or irregular use (Henshaw 1998). Unanticipated sexual intercourse no doubt represents a further risk factor, particularly for single women.

Attitudes towards becoming pregnant are complex and often ambivalent. Typically, two persons are involved, the woman and her husband or partner, whose views do not necessarily coincide. Attitudes may also change over time, particularly between the time before conception and the time following recognition of the pregnancy. For instance, a couple may have no fixed intention to have a baby but nevertheless be delighted when conception occurs. No unambiguous and generally agreed defini-

tion of unintended pregnancy exists. Rather researchers have used a variety of indirect and direct methods of measurement.

The most commonly available and used measure is that employed by the DHS. In this approach women are asked the following question about recent live births and the current pregnancy (if any): "At the time you became pregnant with (NAME OF CHILD) did you want to become pregnant then, did you want to wait until later, or did you want no more children at all?" This question leads to a three-way classification: births wanted at that time; birth not wanted then but later; birth not wanted at any future time. The latter two categories—mistimed and unwanted births—are often grouped together and defined as unintended births or pregnancies. Some authors use the terms unintended and unplanned interchangeably. However the concept of planning implies active preparation for pregnancy (e.g. cessation of contraceptive use, possible dietary changes, etc.) that makes it inappropriate for the large number of countries where contraception is still uncommon.

Unintended pregnancies may be subdivided into those that are unwanted at any time and those that are mistimed. Both categories may lead to induced abortion although in most settings it can be expected that unwanted pregnancies are more likely to be terminated than mistimed ones (Bankole et al. 1999). In 1995, it was estimated that approximately 26 million legal and 20 million illegal abortions occurred worldwide (Henshaw et al. 1999). The legality and safety of abortion are strongly correlated (Rahman et al. 1998). In the developed world where abortion is generally legal, abortion mortality is as low as 0.2 to 1.2 deaths per 100 000 procedures. In non-legal settings, when an unskilled provider, using hazardous techniques, terminates the pregnancy—often in unsanitary conditions—complications for the woman are likely. Of the 20 million illegal abortions globally, 19 million happen in developing countries (Henshaw et al. 1999). Where abortion is either illegal or highly restrictive, abortion mortality averages 390 deaths per 100 000 procedures (but as high as 680 in Africa) (WHO 1998). WHO defines an "unsafe abortion" as a procedure for terminating an unintended pregnancy either by persons lacking the necessary skills or in an environment lacking the minimal medical standards or both (WHO 1998). About a third of unsafe abortions lead to serious complications, and about 13% of the pregnancy-related deaths worldwide are related to complications of unsafe abortion (The Alan Guttmacher Institute 1999).

When unintended pregnancies are not aborted, and no miscarriage or stillbirth occurs, the pregnancy results in a live birth. As mentioned before, in developing countries, the mortality and morbidity risk associated with complications during childbirth is substantial for any birth, whether intended or unintended.

Part of the total burden of obstetric complications during childbirth can, however, be avoided by preventing the unwanted pregnancies (i.e.

those not wanted at any time) through use of effective contraception (Fortney 1987; Winikoff and Sullivan 1987). Avoiding unwanted pregnancies will reduce maternal mortality in two ways: by reducing the number of pregnancies and by reducing obstetric risk (i.e. the risk per pregnancy). Unwanted births tend to occur when women are relatively old and already have several children. Risks to the mother's health of pregnancy and childbirth are higher at older ages. Hence, the obstetric risk as measured by the maternal mortality ratio (maternal deaths per 100 000 live births), is reduced by averting high-risk births based on maternal age and parity but the effect is relatively small (Trussell and Pebley 1984; Winikoff and Sullivan 1987).

In many countries unwanted births are particularly likely to occur to women who have low education, poor nutrition and poor access to health services—all conditions associated with a higher maternal complication rate (Berkley 1998). However, this link between socioeconomic conditions and unwanted births is not universal. In countries with low levels of contraceptive practice, as in much of sub-Saharan Africa, educated women are as likely to report unwanted births as uneducated women (Adetunji 1998). However, these effects of obstetric risk (averting high-risk births), are dwarfed by the impact of reducing the overall incidence of pregnancies through the elimination of unwanted births (Fortney 1987). This elimination will have a huge impact on the maternal mortality rate (maternal deaths per 100 000 women of reproductive age), and the lifetime risk of dying in childbirth or pregnancy.

So is it reasonable to assume that delivery complications associated with unwanted births are the same as those associated with wanted births? Apart from considerations of maternal age and socioeconomic status, another possibility is that mothers neglect unintended pregnancies in ways that put the mother herself at greater risk. The evidence is meagre but a recent analysis using five DHS concluded that unintended pregnancies are not selectively discriminated against in terms of obstetric care and thus probably did not represent excess risk to mother's health and survival (Marston and Cleland 2003a). As these five surveys include enquiries from Africa, Asia and Latin America, it is reasonable to generalize results, at least to developing regions. On balance, therefore, it is justifiable to assume that obstetric complications are the same for wanted and unwanted births.

How should *mistimed* (in distinction to unwanted) births be regarded in relation to delivery complications? Births may be classified as mistimed when the woman is too young and wants to delay the first birth, or when she feels births are too closely spaced or when other conditions are not yet conducive to childbearing. Births to very young women (aged <18 years) do carry a higher risk, and delaying some of those may therefore avert some obstetric risk, although the effect will be small (Trussell and Pebley 1984). However, reducing mistimed births by contraceptive practice will have little influence on the incidence of pregnancies as the

births will merely be delayed rather than averted. Such delay or postponement will thus not reduce the burden of delivery complications. As discussed earlier, the potential contribution of contraception to infant survival through better birth spacing is well known, with babies born within 24 months of an elder sibling being at elevated risk of dying in infancy (Trussell and Pebley 1984). In contrast to the strong evidence regarding childhood risks, it is uncertain whether shorter birth intervals are associated with an increased risk of maternal mortality or morbidity. The only two published studies give conflicting results (Conde et al. 2000; Ronsmans and Campbell 1998). It is therefore not justified to regard short intervals as a risk factor for obstetric complications. It may be concluded, therefore, that prevention of mistimed births through contraceptive use will make no contribution to the reduction of delivery complications.

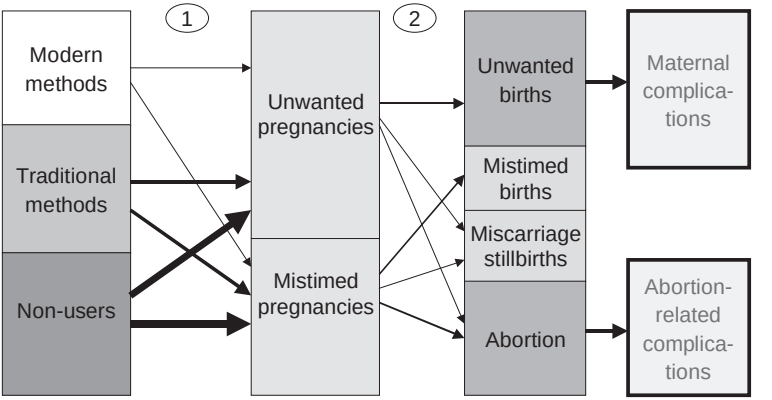
2.3 The pathway from exposure to health outcomes

When assessing the 1990 disease burden attributable to unsafe sex, Berkley (1998) estimated the percentage of women with an unmet need for family planning and attributed an equivalent proportion of the obstetric and abortion burden to non-use or inappropriate use of contraception (Berkley 1998). In order to follow the comparative risk assessment (CRA) methodology, we have modelled the outcomes (intermediate outcome in terms of unwanted births and unsafe abortion and health outcome in terms of obstetric and abortion-related burden) from exposure (i.e. contraceptive behaviour).

The definition of exposure has to take into account the fact that contraceptive use reduces the risk of conception but does not altogether eliminate it. The probability of accidental pregnancy while using a method depends on the intrinsic or theoretical effectiveness of the method itself (method failure) and on whether it is used consistently and correctly (user failure). Some methods (e.g. condoms, oral contraceptives, withdrawal and periodic abstinence) are much more prone to user error than other methods (e.g. contraceptive sterilization, intrauterine devices). Withdrawal, or coitus interruptus, and periodic abstinence are distinguished from all other commonly used methods by their exceptionally high failure rates. In the family planning literature these two methods are often given the label "traditional" because they are not the product of advanced techniques of biochemistry or engineering. To capture this variability by method, exposure was divided into three levels: non-use, use of traditional methods and use of modern methods. The pathway from exposure to intermediate and health outcome is depicted in Figure 15.1. The diagram relates to sexually active women (non-virgins), and for clarity the categories of wanted pregnancies and births are omitted.

All three levels of exposure will lead to both unwanted and mistimed pregnancies, but the probability is lowest for the modern method users.

Figure 15.1 Pathway from exposure to outcome



Failure rates for traditional methods are higher and no protection carries the highest risk. Modern methods have been used as the theoretical-minimum-risk reference category for calculating the relative risks of having an unintended pregnancy (subdivided into unwanted and mistimed) among traditional method users and non-users.

The next step in the model was to examine the reproductive outcomes of these unintended pregnancies: spontaneous fetal loss and stillbirths, abortions, mistimed and unwanted births. Both unwanted and mistimed pregnancies may end in miscarriage or stillbirth, and such events may cause obstetric complications. No evidence exists to suggest that the probability of miscarriage or stillbirth for unintended pregnancies differs from that of intended pregnancies (i.e. the ratio of unwanted over all stillbirths is the same as the ratio of unwanted births over all births). Therefore, these events have been excluded from the calculations of the attributable risk. Since the total burden of maternal complications is a separate input provided by WHO (independent from our model), and complications due to in utero loss are an integral part of this burden, the attributable burden of maternal complications associated with in utero loss has been accounted for. Both unwanted and mistimed pregnancies may be aborted. In terms of ultimate outcome or disease burden, the risk of abortion-related consequences is directly proportional to the risk of an unsafe abortion. Unwanted pregnancies may be carried to term and the burden of maternal complications will be proportional to the percentage of unwanted births among all births. As discussed earlier, mistimed pregnancies carried to term do not contribute to an attributable burden of disease, because this proportion of disease burden would only have been delayed if the pregnancy were not mistimed.

The model in Figure 15.1 was applied to data from 58 DHS to obtain subregional levels of exposure and the relative risk of having unwanted births and unsafe abortions.

3. Data and methods for exposure and hazard

3.1 Data sources

In the analysis, we drew heavily on DHS data on childbearing intentions and contraceptive use at the time of survey. DHS were used for several reasons. First, they are the dominant source of information on fertility intentions and contraceptive use in developing countries, with good representation in all highly populated regions. Most of the most populous developing countries have conducted a recent Demographic and Health Survey: Bangladesh, India, Indonesia and Pakistan in Asia; Brazil and Mexico in Latin America; and Ethiopia and Nigeria in sub-Saharan Africa. Indeed China is the only conspicuous absentee but, as will be shown later, this omission is relatively unimportant because the attributable burden is small in this country.

A further reason for reliance on DHS is that all surveys are nationally representative and executed to a high standard, with abundant technical assistance where needed. Surveys are also highly standardized in content, with the important implication that measures of fertility intentions and contraceptive use are comparable across countries. A final pragmatic reason for using DHS is that well-documented, clean data files are available for public use. Other data, especially the reproductive health surveys by the Centers for Disease Control and Prevention (CDC) were considered, since they cover countries mainly in Latin America and eastern Europe. Because of restricted availability, different age ranges and inadequate detail to make analytically important distinctions, we chose to use DHS data only. Similarly, the use of the National Survey of Family Growth (United States) was briefly considered, but the very small overall burden of maternal complication and the different questions used did not warrant the considerable effort that would have been required.

The surveys used in these calculations were those available at the time of calculation in 2001. Table 15.1 lists the countries by subregion, giving the date of survey, whether the sample was restricted to ever-married women (those who are or have been married) or all women, and the legal status of abortion in each country. The table also indicates what proportion of the total female population aged 15–44 years in the subregion is represented by the country and the weights used for aggregating country-specific data into subregional estimates. Ten of the 58 surveys were done before 1990. Although fertility levels and preferences may have changed considerably in the last decade, these surveys were retained

Table 15.1 DHS used in estimating exposure and relative risks

Subregion	*Country*	*Year*	*Sample (All or ever-married)*	*Legal status of abortion*[a]	*Proportion of subregional total of 15–44*[b] *(%)*	*Weight in subregional estimate*[b] *(%)*
AFR-D	Burkina Faso	1999	All	II	3.9	4.9
	Benin	1996	All	I	2.1	2.6
	Cameroon	1998	All	II	5.0	6.3
	Ghana	1998	All	III	6.9	8.6
	Guinea	1999	All	II	2.5	3.1
	Comoros	1996	All	II	0.2	0.3
	Liberia	1986	All	III	1.1	1.4
	Madagascar	1997	All	I	5.3	6.7
	Mali	1996	All	I	3.6	4.5
	Nigeria	1990	All	I	38.4	48.2
	Niger	1998	All	I	3.4	4.3
	Senegal	1997	All	I	3.2	4.0
	Chad	1997	All	I	2.5	3.1
	Togo	1998	All	I	1.5	1.9
AFR-E	Botswana	1988	All	III	0.5	0.7
	Burundi	1987	All	II	2.0	2.9
	Central African Republic	1994	All	I	1.1	1.6
	Côte d'Ivoire	1994	All	I	4.4	6.3
	Ethiopia	2000	All	II	17.5	25.2
	Kenya	1998	All	I	9.4	13.6

continued

Table 15.1 DHS used in estimating exposure and relative risks (*continued*)

Subregion	*Country*	*Year*	*Sample (All or ever-married)*	*Legal status of abortion*[a]	*Proportion of subregional total of 15–44*[b] *(%)*	*Weight in subregional estimate*[b] *(%)*
	Malawi	1992	All	I	3.1	4.5
	Mozambique	1997	All	II	5.7	8.2
	Namibia	1992	All	III	0.5	0.7
	Rwanda	1992	All	II	2.4	3.4
	United Republic of Tanzania	1996	All	I	10.0	14.4
	Uganda	1995	All	I	6.2	8.9
	Zambia	1996	All	IV	2.8	4.0
	Zimbabwe	1994	All	II	3.8	5.5
AMR-B	Brazil	1996	All	I	40.9	51.8
	Colombia	1995	All	I	9.9	12.5
	Dominican Republic	1996	All	I	1.9	2.4
	El Salvador	1985	All	I	1.4	1.8
	Mexico	1987	All	I	23.3	29.5
	Paraguay	1990	All	I	1.2	1.5
	Trinidad and Tobago	1987	All	III	0.3	0.4
	Bolivia	1998	All	II	11.2	11.2
AMR-D	Ecuador	1987	All	II	18.3	18.3
	Guatemala	1995	All	I	14.7	14.7
	Haiti	1994	All	I	11.3	11.3

	Nicaragua	1997	All	I	6.9	6.9
	Peru	1996	All	II	37.7	37.7
EMR-B	Tunisia	1988	EM	V	7.5	100.0
EMR-D	Egypt	1995	EM	I	20.0	23.7
	Morocco	1992	All	II	9.1	10.7
	Pakistan	1990	EM	II	42.4	50.0
	Sudan	1989	EM	I	8.6	10.1
	Yemen	1997	EM	I	4.6	5.5
EUR-B	Kyrgyzstan	1997	All	V	2.2	4.7
	Turkey	1998	EM	V	32.7	70.8
	Uzbekistan	1996	All	V	11.3	24.5
EUR-C	Kazakhstan	1995	All	V	6.9	100.0
SEAR-B	Indonesia	1997	EM	I	71.8	71.8
	Sri Lanka	1987	EM	I	6.4	6.4
	Thailand	1987	EM	II	21.8	21.8
SEAR-D	Bangladesh	1997	EM	I	11.1	11.9
	India	1993	EM	IV	80.4	86.1
	Nepal	1996	EM	I	1.8	2.0
WPR-B	Philippines	1998	All	I	4.8	100.0

[a] Legal status of abortion: I, permitted only to save the woman's life or prohibited altogether; II, physical health (also to save the woman's life); III, mental health (also to save the woman's life and physical health); IV, socioeconomic grounds (also to save the woman's life, physical health and mental health); V, without restriction as to reason (Source: Center for Reproductive Rights [formerly Center for Reproductive Law and Policy] 1999).

[b] These weights represent the fraction of the total population (of women aged 15–44 years) of countries with data.

in the analysis since the differences in fertility levels within subregions are even wider.

DHS directly provide most of the information needed for the calculation of attributable risk ratios. This includes information on sexual activity (needed to define exposure), fertility intentions, type of contraceptive method used and probability of contraceptive failure. These are discussed below. However, DHS have one important defect: most do not collect information on induced abortion and those that do yield severe underestimates. The difficulty of obtaining reliable information on induced abortion is the most intractable problem in the study of human reproduction, perhaps not surprisingly in view of the fact that abortion is both illegal and stigmatized in many societies.

In this chapter, we have used unpublished 1995 national estimates of unsafe abortions compiled by WHO. These estimates were made indirectly from data on hospital admissions for abortion complications, weighted by the proportion of abortions that are thought to result in complications requiring admission. Information from community surveys and to a lesser extent from abortion providers' surveys and mortality studies have also been used to derive best possible estimates (WHO 1998). While this is the main base of data used to calculate abortion probabilities, DHS data have also been used for three central Asian republics (Kazakhstan, Kyrgyzstan and Uzbekistan) where abortion is legal and survey estimates are considered reliable (Westoff et al. 1998). In addition, data on legal rates of abortion were used to make adjustments in some countries (Henshaw et al. 1999).

The only other non-DHS data source used in the calculations were biological in nature. Monthly probabilities of conception among non-users of contraception (i.e. fecundability) and intrauterine mortality were taken from the published literature (Bongaarts and Potter 1983; Leridon 1977).

3.2 Defining exposure and fertility preferences

Women who do not have sexual intercourse are obviously not at risk of complications of abortion or childbirth. Virgins and others need to be distinguished. Virgins are excluded from exposure and they do not affect the relative risk calculations. The proportions of virgins have been given as a separate input for each subregion. In 12 surveys (Table 15.1), mainly in Asia where premarital sex is relatively uncommon, only ever-married women were interviewed. For these countries never-married women have been categorized as virgins. All non-virgins were included in the appropriate category of exposure variable "contraceptive status". A large proportion of non-users is not exposed to risk of pregnancy for either biological or behavioural reasons. As the calculations involved estimating births over a 12-month period, it was decided to classify women who reported no intercourse in the past 12 months as behaviourally unexposed.

Fertility surveys like the DHS include several questions on fertility intentions: total desired family size, whether more children are wanted, the number of additional children wanted and the intended status of recent births and current pregnancy. Retrospective data on recent births could not be used for our purposes of examining the relationship between unwanted births and contraceptive use, since the women were not asked whether they were using a method of contraception at the time of conception. Instead, we needed to use a forward-looking measure on desirability and timing of any future births. Women's response to the questions "Would you like to have a/another child or would you prefer not to have any (more) children?" and "How long would you like to wait from now before the birth of a/another child?" are considered to be relatively unbiased. Women have no reason to misreport their preference for more children (Bongaarts 1990). Moreover, these future childbearing wishes or intentions are predictive of subsequent childbearing (Westoff 1990).

The data used to link contraception with childbearing intentions are the same that provide the input for calculating the now ubiquitous measure of unmet need for family planning, routinely reported from fertility surveys (Dixon-Mueller et al. 1992; Robey et al. 1996). A woman has an unmet need for birth spacing when she wants to postpone the next birth for at least two years, she is not using contraception and is exposed to risk of conception (i.e. sexually active and menstruating). Similarly, women who want no more children and are not current users are defined to have an unmet need for limiting family size. Unmet need refers to the current status (at the time of interview), but we need to assume a steady state for one year in order to project the fertility implications over the next year for each combined level of exposure and fertility preference.

Since we wanted to estimate the yearly number of expected pregnancies, was this assumption that childbearing intentions stay constant for one year valid? Unless the women are re-interviewed the next year there is no way of knowing the exact fertility implications of the stated preferences and whether these preferences remain constant over a one-year span. Over long periods, childbearing intentions can change substantially, especially in countries progressing through a secular decline in fertility (Freedman et al. 1980). A woman's economic and social circumstances, health status and current marriage/partnership will influence her response about childbearing intentions. As these individual personal circumstances change so may her desire for more children. In the shorter term, however, stability of intentions is reasonably high. For instance, in a prospective study in Peru, aggregate levels of fertility preferences were shown to be consistent over a three-year period (Mensch et al. 1995). Even if circumstances and preferences change for individual women, at the aggregate level the changes are likely to be offset by other women whose life circumstances may change in the opposite direction.

3.3 Estimating the expected number of unintended pregnancies

Referring back to Figure 15.1, the first stage was to estimate how many unintended pregnancies we expected in one year. This was done by first estimating the proportion of women who would become pregnant and combining this with the probability that the pregnancy was unwanted or mistimed. The proportion of women becoming pregnant was based on contraceptive failures among modern and traditional method users and on biological expectations of the number of conceptions among non-users.

Contraceptive failure

Most enquiries of DHS do not collect information that permit the calculation of contraceptive failure rates. However in a subset of 18 developing countries where levels of contraceptive practice are high, the necessary data have been collected in the form of month-by-month calendars of contraceptive use spanning a 60-month period prior to date of interview. The type of methods used, dates of starting and ending episodes of use together with main reason for stopping (including failure) are ascertained. Failure rates can be calculated by application of life table techniques to these data. Though this retrospective method of measurement makes heavy demands on the memory of respondents, recall is aided by prior entry into the calendar of live births, ascertained earlier in the interview, and the contraceptive data appear to be of high quality (Curtis and Blanc 1997).

We used an unpublished analysis of failure rates for all 18 countries where calendar data have been collected. Despite considerable inter-country variability, the ranking of methods according to failure rates is clear-cut and accords with other evidence (Trussell 1998). Failure rates are low for methods requiring no memory or skill from users (sterilization, IUD, implant, injectable), intermediate for theoretically effective methods that do require inputs from users (oral contraceptives, condoms) and high for periodic abstinence and withdrawal.

Failure rates not only vary by method and by type of user but also by duration of use. They tend to be higher during the initial period of use and subsequently decline. The reason for this trend concerns selectivity. Inefficient users of a method have a high probability of an early failure. With the passage of time, continuing users are increasingly selected for their proficiency of use and thus failure rates fall. This tendency is more marked for methods requiring skill or memory than for other methods. We have used the mean method-specific and duration-specific failure rates for these 18 countries to calculate country-specific aggregate failure rates for all 58 countries. In the left-hand panel of Table 15.2 we have presented the yearly probabilities of experiencing failure by duration of use for each method. These were calculated from the single-decrement

Table 15.2 Calculation of average method-specific failure rates from yearly probability of failure by duration of use for women aged 30–44 years, all countries combined

Contraception method	*Yearly probability of experiencing contraceptive failure by duration of use (%)*				*Duration of use in completed years (proportional distribution)*				*Average failure rate (%)*
Year of use	1st	2nd	3rd	4+	0	1st	2nd	3rd	
Pill	5.4	6.1	4.6	4.0	0.43	0.17	0.10	0.30	5.0
IUD	1.4	1.7	1.3	1.4	0.26	0.16	0.14	0.45	1.4
Injections	2.4	2.4	2.1	1.7	0.57	0.17	0.08	0.17	2.2
Diaphragm/foam	20.8	17.9	7.7	5.0	0.44	0.09	0.10	0.37	13.3
Condom	9.6	8.3	7.1	2.7	0.45	0.15	0.10	0.30	7.1
Norplant	0.1	0.2	0.1	0.0	0.37	0.23	0.11	0.30	0.1
Female sterilization	0.1	0.0	0.0	0.2	0.12	0.11	0.10	0.67	0.1
Male sterilization	0.9	0.7	1.8	0.0	0.09	0.10	0.10	0.70	0.3
Periodic abstinence	21.0	19.5	12.7	10.1	0.34	0.09	0.14	0.43	14.9
Withdrawal	17.4	18.2	15.8	10.7	0.41	0.16	0.13	0.30	15.3
Other traditional	16.0	24.0	15.5	8.5	0.37	0.20	0.10	0.33	15.1

life table estimates of failure. The estimates should be interpreted as the cumulated percentage of couples who would experience failure by the end of the 12th month of use, in the absence of other reasons for stopping use. As can be seen, failure rates generally decline by duration of use, although the second-year failure rates are higher than the first year ones for pill, IUD, withdrawal and "other traditional" methods.

For each of the 58 countries, these method–duration-specific failure rates were then combined with data on duration of use of current methods. The resulting country-specific failure rates for each method were calculated separately for the two age groups. Table 15.2 shows the calculation of these method-specific failure rates for women aged 30–44 years (all countries combined).

The final step was to obtain aggregate failure rates for modern and traditional methods by taking into account the relative contribution of the different modern and traditional methods (method-mix) in each country. The country–method-specific failure rates (calculated as in Table 15.2) have been combined with the relative method mix, as shown in Table 15.3. For women aged 30–44 years, the average failure rate experienced by modern method users is 2.3%, while 15.1% of the traditional method users will conceive in a year. For the younger age group these failure rates are higher at 4% and 17.3%, respectively, due to their greater reliance on less effective reversible methods, and a shorter length of time for which current users have been using the methods (calculations not shown).

Table 15.3 Aggregate failure rates, calculated from method-specific failure rates and method mix for women aged 30–44 years, all countries combined

	Average method-specific failure (%)	*Method mix relative distribution*	*Aggregate failure rate (%)*
Modern methods			2.3
Pill	5.0	0.21	
IUD	1.4	0.22	
Injections	2.2	0.11	
Diaphragm/foam/jelly	13.3	0.01	
Condom	7.1	0.06	
Norplant	0.1	0.01	
Female sterilization	0.1	0.36	
Male sterilization	0.3	0.02	
Traditional methods			15.1
Periodic abstinence	14.9	0.60	
Withdrawal	15.3	0.32	
Other traditional	15.1	0.08	

Conception rates

The estimation of the expected conceptions among the non-users was based on fecundability, which is defined as the probability of conceiving in a month among fecundable women (Bongaarts and Potter 1983). Fecundable women are those capable of conceiving. Conception refers to recognizable conception signified by the delay of first menses after fertilization. Some non-users have obvious biological or behavioural characteristics that make them temporarily unexposed to the chance of conceiving: the currently pregnant, amenorrhoeic women, women who have not resumed sex since the most recent childbirth and women who reported no sex in the past year. Other non-users, such as those who have reached menopause or know themselves to be infecund, are permanently unexposed to the risk of conception. These women have thus been excluded and the conception rates applied to fecundable women who have been sexually active over the past year.

Fecundability is difficult to assess empirically and it is usually estimated from waiting times to conception. Most reliable estimates are available from measuring the length of interval between marriage to first birth among couples using no contraception. Coital frequency is the dominant behavioural determinant of fecundability and most of the variation of fecundability by age can be attributed to a decline in intercourse (James 1979). Fecundability has been tabulated either by age or by coital frequency, but not by both (Bongaarts and Potter 1983; Leridon 1977).

Without sex there can be no conception but there are other biological requirements: the woman needs to ovulate, and insemination must lead to a successful fertilization, which then has to result in a recognizable conception. Mathematical modelling of age-specific fecundability shows that the ability to conceive is quite constant between ages 25 and 40, while the ability to maintain a pregnancy starts to decline much earlier (Weinstein et al. 1990; Wood and Weinstein 1988). In our calculations intrauterine mortality was accounted for in the second step (Figure 15.1) when we estimated how many of the unintended pregnancies resulted in unwanted births and unsafe abortions.

While biological determinants of fecundability are fairly constant across populations, differences in frequency of sexual intercourse do have a substantial impact on fertility (Brown 2000; Weinstein et al. 1993). Increased frequencies of sexual intercourse raise fecundability, but the relationship is not linear. When coital frequency is low, the chance of conceiving is proportional to the frequency, but, at higher levels of monthly frequency, further increases are minimal (Potter and Millman 1986; Weinstein et al. 1990).

Should age-based or coitus-based estimates of fecundability be used? DHS data allow the calculation of both. Most surveys enquire about the most recent date of last sexual intercourse and frequency in the last month. We have used the data on the most recent date of the last intercourse to infer coital frequency, because it is less prone to recall error and normative responses than the question on coital frequency in the last month (Becker and Begum 1994). It is also available for more surveys included in our calculations. When the individual probability of coitus is constant throughout the month, the interval between two acts of intercourse is the reciprocal of the frequency. It has been shown—mathematically and empirically—that the distributions of the time since last sex and the interval between two acts have the same mean and variance (Leridon 1993). We could therefore estimate the coital frequency from the mean time since last sex. Our calculations were based on women who have been sexually active in the past year. Among these women, a large proportion did not have sex during the last month. Therefore, the average time since last sex was converted into monthly coital frequency for those who did have sex within the last month, while the others were given the value of 0 in order to derive the average monthly frequency among all women sexually active during the past year. Twelve countries lacked data on time since last sex, and for those we imputed frequencies taking the average values estimated from other countries for the three levels of fertility intention. The calculation of mean coital frequencies was done separately for samples of ever-married and samples that included all women. The frequencies for all women typically give lower values, especially among women who want to space their births (hereafter referred to as "spacers"), which include many women who may not have a regular partner, or not live with a partner.

Since the frequency of intercourse varies considerably according to childbearing intentions, with spacers and women who want to limit their births (hereafter referred to as "limiters") having less frequent sex than women who want a birth within the next two years, we have calculated expected pregnancies in two ways. The first uses model estimates of fecundability based on coital frequency (Bongaarts and Potter 1983), allowing different conception rates by fertility intention, while the second uses simple age-specific fecundability estimates, with monthly fecundability declining from 0.25 in the early 20s to 0 by age 45 years (Leridon 1977).

Table 15.4 shows the differentials in monthly coital frequencies by fertility desire and its impact on fecundability for the two age groups. Fecundability was expressed as the number of women expected to become pregnant at the end of one year, and this was contrasted with the age-specific fecundability as calculated from the monthly model estimates (Leridon 1977).

Overall, the mean coital frequencies derived from data on duration since last sexual intercourse seem very low, but reflect the fact that these are based on all women who had sex during the past year (rather than on those in stable cohabiting relationships). The effect of including unmarried, non-cohabiting women is especially evident among the spacers aged 15–29 years, a group which includes 25% of never-married women, compared with 13% among the limiters and 7% among those who desire a child soon. Single women have sex less frequently—and are likely to underreport sexual activity. For the 12 countries with ever-married samples, higher levels of coital frequency result in higher expected fecundability. But in most societies, sexual intercourse and therefore risk of pregnancy is not restricted to marriage. A study among nine African countries has shown that the time spent in marriage is not always a good proxy for sexual activity either, with high levels of inac-

Table 15.4 Mean monthly coital frequency by age and fertility preference for fecund non-users, and estimates of coitus-based and age-based fecundability

	15–29 years	*30–44 years*
Mean coital frequency		
Want children soon	4.4	3.9
Spacers	2.1	3.0
Limiters	3.5	2.0
Coitus-based fecundability (proportion pregnant in a year)		
Want children soon	0.82	0.78
Spacers	0.54	0.71
Limiters	0.72	0.54
Age-based fecundability (proportion pregnant in a year)	0.97	0.75

tivity recorded by married women, especially in West Africa (Brown 2000). The differential in fecundability according to fertility preference persists when the analysis is restricted to married women only and remains important.

Guided by consistency checks on internal validity of the data (discussed below), the relative risk estimates presented in this chapter were calculated using *age-based* fecundability estimates. Because of the importance of the effect, however, we will return to this topic in the section on uncertainty of estimates at the end of the chapter.

3.4 Combining exposure and fertility intention

The expected number of unwanted pregnancies was calculated separately from mistimed pregnancies. In fact, pregnancies were estimated for all nine combinations of exposure (modern method use, traditional method use, no use) and fertility intention (want birth soon, later, never). As shown schematically in Figure 15.2, contraceptive failures can thus result in pregnancies that are classified as intended. They occur among women who want a birth within the next two years.

More detail is provided in Figure 15.3 showing the calculation of the expected proportion of women having an unwanted pregnancy in the next year for each level of exposure, using data for women aged 30–44 years (all surveys combined).

Among the non-users, 11% were currently pregnant, 19% were amenorrhoeic or had not resumed sex since last birth, 21% were infecund or menopausal and another 10% had not had sex in the past year, leaving

Figure 15.2 Combining data on exposure and fertility intention to estimate pregnancies

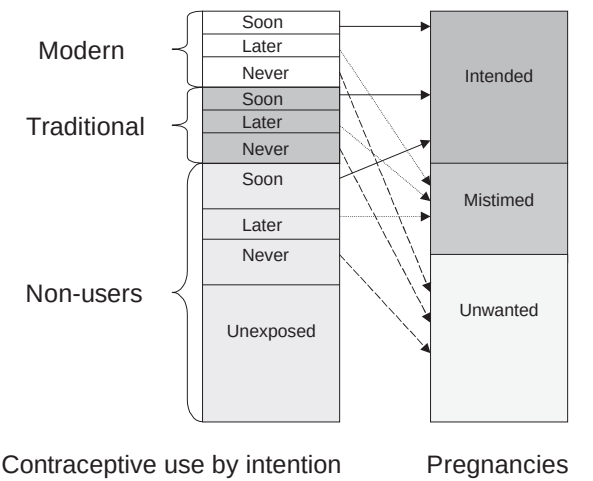
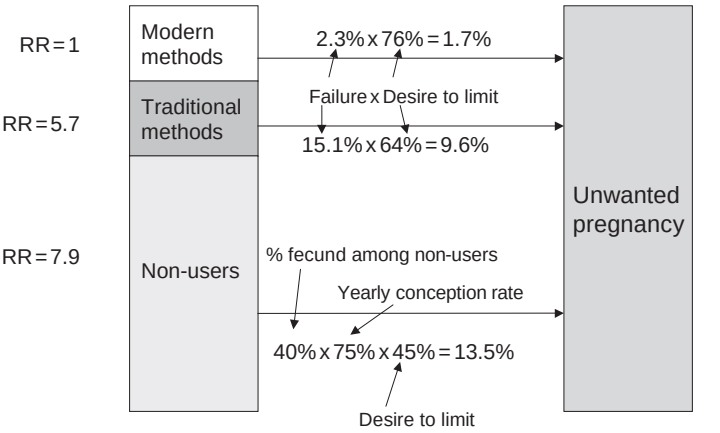


Figure 15.3 Expected proportion of women having an unwanted pregnancy in the next year, by exposure



only 40% of the non-users exposed to the risk of conception. Forty-five per cent of these fecundable women wanted to limit their families (i.e. have no more children) and, based on their coital frequency, they had a 75% probability of conceiving in a 12-month period. The product of these three numbers (40% × 45% × 75%) gave the proportion of all non-users who were expected to have an unwanted pregnancy in the next 12 months, namely 13.5%.

Among traditional (and modern) method users there may also be women who are biologically or behaviourally unexposed, but they are implicit in the failure rates (in contrast to conception rates). In other words, the presence of unexposed women among users will depress failure rates and thus obviate the need to take further account of such women. The calculation of the expected number of unwanted births was simply the product of the 12-month failure probability and the percentage of those who wanted no more children. For all surveys combined, the failure rates for modern and traditional methods were 2.3% and 15.1%, respectively, and the percentage wanting no more children was 76% among modern method users and 64% among traditional method users. The calculations in Figure 15.3 show that a non-user is 7.9 times (13.5% / 1.7%) more likely than a modern method user to have an unwanted pregnancy in the next year.

Table 15.5 illustrates the calculation of the distribution of expected pregnancies in the next year, according to fertility intention. The right-hand panel shows that among the modern method users, 1.7% will experience an "unwanted" failure, while 0.1% will have an "intended" failure. Among 100 non-users, 13.5 will have an unwanted pregnancy and there will be 9.6 intended pregnancies. The last row gives the

Table 15.5 Distribution of expected pregnancies estimated from contraceptive use by fertility intention (women 30–44, all surveys combined)

	Level of exposure: Contraceptive use (%)	*Fertility intention:*[a] *want birth*			*Percentage of women expected to have intended, mistimed and unwanted pregnancies (within level of exposure*		
		Soon[b]	*Later*	*Never*	*Intended*	*Mistimed*	*Unwanted*
Modern users	28.2	5	18	76	0.1	0.4	1.7
Traditional users	8.1	11	25	64	1.7	3.8	9.6
Non-users	63.8	32	22	46	9.6	6.5	13.5
Total	100				6.3	4.6	9.9

[a] Per cent distribution with % wanting birth soon + % wanting birth later + % never wanting birth adding to 100%.

[b] Soon means a birth within the next 2 years.

expected pregnancies, weighted by the contraceptive distribution. In total, 21% of women aged 30–44 years are expected to get pregnant at current levels of contraceptive use. According to prevailing fertility intentions, 10% of women will have an unwanted pregnancy.

3.5 Estimating pregnancy outcomes

This section refers to the second step in Figure 15.1, estimating unwanted births and unsafe abortions from unwanted and mistimed pregnancies.

Spontaneous pregnancy loss

Before applying abortion probabilities to unwanted and mistimed pregnancies, we allowed for miscarriages and stillbirths. Recognizable intrauterine mortality is lowest in the early twenties (16%), but reaches double this rate by age 45 years (Bongaarts and Potter 1983). From these tabulated data we calculated the average spontaneous pregnancy loss as 17% and 27% for the younger and older women, respectively. Pregnancy loss does not vary by fertility intention, and the expected pregnancies in Table 15.5 need to be reduced by 27%, leaving 15.3% of women aged 30–44 years pregnant. Thus after accounting for pregnancy loss, 7.2% of women have an unwanted pregnancy and 3.4% have a mistimed pregnancy. A proportion of these will be aborted while the rest will result in a live birth.

Abortion probabilities

Abortion probabilities have been derived by converting the WHO country estimates of incidence ratios (unsafe abortions per 100 live births) to abortion probabilities (ratio of abortions to abortions plus births). For countries without estimates ($n=10$) we took the WHO

regional abortion estimate (WHO 1998). The average abortion incidence ratio of 17.5 abortions per 100 live births translates into an overall abortion probability per pregnancy of 0.15 (= 17.5/(100+17.5)). Because intended pregnancies are most unlikely to be aborted, this probability needs to be converted to relate to unintended pregnancies only, while keeping the overall abortion ratios constant.

The main reason stated by women for having an abortion is to stop bearing children; following this is the wish to postpone pregnancy (Bankole et al. 1999). Other reasons included disruption of education and the belief that they were too young to have children, especially in Africa where single women are sexually active, all adding to the postponement component of abortion. An in-depth study in Maharashtra, showed that 54% of aborted pregnancies among married women were defined as unwanted and 42% as mistimed at the time of conception (Ganatra et al. 2000). In three central Asian republics, the proportion of unwanted pregnancies that were aborted ranged from 74% to 86%, while two-thirds of all mistimed pregnancies were aborted (Westoff et al. 1998).

Consistent with the limited evidence available, our calculations assumed that the abortion probability of a mistimed pregnancy was half that of an unwanted pregnancy. Though this assumption is essentially arbitrary, it is consistent with the judgements of experts. It can be regarded as reasonable but nevertheless must be viewed with caution.

The relative distribution of mistimed, unwanted and planned pregnancies expected for 100 women (at the current contraceptive prevalence and childbearing intentions) was used to relate abortions to mistimed and unwanted births only. Of the 15.3 pregnancies among 30–44-year olds (after spontaneous pregnancy loss), 3.4 are expected to be mistimed and 7.2 unwanted. So the abortion probability of unintended pregnancies is calculated as $0.15 \times 15.3/(7.2+3.4/2) = 0.25$. This probability implies that one in four unwanted pregnancies will be aborted compared with one in eight of mistimed pregnancies. Although we started from the same incidence ratio for both age groups, the fact that only unintended pregnancies are aborted leads to different abortion probabilities for the two age groups. In the countries with high rates of unsafe abortion and lack of systematic data, we know little about how abortions vary by age. Where official statistics are more complete, there are generally two age patterns of abortion ratios (Bankole et al. 1999). The first takes a U-shape, where abortion is high both among unmarried young women and older women who have reached their desired family size. The second pattern is a steady increase in the abortion ratio with age. As default we have chosen to use the same ratio for both age groups. Where consistency checks indicated negative levels of unwanted births in the younger age group (Guinea, Indonesia and Thailand) or where data on age patterns were available (Kazakhstan, Kyrgyzstan, Uzbekistan) the ratio was adjusted by age. For countries that needed correction, the abortion prob-

ability of (any) pregnancy to women aged 30–44 years was assumed to be three times the probability for pregnancies to younger women (estimate based on the data provided for the latter three countries in the article by Bankole et al. 1999).

We have implicitly assumed that abortion probabilities of unwanted (and mistimed) pregnancies are the same regardless of whether they resulted from method failure or non-use. Common sense suggests that women who act on their intention to prevent an unwanted pregnancy by adoption of contraception are more determined to regulate fertility than other women and thus more likely to seek a termination when pregnant. Very limited empirical evidence on differential abortion probabilities supports this expectation. In Turkey, for instance, 28.5% of unintended pregnancies in 1998 resulting from non-use were aborted, compared with 38.1% and 35.2% of pregnancies resulting from modern and traditional method failure, respectively (Senlet et al. 2000). In Kazakhstan, 51% of unintended pregnancies among non-users were aborted compared with 67% of (all) contraceptive failures (Westoff 2000). The inferred differences are relatively small and their generalizability is unknown. Hence, it was decided to apply the same abortion probabilities for all three contraceptive use categories. Nevertheless, since the desire to discontinue childbearing altogether is highest among modern method users, our simulations did result in higher abortion probability for all modern method failures relative to traditional method failures and conceptions among non-users.

Probability and relative risk of having an abortion

Abortion probabilities have been combined with expected proportions of women who have unwanted and mistimed pregnancies in each exposure category. Table 15.6 provides the worked example of the expected proportion of women aged 30–44 years (all surveys averaged) having an abortion. The expected percentages of women having pregnancies are 27% lower than in Table 15.5, since they are adjusted for spontaneous pregnancy loss.

The abortion probability of 0.25 for unwanted births results in an expected proportion of 3.1% of non-users having an abortion in the next

Table 15.6 Probability of having an abortion in each exposure category and the resulting relative risk ratios

	Expected % of women having pregnancies			*Expected % of women having an abortion*	*RR of having an abortion*
	Intended	*Mistimed*	*Unwanted*		
Modern users	0.1	0.3	1.3	0.4	1.0
Traditional users	1.2	2.8	7.1	2.2	6.0
Non-users	7.1	4.9	9.9	3.1	8.7

year (0.25×9.9% [unwanted pregnancies]+0.25/2×4.9% [mistimed pregnancies]). Following the same logic, only 0.4% of modern method users are expected to have an abortion. Using the modern method users as the reference category the relative risk was derived as the ratio of expected proportion of women having an abortion among non-users compared with modern method users (RR=8.7). Similarly, traditional method users were projected to be six times more likely to have an abortion than modern method users.

Probabilities of women in each exposure category of having an unwanted birth

The proportions of modern and traditional method users and of non-users who are expected to deliver an unwanted birth were calculated by applying the complement of the abortion probability to the unwanted pregnancies. Mistimed pregnancies that end as live births do not contribute to the burden of maternal outcomes. The worked example in Table 15.7 shows that with an abortion probability of 0.25, three-quarters of the 9.9% of non-users with an unwanted pregnancy are expected to carry it to term. With 7.4% of non-users having an unwanted birth in the next year, compared with 0.9% of modern method users, the relative risk is 7.8.

Since the WHO abortion estimates relate to unsafe abortions only (which is required for the calculation of relative risk to have an unsafe abortion), we have potentially overestimated the expected number of unwanted births in countries where legal abortions are common. This does not affect the relative risk of unwanted birth, but it does affect the proportion unwanted among all births. As can be seen from Table 15.1, most surveys were done in countries with highly restrictive abortion laws, and therefore most abortions will be unsafe. For the countries with available data on legal abortions (Henshaw et al. 1999), the proportions of unwanted births among all births were calculated using legal abortion rates. The legal abortion ratio was used for Turkey. For Bangladesh and India, the official legal rates are very low and are underestimates of actual procedures performed (Henshaw et al. 1999). In India, abortion

Table 15.7 Probability of having an unwanted birth in each exposure category and the resulting relative risk ratios

	Expected % of women having pregnancies			*Expected % of women having an unwanted birth*	*RR of having an unwanted birth*
	Planned	*Mistimed*	*Unwanted*		
Modern users	0.1	0.3	1.3	0.9	1.0
Traditional users	1.2	2.8	7.1	5.3	5.6
Non-users	7.1	4.9	9.9	7.4	7.8

has been legal for 30 years, but abortion services by authorized facilities are inadequate, especially in rural areas. Many women are even not aware that abortion is legal and resort to abortions from both unskilled and skilled providers (Ganatra et al. 2000). In Bangladesh, with highly restrictive abortion laws, the legal rate refers to menstrual regulation services (manual aspiration evacuation of the uterus without prior confirmation of pregnancy). They are widely available, effectively providing abortion up to eight weeks of a woman's last menstrual period (Rahman et al. 1998). For India and Bangladesh we used the WHO estimates on unsafe abortions.

3.6 Consistency checks

Two basic consistency checks guided the assessment of the plausibility of the data inputs and the method assumptions, and informed the adjustments done for a few individual countries. We compared both the total numbers of projected births and the proportion of unwanted births among all births with retrospective estimates based on the question "At the time you became pregnant with (NAME OF CHILD) did you want to become pregnant then, did you want to wait until later or did you want no more children at all?" Age-specific fertility rates published by DHS allowed the calculation of the yearly number of births in each age group. These rates were averaged for a period of three or five years prior to survey data in order to reduce sampling error. The number of recent births per year should agree with our expected number of births per 1000 women in each age group, after adjusting for the proportions of women who never had sex. Exact matches for each survey are not expected, but systematic patterns of excess or shortfall of births may imply regional biases.

Compared with the current births calculated from the age-specific fertility rates, the projection using age-based fecundability estimates (averaging the outcomes for all surveys) results in a 15% shortfall of expected births for the younger age group, and 2% excess for the older age group. What may contribute to the underestimation of expected births? A minor factor may be that we discount miscarriages and stillbirths before calculating abortions. Since some of the aborted pregnancies might have resulted in a spontaneous intrauterine death, we may have underestimated the pregnancies carried to term. A second factor may be abortion probabilities that are too high. While this is unlikely overall, the assumption of constant incidence ratios by age may contribute to the shortfall of expected births in the youngest age group. Continuing fertility decline would also contribute to the discrepancy between projected and retrospective fertility levels. Finally, underreporting of sexual exposure (both overstating virginity and time since last sex) may well be a major cause of the deficit of births to young women.

It should be pointed out that in terms of estimating the relative risk ratios of having an unwanted birth or an abortion, overall numbers of

births are less important than the relative distribution of births across fertility intention. As we indicated before, frequency of sexual intercourse is not the same for women who want a birth soon and those who either want to space or stop childbearing. Moreover, there is no reason to doubt the validity of these *relative* differences. This is an important factor contributing to the uncertainty around the estimates (discussed later).

This leads us into the second consistency check, the comparison of projected and retrospective estimates of the proportion of births that are unwanted. Compared with the proportion of the most recent births that were reported as unwanted, serious discrepancy was apparent with projected proportions of unwanted births: they were 63% and 143% higher for the 15–29-year olds and the 30–44-year olds, respectively. We discuss below potential biases that may account for the discrepancy.

The proportion unwanted among most recent births is available if this birth occurred within the last five, or in some surveys, three years. It is derived from the question "At the time you became pregnant with (NAME OF CHILD), did you want to have more children then, did you want to wait until later, or did you want no more children?" The question is meant to draw on the memory of the feelings that were held at the time of conception. Though little evidence exists to evaluate to what extent respondents later report as being wanted those children whose conception was initially unwanted, the most important reason why the "wantedness" of the most recent birth may not agree with our projected number of unwanted births is undoubtedly ex-post rationalization (Bongaarts 1990; Westoff 1981). Panel data in Morocco allowed the comparison of reports on the "wantedness" of 0–2-year-old children in 1992 with reports on the same births three years later (Westoff and Bankole 1998). Whereas 6% of children reported in 1992 as wanted at the time of conception were later described as unwanted, as many as 62% of the unwanted pregnancies in the first round were reported as wanted in 1995. The older the child, the more likely reports on the feelings or intentions had changed from unwanted to wanted. A comparison among five DHS confirms the result from Morocco. The percentage unwanted declines as the age of children increased, though the trend was less pronounced than in the Morocco data (Montgomery et al. 1997). Data on change in perception before and after a baby is born are not available. However, in a recent qualitative study in the United Kingdom of Great Britain and Northern Ireland among pregnant women and those who recently aborted, women generally agreed that conceptions which were initially "unplanned" or "unintended" could become "wanted", a term many women associated with the decision to carry the pregnancy to term (Barrett and Wellings 2002). Since we do not know how these concepts are translated and understood by different cultures, the magnitude of the rationalization bias cannot be quantified.

An additional explanation is that the most recent birth refers to a child of lower birth order than the projected next birth. Calculating order-specific estimates on the "wantedness" of the last child and applying these to the birth order distribution shifted by one child allows assessment of the magnitude of this order effect on fertility preferences. The average effect is surprisingly low: 9% of women aged <30 years report their most recent birth as being unwanted, while this would be raised to only 11% by shifting the order by one child. For the older women the effect was equally small.

A last potential factor contributing to the discrepancy might be the underestimation of the proportion of unwanted pregnancies that are terminated. By considering unsafe abortions, rather than all abortions, we may have overestimated the expected unwanted pregnancies carried to term. When the relative abortion probability of unwanted over mistimed pregnancies is lower than two, then we could again overestimate the number of unwanted births.

Given the considerable differences between the various measures of preferred fertility and the known bias in underreporting of unwanted births (Bongaarts 1990; Westoff 1981), we have to tolerate a level of inconsistency between projected and retrospective estimates of "wantedness". Important biases other than our model assumptions are operating and this precludes the use of the extent of discrepancy as a guide to make adjustments. Nevertheless, for selected countries with available data on legal rates and age pattern of abortion, adjustments were done (as explained above). We have been guided more by disparity between the age groups, than any discrepancy in expected level of fertility. While the abortion adjustments do not affect relative risk ratios, they do lower the proportion of unwanted among all births, which is used in projecting how much of the obstetric burden can be avoided by reducing unwanted births.

3.7 Counterfactual scenario

Counterfactual distribution of contraceptive use

The burden of maternal outcomes, including abortion, attributable to lack of effective modern contraception was calculated as the reduction in current burden that would be observed if levels of exposure were reduced to a counterfactual distribution of contraceptive use. Theoretical-minimum-risk distribution of contraception does not mean 100% modern use, but rather that all women with a desire to either stop or postpone childbearing for at least another two years, adopt an effective modern method of contraception. Perfect implementation of fertility preferences among limiters and spacers obviously results in a higher proportion of modern method users and fewer women using traditional contraception or no contraception at all. All traditional method users and fecund non-users now consist of women who want a birth in the next

two years. This theoretical minimum level of exposure was thus simply calculated by shifting all spacers and limiters into modern method use.

The potential impact fractions were used to estimate the proportional reduction in the total number of unwanted births and unsafe abortions by a change in contraceptive prevalence. Potential impact fractions generally assume that only exposure changes, while the relative risk of the outcome for each level of exposure stays the same. However, as we demonstrated in the previous section, the relative risk of an abortion (or unwanted birth) among the traditional method users not only depends on the failure rates, but also on the fertility preference among these users. The factors determining the relative risk among non-users are conception rates, fertility preferences and the proportion fecund among the non-users. While both the failure and conception rates remain constant, the relative fertility preferences in each exposure category and the proportion of women who are fecund among non-users will change with a change in contraceptive distribution.

Counterfactual relative risk ratios

How do we expect the relative risk to vary with a change in exposure? Under the scenario of theoretical-minimum-risk, all women with a desire to stop or space childbearing will adopt modern methods and all expected conceptions in traditional method users and non-users will now be intended pregnancies. Because only the reference group (modern users) is at risk of an unintended pregnancy, the relative risk of unwanted births and abortions will be 0 in other groups. For intermediate levels of shifting the counterfactual relative risk will be between 0 and the current level of relative risk.

In deriving these counterfactual levels of relative risk, the degree of shifting was assumed to be the same for spacers and limiters. The calculation involved computing counterfactual proportions of fecundable women among the non-users, and counterfactual fertility preferences in all three categories of exposure.

For each level of shifting, the counterfactual distribution of contraceptive use was determined by subtracting the number of women using traditional methods and no contraception at all and adding that to those using modern methods. The distribution of non-users was recalculated by keeping the number of infecund/menopausal women and those that were not sexually active in the past year constant; the number of fecundable women changed by subtracting the number of limiters and spacers who shifted to using modern methods; the number of pregnant and amenorrhoeic women changed since more effective contraception implies a reduction in the number of births, with at any one time a smaller proportion of women pregnant or in the amenorrhoeic state. Through a process of iteration we imputed the ratio of births in the counterfactual over current population and assumed that the number of pregnant and amenorrhoeic change to the same extent. These numbers were combined

in the counterfactual proportion of fecund women among non-users. The counterfactual distribution of fertility desires among the three categories of exposure was calculated taking account of the level of shifting and counterfactual proportion of fecundity among the non-users. Table 15.8 compares current with counterfactual exposure and fertility intentions, calculated for women aged 30–44 years, all countries combined, assuming 50% shifting.

These counterfactual contraceptive distributions, fertility desires and proportion fecund among non-users were then combined to obtain relative risks in the same way as explained in the previous section for the "current" relative risk levels. Table 15.9 contrasts the counterfactual relative risk for the same scenario of 50% shifting, among women aged 15–29 years and 30–44 years, with the relative risk levels under current exposure.

The relative risks among non-users initially decline faster with the degree of shifting, compared with levels among traditional method users. When half of all non-users and traditional users who currently have an unmet need for spacing or limiting have adopted a modern method, the relative risk for a non-user to have an unwanted birth has decreased from 7.8 to 5.0. The relative risk for traditional users drops to a lesser extent (from 5.6 to 5.2). The steeper decline among the non-users can be explained mainly by the fact that the infecund, the menopausal and the sexually inactive women gradually become a larger proportion of all non-users. The pattern of the counterfactual relative risk levels by degree of shifting varies across countries. The average pattern (for all surveys combined, women aged 30–44 years) shows the most common pattern as depicted in Figures 15.4 and 15.5 of a near-linear decline for

Table 15.8 Current and counterfactual contraceptive distribution and fertility preferences among women aged 30–44 years, all surveys combined

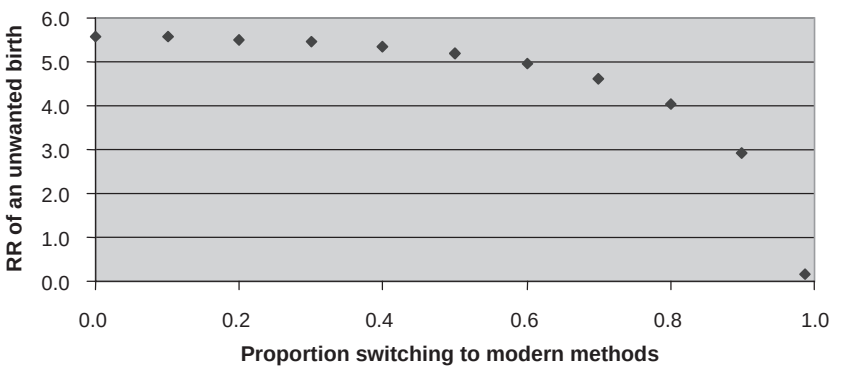
	Current exposure and intention			
	Contraceptive use	*Want birth soon*	*Want to space*	*Want to limit*
Modern	28.2	0.05	0.18	0.76
Traditional	8.1	0.11	0.25	0.64
Non-use[a]	63.8	0.32	0.22	0.46
	Counterfactual exposure and intention (50% shifting)			
Modern	37.4	0.04	0.22	0.74
Traditional	4.6	0.20	0.22	0.58
Non-use[a]	58.1	0.49	0.17	0.34

[a] The percentage fecund among the non-users decreases from 40% to 32%.

Table 15.9 Relative risk ratios for unsafe abortions and unwanted births under current and counterfactual exposure, both age groups, all surveys combined

	RR under current exposure		*RR under counterfactual exposure assuming 50% shifting*	
Abortions	15–29 years	30–44 years	15–29 years	30–44 years
Modern	1.0	1.0	1.0	1.0
Traditional	3.8	6.0	3.2	5.4
Non-use	6.2	8.7	3.8	5.4
Unwanted birth				
Modern	1.0	1.0	1.0	1.0
Traditional	2.9	5.6	2.7	5.2
Non-use	4.5	7.8	3.1	5.0

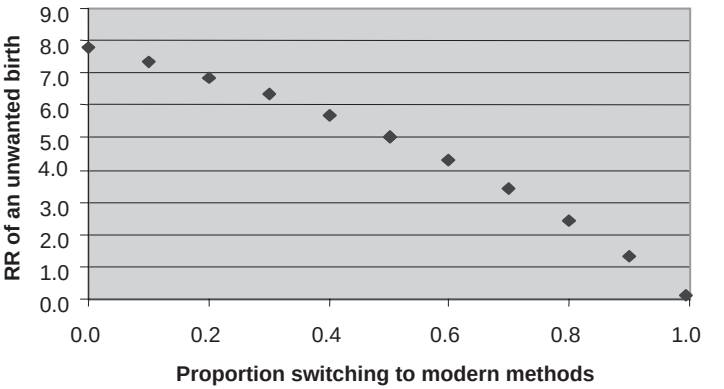
Figure 15.4 Pattern of change in relative risk of an unwanted birth by level of switching for traditional method users, women aged 30–44 years



non-users, while the relative risks for traditional method users tend to stay relatively constant up to 50–60% shifting, declining rapidly to 0 thereafter.

For traditional use the relative risk of unwanted births and abortions sometimes increases initially at lower levels of shifting. This seems to happen when more spacers than limiters move into the modern method users, which has little effect on the distribution of fertility preferences among the traditional users, but affects the expected proportion of modern users who will be having relatively more "mistimed failures"

Figure 15.5 Pattern of change in relative risk of an unwanted birth by level of switching for non-users, women aged 30–44 years



rather than "unwanted failures". Thus the expected proportion of modern method users having unwanted births actually decreases slightly. Since the expected proportion of unwanted births among the non-users changes much faster, this pattern of an initial relative risk rise is not detected so much among the non-users.

3.8 Aggregating country-specific estimates into subregions

The country level results in terms of contraceptive exposure, relative risks of unsafe abortions and unwanted births were aggregated into subregional averages, weighting each country according to the size of the population (see Table 15.1 for weighting factors). The derived estimate was then taken as subregional average, assuming that the countries were reasonably representative of the whole subregion. Table 15.10 presents the coverage of subregional population by countries with data.

As can be seen from the subregional coverage presented in Table 15.10, the African, Latin American and some of the Asian subregions are well represented. The three subregions where relevant data are totally lacking are AMR-A, EUR-A and WPR-A, which mainly consist of industrialized low-fertility countries with small burdens of maternal mortality or unsafe abortion. The exposure distributions for these subregions were imputed from data provided in the United Nations (UN) report on levels and trends in contraceptive use (UN 1999). As the UN data were for married women aged 15–49 years, we assumed 90% of total estimate for the age group 15–29 years and 110% of estimate for 30–44-year olds. Since safe abortion is widely available in most of these countries, the relative risk levels were arbitrarily set at 1.5 for unwanted births (both for traditional users and non-users). For abortion we took relative

Table 15.10 Coverage of subregional population by countries with data from DHS

Subregion	*Number of DHS per subregion*	*Percentage of the subregional population covered by countries with DHS data*
AFR-D	9	80
AFR-E	14	69
AMR-A	0	0
AMR-B	6	79
AMR-D	6	100
EMR-B	1	8
EMR-D	5	85
EUR-A	0	0
EUR-B	3	46
EUR-C	1	7
SEAR-B	3	100
SEAR-D	3	93
WPR-A	0	0
WPR-B	1	5

risks of 2 and 3, respectively, for traditional method users and non-users. It matters little what the real risks are since the burden of obstetric and abortion complications is negligible. Subregions that are cause for greater concern are EMR-B, EUR-C and WPR-B. Tunisia is the only country representing EMR-B and Kazakhstan the only one in EUR-C. Most worrying is that WPR-B, a subregion in which 83.5% of the population live in China, is represented solely by the Philippines. In China, because of the strict anti-natal policies, there will be very few unwanted births and most abortions are legal and safe. Therefore an adjustment was needed for WPR-B, which is discussed in the next section on deriving attributable burden.

3.9 Subregional estimates of input parameters

The inputs provided for calculating the attributable burden are the contraceptive distributions and relative risks under the current situation and counterfactual (theoretical minimum) scenario, and the proportion of all births that are unwanted.

We first estimated the proportions of women who were excluded from the analysis because they never had sex and were therefore not exposed to the possibility of pregnancies (i.e. virgins). Table 15.11 gives the proportion of the female population that is presumed to be virgins. Our method overestimates virginity in two ways. In the countries with ever-married samples, single women do not enter the analysis and are assumed to be virgins. Clearly, this assumption is not entirely valid. In the other

Table 15.11 Average proportion of women who are virgins, by subregion

Subregion	*15–29 years*	*30–44 years*
AFR-D	23.6	0.5
AFR-E	26.1	0.5
AMR-A	15.0	1.5
AMR-B	39.4	4.1
AMR-D	43.0	3.7
EMR-B	65.1	7.4
EMR-D	48.8	4.2
EUR-A	15.0	1.5
EUR-B	46.5	3.1
EUR-C	42.1	2.1
SEAR-B	45.8	5.6
SEAR-D	27.6	1.2
WPR-A	15.0	1.5
WPR-B	59.8	8.7

countries virginity may be overstated to the extent that single women underreport sexual activity. The estimates for the three subregions where we lacked data, AMR-A, EUR-A and WPR-A, were informed by data from the British Sex Survey in 1990/91 (Johnson et al. 1994).

Current contraceptive use, averaged for each subregion, is shown in Table 15.12 and the counterfactual scenario in Table 15.13. The second table thus provides the contraceptive status that would have been observed if all women with a current need for spacing and limiting were to adopt a modern method. Note that a small residue of traditional method users remains; these are women who want a child within the next two years. The subregional variation in *current* contraceptive use is thus explained by variations in desired family size and extent of implementation of these desires through contraception (and abortion). Subregional variations in *counterfactual* contraceptive use reflect differences in desired family sizes and use of abortion as alternative means to implement fertility intentions.

Tables 15.14 and 15.15 show the relative risk estimates for having an unwanted birth and an unsafe abortion, respectively, under the current regime of contraceptive practice and fertility preferences. The levels for the counterfactual minimum risk have not been provided since they are all 0, as explained earlier. As can be seen, there is wide subregional variation in relative risks, both for traditional method users and non-users. The subregion that stands out is SEAR-D, where India represents a large part of the population. Modern contraception in India (and

Table 15.12 Averages of current contraceptive distribution among women who ever had sex, by subregion

	15–29 years			*30–44 years*		
Subregion	*Modern method*	*Traditional method*	*Non-use*	*Modern method*	*Traditional method*	*Non-use*
AFR-D	8.0	9.1	82.9	10.7	7.7	81.6
AFR-E	13.0	5.4	81.6	15.6	5.6	78.8
AMR-A	60.6	2.6	26.8	74.1	3.2	32.8
AMR-B	48.1	6.9	45.0	62.1	6.8	31.1
AMR-D	27.9	11.6	60.5	36.2	14.6	49.2
EMR-B	30.4	7.7	61.9	44.8	10.8	44.5
EMR-D	14.9	2.6	82.4	23.5	4.1	72.4
EUR-A	62.1	5.2	22.7	75.9	6.4	27.7
EUR-B	36.0	16.6	47.4	46.9	21.8	31.3
EUR-C	42.4	12.4	45.2	55.5	12.2	32.3
SEAR-B	54.4	2.6	43.0	55.9	4.1	40.0
SEAR-D	23.5	4.0	72.5	48.3	5.3	46.4
WPR-A	48.7	5.0	36.3	59.5	6.1	44.4
WPR-B	24.1	18.0	57.9	29.5	20.3	50.2

Table 15.13 Averages of counterfactual contraceptive distribution (theoretical minimum) among women who ever had sex, by subregion

	15–29 years			*30–44 years*		
Subregion	*Modern method*	*Traditional method*	*Non-use*	*Modern method*	*Traditional method*	*Non-use*
AFR-D	43.1	2.5	54.4	42.5	1.8	55.8
AFR-E	47.7	1.5	50.8	49.9	0.8	49.4
AMR-A	60.6	2.6	26.8	74.1	3.2	32.8
AMR-B	84.7	1.1	14.3	84.4	0.7	14.9
AMR-D	72.9	2.0	25.1	72.8	1.7	25.5
EMR-B	66.2	1.4	32.5	82.6	1.0	16.4
EMR-D	48.6	0.4	51.0	62.3	0.4	37.4
EUR-A	62.1	5.2	22.7	75.9	6.4	27.7
EUR-B	70.3	2.7	27.0	82.4	0.8	16.8
EUR-C	76.1	2.0	21.9	79.5	0.9	19.6
SEAR-B	71.2	0.3	28.5	73.0	0.5	26.6
SEAR-D	60.7	0.9	38.4	74.0	0.2	25.8
WPR-A	48.7	5.0	36.3	59.5	6.1	44.4
WPR-B	72.7	1.2	26.1	70.0	1.0	28.9

Table 15.14 Average relative risk of having an unwanted birth under the current contraceptive scenario, by subregion

	Traditional method users		*Non-users*	
Subregion	*15–29 years*	*30–44 years*	*15–29 years*	*30–44 years*
AFR-D	2.3	3.2	2.8	3.5
AFR-E	1.6	4.4	3.0	4.9
AMR-A	1.5	1.5	1.5	1.5
AMR-B	3.2	7.9	8.3	14.3
AMR-D	4.1	6.4	6.2	9.8
EMR-B	3.2	5.9	3.9	8.4
EMR-D	3.2	5.1	3.4	6.8
EUR-A	1.5	1.5	1.5	1.5
EUR-B	4.0	6.3	5.1	11.1
EUR-C	5.1	8.6	6.9	12.7
SEAR-B	4.1	6.1	4.3	8.0
SEAR-D	4.6	22.2	4.6	27.8
WPR-A	1.5	1.5	1.5	1.5
WPR-B	3.7	5.9	5.8	7.7

Table 15.15 Average relative risk of having an unsafe abortion under the current contraceptive scenario, by subregion

	Traditional method users		*Non-users*	
Subregion	*15–29 years*	*30–44 years*	*15–29 years*	*30–44 years*
AFR-D	2.6	3.5	3.9	4.2
AFR-E	3.0	4.7	4.7	5.4
AMR-A	2.0	2.0	3.0	3.0
AMR-B	4.1	8.4	10.6	16.2
AMR-D	4.3	6.8	7.6	10.6
EMR-B	4.6	6.0	6.4	10.0
EMR-D	3.9	5.2	5.0	7.7
EUR-A	2.0	2.0	3.0	3.0
EUR-B	4.9	6.3	7.5	11.5
EUR-C	5.0	8.3	10.5	13.3
SEAR-B	5.5	6.4	6.4	9.2
SEAR-D	6.6	22.7	10.9	30.9
WPR-A	2.0	2.0	3.0	3.0
WPR-B	4.1	6.2	6.6	8.5

Nepal) is dominated by sterilization, which has a very low failure rate. As very few modern method users become pregnant, the relative risks for the other two categories are raised. This observation has complicated the interpretation of the counterfactual relative risks for abortions, as one of our model assumptions has obviously been violated. Keeping failure rates and contraceptive method mix constant with an increase in contraceptive prevalence has become internally inconsistent with the fertility preferences. Where sterilization is the dominant modern contraceptive method, it is impossible to accommodate the needs of spacers at the current method mix. Increased contraceptive use among spacers in such settings would have to involve uptake of reversible methods but this trend would reduce the average effectiveness of modern method use, because reversible methods have higher failure rates than sterilization. And this, in turn, would reduce the relative risk of an abortion for traditional method users and non-users. The attributable fractions derived from these relative risks of unsafe abortions for these countries with high use of sterilization will thus be a slight overestimate. However, since spacers do not contribute to unwanted births, the relative risks for unwanted births remain unaffected. Only when limiters adopt methods other than sterilization, as method choice widens, would the relative risks for unwanted births be affected.

4. Attributable burden

By combining the relative risk values for an unsafe abortion (Table 15.15) with the data on current and counterfactual distributions of contraceptive use (Table 15.12 and Table 15.13), we have derived the attributable fractions for unsafe abortions (Table 15.16). These estimates show that a very large proportion of the disease burden due to abortion complications is attributable to unprotected sex or use of less effective traditional methods. The residual is the "unavoidable" burden associated with modern method failure.

Table 15.17 gives the estimated burden of disease attributable to unsafe abortions by subregion. Both deaths and total DALYs have been presented as well as the relative burden in each subregion, expressed as DALYs per 1000 women, to allow better comparison between the subregions. The estimated burden of abortion attributable to non-use and use of ineffective methods of contraception is 4.4 million DALYs, with 82% of the burden falling on women aged <30 years. South Asia with its large population (SEAR-D) has the highest abortion burden at about 35% of the total abortion burden in both age groups. Although AFR-D and AFR-E include smaller populations than SEAR-D, women in these subregions have the highest relative burden.

The calculation of attributable burden for unwanted births is slightly more complicated. The attributable fraction for unwanted births could not be applied to the total burden of obstetric complications in

Table 15.16 Estimates of attributable fraction for unsafe abortions, by subregion

	Attributable fraction (%)	
Subregion	*15–29 years*	*30–44 years*
AFR-D	88	89
AFR-E	88	90
AMR-A	59	59
AMR-B	85	87
AMR-D	87	89
EMR-B	86	85
EMR-D	90	90
EUR-A	56	56
EUR-B	86	85
EUR-C	87	86
SEAR-B	79	84
SEAR-D	93	95
WPR-A	68	71
WPR-B	85	88

childbirth, since the many wanted births also contribute to maternal deaths and morbidity. We thus had to restrict the burden to that proportion of all births that was unwanted as simulated from contraceptive failure, conception rates and fertility preferences. These proportions are presented in Table 15.18, together with the unadjusted attributable fractions. These fractions express what proportion of unwanted births can be avoided by perfect implementation of fertility preferences. The proportion of unwanted among all births is much higher for the older age group, as the desire to limit family size is much more prevalent than among younger women. The negative values for women aged 15–29 years in EUR-C and SEAR-B reflect the fact that the abortion probabilities assumed for these women were probably still too high. In calculations the proportion was set to 0, and in the subregions concerned we may have slightly underestimated the burden of maternal complications in childbirth for the younger age group while overestimating it for older women. Finally, these proportions have been combined with the unadjusted attributable fractions (unwanted births only), to give the attributable fractions of all births that could be averted if all women who wish to stop childbearing used a modern method.

As mentioned earlier, the estimate for WPR-B needed adjusting since China, with very low rates of unwanted births, dominates the subregion. The easiest adjustment procedure was to keep the relative risk and contraceptive prevalence estimates for the Philippines, and adjust the

Table 15.17 Estimated burden of disease attributable to unsafe abortions (due to non-use of contraception), by subregion and age group in 2000

	Women 15–29 years				*Women 30–44 years*			
Subregion	*Attributable fractions (%)*	*Deaths*	*DALYs*	*DALYs per 1000 women*	*Attributable fractions (%)*	*Deaths*	*DALYs*	*DALYs per 1000 women*
AFR-D	88	5437	604075	14944	89	2671	91241	4062
AFR-E	88	10133	895872	18686	90	3717	124403	4741
AMR-A	59	12	496	15	59	14	370	10
AMR-B	85	808	69473	1151	87	435	13608	297
AMR-D	87	461	42841	4197	89	393	11887	1843
EMR-B	86	65	36873	1880	85	180	7894	680
EMR-D	90	2976	385594	8468	90	3601	115620	4018
EUR-A	56	8	1466	37	56	10	363	8
EUR-B	86	79	16217	574	85	37	1980	87
EUR-C	87	117	21562	788	86	103	4014	144
SEAR-B	79	1009	196086	4714	84	1745	62122	1952
SEAR-D	93	6445	1230514	7549	95	8507	310247	2647
WPR-A	68	0	31	2	71	2	49	3
WPR-B	85	834	86591	456	88	999	31689	176
World	89	28383	3587692	4710	91	22413	775486	1247

Table 15.18 Derivation of attributable fractions for unwanted births, proportions of unwanted births and attributable fraction for all births, by subregion and age

	Women 15–29 years			*Women 30–44 years*		
Subregion	*Attributable fraction (unwanted)*	*Unwanted births (%)*	*Attributable fraction (all)*	*Attributable fraction (unwanted)*	*Unwanted births (%)*	*Attributable fraction (all)*
AFR-D	84%	2.4	2%	87%	23.7	21%
AFR-E	82%	6.8	6%	88%	34.6	31%
AMR-A	42%	2.0	1%	42%	2.0	1%
AMR-B	81%	23.5	19%	85%	58.0	49%
AMR-D	84%	27.3	23%	88%	61.7	54%
EMR-B	78%	9.5	7%	83%	51.8	43%
EMR-D	85%	11.9	10%	89%	53.3	47%
EUR-A	40%	2.0	1%	40%	2.0	1%
EUR-B	79%	11.6	9%	85%	67.9	58%
EUR-C	82%	–0.8	0%	86%	38.5	33%
SEAR-B	71%	–0.7	0%	82%	39.9	33%
SEAR-D	84%	9.6	8%	95%	67.2	64%
WPR-A	56%	2.0	1%	56%	2.0	1%
WPR-B	83%	21.7	5%	87%	53.6	12%

proportion of all births that are unwanted. The whole subregion has an average total fertility rate (TFR) of 2, and we can therefore calculate that 75% of the subregional births occur in China (with a TFR of 1.8), all of them assumed as wanted. Of course, this assumption cannot be totally correct nor can it be verified, but it is likely to be a close approximation to the truth because of the strict birth control policies that have been applied in China since 1979.

The burden of disease attributable to unwanted births totals 4.6 million DALYs (see Table 15.19). In contrast to abortion, the largest part of the burden befalls women aged >30 years (75%). It is again the same subregions that are most affected, with those in Africa having the highest relative burden.

5. Sources of uncertainty

The calculation of aggregate level attributable fractions inevitably involves numerous uncertainties. One of the most serious concerns is the limited availability of data and the need to extrapolate results for a few countries to an entire subregion. Others relate to the quality of the input data, and to the method for simulating the expected numbers of

Table 15.19 Estimated burden of disease attributable to unwanted births (due to non-use of contraception), by subregion and age group in 2000

	Women 15–29 years				*Women 30–44 years*			
Subregion	*Attributable fractions (%)*	*Deaths*	*DALYs*	*DALYs per 1000 women*	*Attributable fractions (%)*	*Deaths*	*DALYs*	*DALYs per 1000 women*
AFR-D	2	938	53279	1318	21	7284	248587	11068
AFR-E	6	3527	188279	3927	31	15812	523605	19957
AMR-A	1	2	1092	34	1	2	481	13
AMR-B	19	1231	139381	2308	49	2512	152705	3338
AMR-D	23	814	60070	5884	54	1916	88018	13647
EMR-B	7	91	18566	946	43	835	56151	4839
EMR-D	10	2719	194874	4280	47	13663	513596	17849
EUR-A	1	1	760	19	1	1	479	10
EUR-B	9	86	20887	740	58	214	44212	1942
EUR-C	0	0	0	0	33	146	21774	782
SEAR-B	0	0	0	0	33	3827	138968	4368
SEAR-D	8	4491	369244	2265	64	36894	1443505	12314
WPR-A	1	0	396	25	1	1	270	17
WPR-B	5	446	86893	457	12	997	84396	469
World	7	14346	1133721	1488	40	84106	3316749	5335

unwanted births and unsafe abortions. Finally, the counterfactual estimates involved further assumptions.

While it was not possible, with one exception, to quantify the magnitude of the effect caused by the nature of the empirical data and assumptions, we can predict the direction in which they operate: whether they lead to an overestimation or an underestimation of the burden of maternal ill-health attributable to lack of effective contraceptive use. We have briefly reviewed the most important uncertainties and their likely effect on our estimates of relative risk and attributable fractions.

5.1 Robustness of data

Current contraceptive use

Survey data on current contraceptive use have been routinely collected by means of national surveys for 30 years. Their quality is considered high. Trends over time are plausible and the relationship between contraceptive prevalence and fertility rates is strong. Measurement error, where it exists, is likely to take the form of underreporting. In some societies, clandestine use by women occurs (Biddlecom and Fapohunda 1998; Castle et al. 1999) and this is likely to be concealed in conventional surveys. Moreover, some evidence exists to suggest that women underreport male methods, such as condoms, because of shyness and embarrassment (Koenig et al. 1984). Thus some users may be misclassified as non-users, and exposure may be slightly overestimated. The effect on relative risk estimates of such errors will be in the direction of underestimation but is likely to be small compared with other errors.

Contraceptive failure rates

Contraceptive failure rates were derived from enquiries of 18 DHS where detailed month-by-month information on contraceptive use episodes had been collected. Although this data source is undoubtedly the most appropriate, two types of uncertainty apply: the accuracy of the information and their representativeness. With regard to accuracy, the estimates (Table 15.2) are in general consistent with evidence from more carefully controlled prospective studies, with the exception of the condom where rates were lower than expected (Trussell 1998). However condoms are not a common method of contraception in most developing countries and any error is of minor significance. There are also doubts about representativeness. Calendar data are collected only in countries with high prevalence of use but the average results have been applied to all 58 countries. The validity of the underlying assumption—that failure rates are unrelated to overall levels of use—is unknown, but, again, the error is unlikely to be serious.

Induced abortion

In countries where abortion is illegal, or highly restricted and heavily stigmatized, it is impossible to obtain reliable information on incidence by means of conventional direct questioning. Hence, in this chapter, we have had to rely on WHO's indirect estimates of unsafe abortion (WHO 1998). While these estimates are widely accepted and cited at global level, a very considerable band of uncertainty surrounds them. However, no means exist of assessing the possible magnitude, or even direction, of error. In a few countries, the number of legal terminations could be taken into account when estimating the projected number of unwanted births. In yet other countries, legal terminations are carried out but no information was available on their number and therefore no allowance could be made. This gap in data leads to an overestimate of relative risks of unwanted births.

Fertility preferences

The method used to derive attributable and avoidable burden of disease depends heavily on women's statements about their future fertility desires or intentions. While experts agree that this way of measuring preferences is the least problematic of the several alternatives, interpretation is far from straightforward. For instance, attitudes toward future childbearing may be weakly held and ambivalent. Moreover, the attitude of the spouse, or male partner, is not taken into account. These considerations may partly explain why projected estimates of unwanted and mistimed pregnancies based on the pair of questions "Would you like to have a/another child or would you prefer not to have (more) children?" and "How long would you like to wait from now before the birth of a/another child?" are much larger than retrospective estimates of the "wantedness" of recent births and current pregnancies. While there is good reason—and some empirical evidence—to believe that the retrospective estimates are biased downwards by post facto rationalization, the size of the discrepancy between the prospective and retrospective estimates (63% and 143% for younger and older age groups, respectively) is a matter of concern. The direction of potential bias is clear. To the extent that preferences for future childbearing do not translate into unwanted and mistimed pregnancies, we will have overestimated relative risks as well as exposure.

Sexual exposure

The method of estimating attributable burden involved the exclusion of virgins from the calculations and the classification of women who report no sexual intercourse in the past 12 months as behaviourally not at risk of pregnancy. In countries with a strong traditional emphasis on pre-marital chastity for women, it is to be expected that single women will underreport sexual activity, which would lead to an underestimate

of expected birth among non-users, thereby underestimating relative risks.

In countries where DHS field staff interviewed ever-married women only, the single women were implicitly categorized as virgins. Countries with no data on single women are typically those where it would be socially unacceptable to ask young unmarried women about sex and reproduction. Although levels of sexual activity will no doubt be low in these countries, the resulting pregnancies are very likely to be unwanted. Most of them will be terminated, often clandestinely and thus most probably "unsafe". Insofar that the WHO estimates on unsafe abortion include procedures to unmarried girls (as estimates are based to a large extent on hospital admissions for abortion complications), these abortions have been attributed to the married women. However, since the exposure for single women is more skewed towards non-users we will have underestimated the attributable burden.

Fecundability

A major dilemma arose in the estimation of projected pregnancies among non-users of contraception. We had to choose biological estimates of the monthly probability of conceiving based on woman's age or on reported coital frequency. The reason for preferring age-based estimates was that they gave a closer fit between expected overall births and observed births in the recent past. The deficit in births was much bigger when estimates of fecundability were based on coital frequency, with expected births 39% and 11% lower than recent observed births for the 15–29 and 30–44 group, respectively. At low levels of intercourse, the impact of frequency of intercourse on fecundability is substantial with coitus-based conception rates well below the age-based rates, as shown in Table 15.4. The large shortfall in expected births compared with recent age-specific fertility does cast doubt on the reliability of sexual activity data, which appear to be too low to explain current fertility. Brown (2000) in his comparative study in Africa used the same Bongaarts and Potter model estimates of coitus-dependent fecundability with reported coital frequencies in the last month, and came to the same conclusion. However, in defence of the data he has shown good internal consistency between reported monthly frequency and time since last sex. Of course, at low frequencies, sex could be targeted to coincide with ovulation, increasing the probability of a pregnancy, although this would only affect women who desire a pregnancy soon and who are knowledgeable about the timing of ovulation.

The use of fecundability based on reported coital frequency would make a substantial difference to relative risks, reducing them by about a quarter. The relative risk of having an abortion for non-users would decrease from 8.7 to 6.7, while the relative risk of an unwanted birth would decrease from 7.8 to 5.7. This is explained by the fact women aged 30–44 years who say they want to have no more children have less

sex (two times a month) than those who want to space (three times) or those who desire a child within the next two years (3.9 times).

Whereas coital frequency among contraceptive users also varies with fertility intention, failure rates were kept constant regardless of preference. For modern methods, the most important determinant of failure is imperfect use. However, among perfect users, frequency of intercourse is the most important characteristic determining method failure (Trussell 1995). Traditional method failure is likely to be more dependent on coital frequency. This assumption, though far less important than the choice of age-based rather than coitally-based estimates of fecundability for non-users, will act to bias relative risks upwardly.

5.2 Assumptions in the basic model

In addition to concerns about the robustness of the empirical data, we had to make several assumptions in the basic model that links exposure to outcomes. The most important of these were:

- The burden is limited to direct obstetric events.
- Obstetric morbidity and mortality are the same for wanted and unwanted births.
- Abortion probabilities are zero for those who want another child in the next two years and are twice as high for limiters than spacers; probabilities are not affected by the proximate cause (failure vs non-use) and are the same for the two age groups.

Choice of outcomes

The crucial dilemma in defining the burden was whether or not to include perinatal mortality, much of which stems from unwanted pregnancies. Expert opinion was divided. The final decision to exclude perinatal mortality was based on the judgement that its inclusion would open up a Pandora's box of other intergenerational effects, going well beyond the perinatal period into infancy and childhood. Beyond the mortality of the unwanted children, short interbirth intervals are known to be a major risk factor for infant mortality and can be prevented by contraceptive use to cause better child spacing.

Obstetric burden is the same for wanted and unwanted births

The evidence base for judging whether the obstetric burden is the same for wanted and unwanted birth was meagre. To the extent that unwanted births are concentrated among older women of low socioeconomic status, it would have been justifiable to assume a higher risk. However, because births at late maternal ages constitute a small fraction of all births and because the link between socioeconomic status and unwanted childbearing varies between subregions, it was decided, by default, to

assume no difference. The effect of errors in this assumption would be to raise relative risks and the attributable burden.

Abortion probabilities

Several potential biases stem from assumptions that had to be made about the distribution of abortions by age group and exposure status. We made the simplifying assumption that no pregnancies occurring to women who report the desire for a child in the next two years are aborted. Because life circumstances change, this is no more than a close approximation to the truth and a small upward bias on relative risks is possible. A more important possible form of bias operating in the same direction is the assumption that abortion probabilities for unintended pregnancies are the same for non-users and for users who experience contraceptive failure. The available empirical evidence on this matter was insufficient to propose differing probabilities but it is nevertheless likely that modern method users do have a greater propensity to seek terminations than traditional method users and non-users, in which case relative risks would be overestimated. This bias may be offset to the extent that modern method users are more likely than others to seek safe abortions rather than unsafe, illicit abortions.

A bias operating in the opposite direction arises from the assumption that women who experience an unintended pregnancy when they want no more children are twice as likely to seek an abortion than those who wish to postpone the next pregnancy. Such a differential accords with common sense, at least for married women, and is consistent with the available shreds of evidence, but the size of the assumed difference is essentially arbitrary and may be too high. Because modern method users contain a disproportionately large number of limiters, relative risks may be underestimated.

The age pattern of abortion is known to vary between countries as a reflection of large differences in the proportion of young single women who are exposed to the risk of unintended pregnancy and differences in age at marriage. The simplifying assumption that abortion probabilities were constant by age may have led to an overestimation of abortion in the younger age group and an underestimation among older women. This will not affect the relative risks of unsafe abortion or unwanted birth by contraceptive use status but for the younger age group we may have underestimated the proportion of all births that are unwanted and thus the burden of obstetric complications.

5.3 Summary of uncertainty

Table 15.20 attempts to summarize in a necessarily crude manner the possible magnitude and direction of data defects and model assumptions on relative risks. A positive symbol (+) indicates that the effect may be to bias risks upwardly and a negative symbol (–) the opposite. A zero

Table 15.20 Possible effects of data limitations and assumptions on relative risks

	RRs Abortion	*RRs Unwanted births*
Robustness of empirical data		
• Contraceptive use	–	–
• Failure rates	–	–
• Induced abortion	0	+
• Fertility preferences	+++	+++
• Sexual exposure	––	––
• Fecundability	+++	+++
Assumptions in the model		
• Definition of burden	0	0
• Obstetric burden is same for unwanted and wanted births	––	––
• Abortion probabilities		
— No abortions among those who want birth soon	–	0
— Same for failure and non-use	++	0
— Twice as high for limiters than spacers	––	0
— No age pattern	0	0

(0) denotes that the direction of the uncertainty, or possible bias, cannot be established. The number of symbols represents our judgement on the magnitude of the possible bias. As may be seen, positive biases are broadly balanced by negative biases.

Quantifying a range around our estimates is not an easy task, and beyond the scope of this exercise. Varying fecundability by fertility desire alone could lead to a 25% reduction in relative risk levels. Allowing for other biases that work in the same direction we may set 25% as a minimum range of uncertainty around the estimates at country level. The extrapolation to subregional level may well introduce the biggest cause of uncertainty. Given these inherent limitations in the data and in the complexity of the various assumptions adopted to apply the methodology, the subregional estimates presented are approximate and reflect actual disease burden in general terms.

6. Discussion of attributable burden

The calculation of the burden of disease attributable to non-use of modern contraception methods has required a long and complicated series of steps, mainly arising from the fact that exposure has two dimensions: a behavioural one (use or non-use of contraception) and an attitudinal one (the desire to avoid or delay childbearing). Despite the inevitable degree of uncertainty surrounding estimates, some stemming

from inadequacies of empirical evidence and others from necessary assumptions, the key results make good intuitive sense and certainly provide a reliable basis for setting priorities at global and regional levels.

It is estimated that about 57 000 women die each year and that 4.9 million healthy life years (measured in DALYs) are lost because of abortions. Globally, about 90% of this burden is attributable to non-use of modern contraception. Regional differences in the attributable burden are strikingly large. In east and southern Africa (AFR-E), the estimated annual attributable burden exceeds 18 500 DALYs per 1000 women aged 15–29 years, and is also high in West Africa, South Asia and some Middle Eastern countries (AFR-D, SEAR-D and EMR-D). By comparison, it is under 40 per 1000 women aged 15–29 years in the industrialized low-fertility subregions AMR-A, EUR-A and WPR-A. Of course, one reason stems from differences in exposure: variations in the propensity of women who want to delay or avoid pregnancy to use modern contraception. But the more important reason concerns differences in access to legal and safe abortion services. Regions with high attributable burden of abortion-related mortality and morbidity are characterized by restrictive abortion laws, and vice versa. From a public health perspective, both issues (low contraception access and use and restrictive abortions) have important policy implications.

The magnitude of abortion-related mortality and morbidity is dwarfed by the obstetric burden stemming from complications of pregnancy and childbirth. It is estimated that about 415 000 women die each year from obstetric causes and that about 25 million healthy life years (measured in DALYs) are lost to these conditions. However, only a minority of these pregnancies are unwanted and hence the proportion of this overall disease burden attributable to non-use of modern methods is much lower than for the abortion-related burden: 7% among younger women rising to 40% in older women, among whom the desire to avoid all further childbearing is much more common. For both age groups combined, the estimates suggest that 98 000 obstetric deaths, representing nearly 20% of all such deaths, could be prevented each year if all women who desire no more children were to use modern contraceptives. The attributable burden is thus appreciably larger than the attributable abortion-related burden. Huge subregional differences are again apparent. In five subregions (AFR-D, AFR-E, AMR-D, EMR-D and SEAR-D) over 10 000 healthy life years are lost per year per 1000 women aged 30–44 years. The equivalent figure for the industrialized subregions (AMR-A, EUR-A and WPR-A) is below 20. In addition to access to and the use of modern contraception, these stark contrasts stem largely from variations in the coverage and quality of obstetric services.

7. Avoidable burden

There is little time lag between a change in contraception and the effect on burden of maternal complications. Current abortions and unwanted births are due primarily to non-use of contraception in the previous year. Thus, determining avoidable risk is very much like calculating attributable risk, but for "exposure in the future". Since fertility in today's medium- and high-fertility subregions is expected to drop in the future, the risk for each woman of death from an obstetric complication is also expected to decrease. However, the total burden of abortion-related complications and maternal outcomes may continue to increase in the next three decades, because the absolute number of women of reproductive age and the total number of births will continue to increase in the high-fertility subregions with the highest burden of maternal mortality.

In calculating counterfactual scenarios and attributable burden, the level of obstetric care and the quality of abortion services available were assumed to remain constant at current levels. Only the numbers of unwanted births and abortions determine the potential decrease in burden by uptake of contraception. However a reduction in unintended pregnancies is not the only pathway to lower levels of disease burden. In industrialized countries, there are still high levels of unintended pregnancies and abortions, but the disease burden associated with these is minimal because of the high quality of obstetric and abortion services. Indeed, the avoidable burden in absolute numbers may change more through a decline in the risk attributed to each pregnancy—by improvements in quality and provision of safe obstetric and abortion services—than through a decline in unintended pregnancies resulting from the use of effective contraception. It should be emphasized that improvement in risks related to abortions in many low-income countries requires above all a political will to change restrictive laws. Whatever the future may hold in terms of the risk attached to a single pregnancy or birth, it remains relevant and valid to estimate the proportion of the burden avoidable by increased effective use of contraception to avert unwanted births and abortions.

How should future attributable fractions be calculated and what are the necessary assumptions about fertility decline and levels of exposure? Fertility is expected to vary over the next three decades according to the UN medium-variant projections. Specifically all developing countries are now projected to reach replacement level fertility of 2.1 births per woman in the course of this century. Indeed the next UN projection will assume declines to 1.85 births per woman (UN 2002). These projections are rooted in evidence from the past 100 years that suggests that, once fertility has started to decline, it continues to fall until the achievement of low levels. This process of fertility transition appears to be relatively impervious to socioeconomic development. For instance, fertility has declined under conditions of rapidly improving standards of living, as in

many east Asian countries. It has also declined under conditions of economic stagnation or decline, as in Europe in the 1930s and much of east Africa in the past 20 years. Whatever the underlying forces of change, these fertility declines will be achieved primarily through increased levels of contraceptive use (and perhaps abortion), accompanied in some countries by rising age at marriage.

We thus need to project a future contraceptive distribution based on expected declines in fertility in the next three decades. Cross-country comparisons show that a fall in TFR of one child roughly corresponds to an increase in contraceptive prevalence of 15 percentage points (Ross and Frankenberg 1993). However, inferring contraceptive use from future fertility is complicated by the fact that abortion is an alternative means to regulate fertility. In countries experiencing simultaneous fertility decline and rapid changes in desired family sizes, unwanted births, abortion and contraception levels may all rise in parallel. This counterintuitive trend reflects the fact that in societies where couples want large families of, say, five or six children, exposure to the risk of an unwanted pregnancy is bound to be low. As fertility desires fall, the risk increases. For instance, in a society where couples want two children and women marry at 20 years of age, the desired family size will typically be achieved when the wife is in her mid-20s, leaving her exposed to the potential risk of unwanted pregnancy for the next 20 years. Thus it is not surprising that rapid declines in desired fertility can give rise to situations where increased contraceptive practice is unable to meet the growing need for fertility regulation.

One of the clearest examples is the Republic of Korea. As documented from longitudinal data in this country (Bongaarts and Westoff 2000), early on in the fertility transition, both levels of contraceptive use and the incidence of abortion rose in parallel, which in itself provides evidence for a growing unmet need for contraception. While abortion levels reached a peak and declined, contraceptive prevalence continued to rise, as the Republic of Korea progressed through the fertility transition.

The sequence of events in the Republic of Korea is not inevitable. In countries where effective contraception has not been promoted, and is thus relatively inaccessible, heavy reliance on abortion may persist. This is true in Japan and much of eastern Europe and central Asia (Henshaw et al. 1999). But in countries where abortion is very common, evidence suggests that improved availability of family planning services and wider choice of effective contraception can cause a rapid decline in abortion (Henshaw et al. 1999). This is the trend observed in central Asia, where abortion is being replaced by contraceptive use (Westoff et al. 1998). Thus widely varying patterns of change in population levels of contraception and abortion levels are evident in different populations (Marston and Cleland 2003b).

We therefore needed to make assumptions about the level of unmet need at future expected levels of TFR and contraceptive use. What

change in unmet need is to be expected from recent trends? There has been a steady increasing potential need for limiting births (adding the met need and unmet need), in the 1980s and the 1990s in developing countries (Westoff and Bankole 2000). Whereas the average proportion of women using contraception for limiting births increased faster than the potential need in Latin America, Asia and north Africa, this was not true for sub-Saharan Africa, where the proportion of women with an unmet need for limiting increased during the past two decades (Westoff and Bankole 2000).

As explained before, with rapid rates of social change, the need for contraception can grow faster than contraceptive use itself, resulting in rising incidence of unintended pregnancies, unwanted births and unsafe abortions. This implies that at the same levels of future fertility (TFR), populations could have different levels of unmet need, and that women will make trade-offs between abortion and contraception, depending on the legality, availability and perceived quality of services. However, the prediction of changes in unmet need (reductions in Asia and Latin America vs increases in Africa) made the calculation of avoidable risk too complex to operationalize.

The most effective and practical means of obtaining estimates of future exposure levels was to start from the relationship between fertility desires and TFR, and then to infer future contraceptive levels from the projected fertility desires. Fertility changes because of a decline in fertility desires and/or a better implementation of fertility preferences by couples. Future exposure to the risk factor considered here will depend on both. Therefore, the decrease in the maternal burden of disease (linked to a decline in pregnancies) can be split in two components: a reduction due to a lower desired family size and one due to a better implementation of fertility desires. Subregions with a high burden of maternal complications are going through a fertility transition and the desire for smaller families will continue to increase.

We have thus inferred the change in family size desires from the expected decline in TFR. Whereas the fertility preferences within each category of exposure will shift to higher proportions of women with an intention to stop or postpone childbearing, we can keep the propensity to translate desire into effective contraceptive protection constant. We have assumed that within each level of fertility intention (want children now, spacers, limiters), the relative distribution of non-users, traditional and modern method users stays the same, providing us with an estimate of future exposure. The future levels of contraceptive use implicit in the lower family size desires can be taken as the "business-as-usual" exposure. This reflects a rise in use expected from declines in desired family sizes, but is obviously dependent to a large extent on continued or increased investment in subsidized contraceptive services.

Based on this business-as-usual scenario, the avoidable burden then refers to the burden associated with unwanted births and unsafe abor-

tions that could be avoided through perfect implementation of future fertility preferences, over and above the burden that is avoided by the general trend towards lower fertility and the desire to have smaller family sizes. Simulating business-as-usual exposure levels for different time points in the future also requires associated business-as-usual relative risk levels and proportions unwanted among all births.

The step-by-step derivation of all the input data needed for the calculation of avoidable risk is given, followed by a discussion of the trends in input data. Avoidable risk was calculated for 2001, 2005, 2010, 2020 and 2030. For 2001, the 2000 input levels have been used.

7.1 Business-as-usual exposure and other inputs to calculate avoidable burden

Establishing the association between fertility desire and TFR

We have related the levels of fertility intention (want more children soon, spacing and limiting) aggregated across the three levels of exposure (modern methods, traditional methods, fecund non-users) to the TFR in the following way. Using the data from the 58 DHS, the cross-sectional linear associations between each of the current aggregate levels of fertility preference and the TFR were assumed to represent rates of change in the fertility desire for a one-unit change of TFR. Among the non-users, the proportions of women that are not currently exposed to the risk of pregnancy, either behaviourally or biologically, were also correlated to the TFR. With a fall in fertility there will also be a decline in aggregate levels of women who are pregnant and amenorrhoeic at any point in time; these were regressed as one category against the TFR. Amenorrhoeic women include those who abstain sexually after childbirth. Since both the numbers of women who did not have sex in the past 12 months (other than those who are also amenorrhoeic) and menopausal or infecund women are expected to change little with levels of TFR they were combined and then regressed together against the TFR. The regression coefficients were calculated separately for the 15–29- and 30–44-year age groups. Table 15.21 shows the results in terms of the slopes (rate of change along the regression line) and also the R-squared values.

Projected fertility decline

The expected decline in fertility was calculated from current TFRs and projected TFR levels using the UN projections (medium-variant) for each time point in future. Table 15.22 shows the subregional average TFR—current and projected—and the resulting projected decline at each time point. For 2001, the current levels were kept at calculated 2000 levels.

The sudden drop in fertility from current levels of 4.2 to 2.1 in 2005 in the EMR-B subregion is spurious and explained by the fact that

Table 15.21 Association of fertility intentions by TFR

	% change for 1 unit change in TFR: regression coefficients (R-squared values)				
Age group (years)	*Want birth soon*	*Want to space*	*Want to limit*	*Pregnant and amenorrhoeic*	*Infecund + no sex in past year*
15–29	1.42 (0.14)	–2.87 (0.24)	–6.56 (0.60)	7.16 (0.69)	0.85 (0.12)
30–44	2.38 (0.30)	1.34 (0.16)	–13.4 (0.73)	8.39 (0.83)	1.28 (0.09)

Table 15.22 Current and projected future levels of TFR, by subregion

		Projected future TFR UN median variant				*Projected change in TFR*			
Subregion	*Current TFR*	*2005*	*2010*	*2020*	*2030*	*2005*	*2010*	*2020*	*2030*
AFR-D	5.9	5.4	4.9	4.0	3.1	–0.5	–0.9	–1.9	–2.8
AFR-E	5.7	5.5	5.1	4.2	3.4	–0.2	–0.6	–1.5	–2.3
AMR-A	—	—	—	—	—	—	—	—	—
AMR-B	3.1	2.3	2.2	2.1	2.1	–0.8	–0.9	–0.9	–1.0
AMR-D	4.1	3.1	2.8	2.4	2.2	–0.9	–1.2	–1.7	–1.9
EMR-B	4.2	2.1	2.1	2.1	2.1	–2.1	–2.1	–2.1	–2.1
EMR-D	4.6	4.2	3.8	3.1	2.5	–0.4	–0.8	–1.5	–2.1
EUR-A	—	—	—	—	—	—	—	—	—
EUR-B	2.8	2.2	2.1	2.1	2.1	–0.6	–0.7	–0.7	–0.7
EUR-C	2.5	1.9	1.9	1.9	1.9	–0.6	–0.6	–0.6	–0.6
SEAR-B	2.7	2.1	2.0	2.0	2.0	–0.5	–0.6	–0.6	–0.6
SEAR-D	3.4	2.9	2.5	2.2	2.1	–0.5	–0.9	–1.2	–1.3
WPR-A	—	—	—	—	—	—	—	—	—
WPR-B	3.7	3.0	2.6	2.1	2.1	–0.7	–1.1	–1.6	–1.6

— No data.

Tunisia is the only country represented. Fertility has changed dramatically since the last Demographic and Health Survey in 1988.

Projected levels of fertility preferences

The calculation involved different steps, explained here in detail for women who want to limit family size, using Ghana for illustrative purposes.

1. *The current overall percentage of women who want to limit childbearing:* This was calculated from the distribution of fertility preferences within each level of exposure. For example, in Ghana, 11.4% of women aged 15–29 years are modern method users, 10.1% use tra-

ditional methods and 78.5% are not using any contraception at all. Among the modern method users, 11% want to limit their family size. So 1.2% (= 11.4% × 11%) of all women aged 15–29 years are modern method users with a desire to limit family size. Since 7% of traditional method users want no more children, 0.7% of all women are traditional method users with desire to limit family size. Among the non-users 47% are currently fecund and, among these, 6% want to stop childbearing. This gives 2.2% (= 78.5% × 47% × 6%) fecund non-users with a desire to limit family size. Adding all limiters together gives an aggregate percentage of 4.1% (= 1.2% + 0.7% + 2.2%) of all 15–29-year olds who want no more children.

2. *Projected decline in fertility*: Fertility in Ghana is projected by the UN to fall by 0.4 children from a current TFR of 4.4 to 4.0 in 2005.
3. *Future overall percentage of women who want to limit childbearing in 2005*: The projected fertility decline (0.4) was multiplied by the coefficient representing the change in the percentage of limiters (see Table 15.21: –6.56) and added to the current percentage (4.2%). So in 2005, 4.2 + 6.56 × 0.4 = 6.6% of 15–29-year olds are projected to want to limit their family size.
4. Steps 1 to 3 were repeated for the four other variables, using the appropriate coefficients in Table 15.21. By 2005, the percentage of women who want more children is projected to decline from 12.9% to 12.4% and women who want to space to increase from 41% to 42.1%. The percentage pregnant and amenorrhoeic would decline from 29% to 26.3%, while the percentage of infecund and not sexually active women would change from 13% to 12.7%.

Projected business-as-usual levels of contraceptive use

Having calculated the new overall distribution of fertility preferences in the population, we used the relative distribution of modern method users, traditional method users and fecund non-users within each level of fertility desire (current scenario), to estimate our business-as-usual exposure variable. For example, 29% of the limiters were modern method users, 17% used traditional methods and 54% were fecund non-users. Keeping the propensity to translate fertility intention into contraceptive practice constant, we have 29% of the 6.6% limiters aged 15–29 years, i.e. 1.9% using modern methods in 2005. Adding this to the 12% of the 12.4% women who want a child soon and the 21% of the aggregate 42.1% spacers who are using a modern method, we have derived an aggregate percentage of 12.3% modern method users in 2005 (= 1.9% + 1.4% + 8.9%). In total, 10.7% are traditional method users and 77.1% are non-users (this last category includes the 26.3% pregnant and amenorrhoeic and 12.7% infecund and not sexually active).

Other input for business-as-usual scenario

The relative distribution of fertility preference within each exposure level was calculated (e.g. 1.9 of the 12.3% or 16% of modern method users want no more children). The relative composition of the non-users in five categories of fecund, pregnant, amenorrhoeic, infecund/menopausal and no sex in the past year was also recalculated, and together with the projected contraceptive use levels used as input into the simulations. From this we derived the business-as-usual levels of the relative risk of having an abortion, the relative risk of having an unwanted birth as well as the proportions of unwanted among all births.

These calculations were done for each of the 58 countries at different time points (2005, 2010, 2020, 2030). For 2001, we used the current (2000) levels of contraceptive use and fertility desire. For each time point in the future we estimated five contraceptive prevalence distributions: business-as-usual exposure, theoretical minimum, and three other counterfactuals, shifting 10%, 20% and 30% of fecund non-users and traditional method users into the modern use category. For each of these we used the corresponding relative risks for abortions, unwanted births and proportion of unwanted births among all births.

7.2 Trends in business-as-usual exposure and relative risk levels

We have presented the trend in modern method use as derived through aggregating the country data into subregional estimates from 2001 (which was kept as the same as the 2000 level) to 2030. Table 15.23 shows these trends separately for the two age groups. For the three subregions with missing data, we kept the contraceptive distribution constant at the estimated 2000 level.

As expected from the fertility decline (Table 15.22), the contraceptive levels, reflecting the associated decline in fertility desire, are increasing steadily. As suggested by the regression coefficients (Table 15.21), the increase is more marked for the older age group. Since the expected fall in fertility by 2030 is largest for the two African subregions, the business-as-usual contraceptive levels are predicted to increase most steeply here.

In the business-as-usual scenario, the relative risk levels also varied from the current one, as shown in Tables 15.24 and 15.25 for unwanted births and abortions, respectively. The increase is more marked for non-users than it is for traditional method users. The increase is explained by the fact that the proportion of limiters increases among all three levels of exposure, but relatively more among the non-users.

Logically, as fertility desires are projected to go down with time, and keeping the propensity to translate desire into contraceptive behaviour at 2000 levels, the proportion of unwanted births among all births increases steadily, as can be seen from Table 15.26.

Table 15.23 Projected trends in the proportion of women using a modern method, business-as-usual scenario

	Women 15–29 years					*Women 30–44 years*				
Subregion	*2001*	*2005*	*2010*	*2020*	*2030*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	8.0	9.2	10.2	12.4	14.7	10.7	13.1	14.9	18.9	22.9
AFR-E	13.0	13.9	15.4	18.5	21.1	15.6	16.9	18.9	23.1	26.7
AMR-A	60.6	60.6	60.6	60.6	60.6	74.1	74.1	74.1	74.1	74.1
AMR-B	48.1	52.1	52.5	53.0	53.1	62.1	68.6	69.3	70.0	70.3
AMR-D	27.9	31.9	33.1	34.9	35.7	36.2	42.4	44.4	47.3	48.7
EMR-B	30.4	42.7	42.7	42.7	42.7	44.8	61.5	61.5	61.5	61.5
EMR-D	14.9	17.3	18.7	20.4	21.5	23.5	26.6	29.0	32.1	34.7
EUR-A	62.1	62.1	62.1	62.1	62.1	75.9	75.9	75.9	75.9	75.9
EUR-B	36.0	39.4	40.0	40.0	40.0	46.9	51.2	51.9	51.9	51.9
EUR-C	42.4	45.5	45.6	45.6	45.6	55.5	59.9	60.1	60.1	60.1
SEAR-B	54.4	58.5	59.1	59.1	59.1	55.9	60.9	61.7	61.8	61.7
SEAR-D	23.5	26.1	28.0	30.0	30.3	48.3	53.5	56.9	60.6	61.1
WPR-A	48.7	48.7	48.7	48.7	48.7	59.5	59.5	59.5	59.5	59.5
WPR-B	24.1	26.8	28.5	30.3	30.3	29.5	33.6	36.3	39.1	39.1

7.3 Business-as-usual and counterfactual scenarios

The main assumptions underlying the projected exposure and risk are as follows:

- Fertility in developing countries will fall in line with the UN median-variant projections.
- The cross-sectional relationship between fertility desires and fertility itself can be used to project future changes in fertility desires.
- In the business-as-usual scenario, the propensity to translate fertility desires into contraceptive use will remain unchanged.
- The relative popularity of different contraceptive methods and failure rates will remain unchanged.
- The proportion of infecund and sexually inactive non-users will remain constant.

Fertility trends

The UN medium-variant fertility projections are widely accepted as a reasonably dependable guide to the future. The UN's past record of successful projection over the short term of 20–30 years is impressive and no specific reason exists to doubt recent projections (Bongaarts and

Table 15.24 Projected trends in relative risk of having an unwanted birth, business-as-usual scenario

	Traditional method users 15–29 years					*Non-users 15–29 years*				
Subregion	*2001*	*2005*	*2010*	*2020*	*2030*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	2.3	2.3	2.3	2.4	2.4	2.8	3.3	3.8	4.8	6.0
AFR-E	1.6	1.6	1.7	1.8	1.9	3.0	3.2	3.7	4.9	6.0
AMR-A	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
AMR-B	3.2	3.3	3.3	3.3	3.3	8.3	10.4	10.6	10.9	11.0
AMR-D	4.1	4.1	4.2	4.2	4.2	6.2	7.9	8.5	9.4	9.9
EMR-B	3.2	3.5	3.5	3.5	3.5	3.9	7.2	7.2	7.2	7.2
EMR-D	3.2	3.2	3.3	3.4	3.4	3.4	4.0	4.6	5.4	6.1
EUR-A	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
EUR-B	4.0	4.1	4.1	4.1	4.1	5.1	6.3	6.5	6.5	6.5
EUR-C	5.1	5.1	5.1	5.1	5.1	6.9	8.1	8.1	8.1	8.1
SEAR-B	4.1	4.2	4.2	4.2	4.2	4.3	5.1	5.2	5.3	5.2
SEAR-D	4.6	4.9	5.0	5.1	5.1	4.6	5.4	5.9	6.6	6.7
WPR-A	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
WPR-B	3.7	3.7	3.7	3.7	3.7	5.8	6.9	7.8	8.8	8.8
	Traditional method users 30–44 years					*Non-users 30–44 years*				
	2001	*2005*	*2010*	*2020*	*2030*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	3.2	3.3	3.4	3.6	3.7	3.5	4.3	5.2	7.2	9.6
AFR-E	4.4	4.4	4.6	4.8	4.9	4.9	5.4	6.4	9.1	11.7
AMR-A	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
AMR-B	7.9	8.4	8.4	8.4	8.5	14.3	21.3	22.3	23.4	23.8
AMR-D	6.4	6.8	6.9	7.2	7.3	9.8	14.1	15.8	18.9	20.9
EMR-B	5.9	6.3	6.3	6.3	6.3	8.4	20.8	20.8	20.8	20.8
EMR-D	5.1	5.1	5.2	5.3	5.4	6.8	8.2	9.7	11.9	14.0
EUR-A	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
EUR-B	6.3	6.3	6.3	6.3	6.3	11.1	15.3	16.1	16.1	16.1
EUR-C	8.6	8.6	8.6	8.6	8.6	12.7	16.5	16.7	16.7	16.7
SEAR-B	6.1	6.3	6.3	6.3	6.3	8.0	10.0	10.3	10.4	10.4
SEAR-D	22.2	23.0	23.4	23.8	23.9	27.8	35.8	41.9	49.5	50.0
WPR-A	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
WPR-B	5.9	6.1	6.1	6.2	6.2	7.7	10.0	11.9	14.4	14.4

Table 15.25 Projected trends in relative risk of having an unsafe abortion, business-as-usual scenario

	Traditional method users 15–29 years					*Non-users 15–29 years*				
Subregion	*2001*	*2005*	*2010*	*2020*	*2030*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	2.6	2.6	2.7	2.6	2.6	3.9	4.3	4.8	5.8	7.1
AFR-E	3.0	3.2	3.0	2.9	2.9	4.7	5.0	5.4	6.7	8.0
AMR-A	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
AMR-B	4.1	4.2	4.2	4.2	4.2	10.6	13.1	13.4	13.7	13.8
AMR-D	4.3	4.4	4.4	4.5	4.5	7.6	9.5	10.2	11.2	11.7
EMR-B	4.6	4.7	4.7	4.7	4.7	6.4	10.7	10.7	10.7	10.7
EMR-D	3.9	3.9	3.9	4.0	4.0	5.0	5.7	6.3	7.3	8.1
EUR-A	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
EUR-B	4.9	4.9	4.9	4.9	4.9	7.5	9.1	9.3	9.3	9.3
EUR-C	5.0	5.0	5.0	5.0	5.0	10.5	12.1	12.1	12.1	12.1
SEAR-B	5.5	5.5	5.5	5.5	5.5	6.4	7.5	7.7	7.7	7.7
SEAR-D	6.6	6.8	6.9	7.0	7.0	10.9	12.5	13.5	14.7	14.8
WPR-A	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
WPR-B	4.1	4.1	4.1	4.1	4.1	6.6	7.8	8.8	9.9	9.9
	Traditional method users 30–44 years					*Non-users 30–44 years*				
	2001	*2005*	*2010*	*2020*	*2030*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	3.5	3.6	3.7	3.8	3.8	4.2	5.0	5.7	7.7	10.2
AFR-E	4.7	4.7	4.8	4.9	5.0	5.4	5.9	7.0	9.6	12.1
AMR-A	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
AMR-B	8.4	8.8	8.8	8.8	8.9	16.2	23.3	24.3	25.4	25.9
AMR-D	6.8	7.1	7.2	7.4	7.5	10.6	14.9	16.6	19.6	21.6
EMR-B	6.0	6.3	6.3	6.3	6.3	10.0	22.9	22.9	22.9	22.9
EMR-D	5.2	5.2	5.3	5.3	5.4	7.7	9.1	10.5	12.7	14.7
EUR-A	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
EUR-B	6.3	6.2	6.2	6.2	6.2	11.5	15.7	16.5	16.5	16.5
EUR-C	8.3	8.3	8.3	8.3	8.3	13.3	17.1	17.3	17.3	17.3
SEAR-B	6.4	6.6	6.6	6.6	6.6	9.2	11.2	11.6	11.7	11.6
SEAR-D	22.7	23.4	23.7	24.0	24.1	30.9	38.5	44.2	51.4	51.9
WPR-A	2.0	2.0	2.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0
WPR-B	6.2	6.3	6.4	6.4	6.4	8.5	10.9	12.9	15.3	15.3

Table 15.26 Projected trends in the proportion of births that are unwanted among all births, business-as-usual scenario

	Women 15–29 years					*Women 30–44 years*				
Subregion	*2001*	*2005*	*2010*	*2020*	*2030*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	2.4	5.8	8.7	15.0	20.6	23.7	30.8	37.6	51.1	62.3
AFR-E	6.8	7.2	9.8	15.9	20.9	34.6	35.2	42.3	56.7	67.5
AMR-A	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
AMR-B	23.5	26.5	26.8	27.1	27.2	58.0	67.5	68.5	69.4	69.8
AMR-D	27.3	32.0	33.4	35.2	36.1	61.7	73.5	77.0	81.9	84.2
EMR-B	11.7	20.2	20.2	20.2	20.2	53.9	78.0	78.0	78.0	78.0
EMR-D	11.9	13.6	16.3	20.4	23.7	53.3	58.3	64.9	74.4	81.5
EUR-A	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
EUR-B	18.6	21.0	21.3	21.3	21.3	72.3	78.9	79.9	79.9	79.9
EUR-C	6.0	7.9	8.0	8.0	8.0	47.6	54.9	55.3	55.3	55.3
SEAR-B	0.4	0.6	0.7	0.7	0.7	39.9	48.6	50.0	50.1	49.9
SEAR-D	9.6	11.1	12.1	13.3	13.5	67.2	78.5	86.1	94.6	95.3
WPR-A	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
WPR-B	21.7	25.7	28.1	30.5	30.5	53.6	63.7	70.1	76.3	76.3

Butatao 2000). In our view, fertility projections are therefore not a major source of uncertainty, nor is it possible to conjecture about possible departures from projections and their effect on our estimates.

Link between fertility and fertility desires

The relationship between achieved and desired fertility is not straightforward. Nevertheless, over the longer term, they tend to move in parallel. To represent this link we calculated cross-sectional correlations in the 58 study countries between fertility rates and fertility intentions and then assumed that these relationships would remain constant over the next 30 years. While we accept that a degree of uncertainty surrounds this procedure, we doubt whether it represents a serious bias.

Propensity to translate fertility desires into contraceptive practice

In the business-as-usual scenario it was assumed that the proportion of limiters and spacers who use contraception to achieve their intentions remains constant. This assumption was made in adherence to the business-as-usual concept that exposure changes with changing fertility desires, while the propensity to translate these into effective contraceptive use remains the same. Yet the empirical record of the past 40 years suggests that it is an artificial and unrealistic assumption. Fertility has declined in the past both because desired fertility has fallen and because

desires have been better implemented (Feyisetan and Casterline 2000). A more realistic representation of business-as-usual, of course, would act to reduce the avoidable burden.

Method mix and failure rates

Estimation of future scenarios used the assumption that choice of methods—or the method mix—within each country would remain the same, as with failure rates. Several forms of uncertainty may be identified: the development and uptake of newly developed forms of contraception; the possibility of a drift towards more effective existing methods; and greater resort to barrier methods in response to the HIV pandemic. The development and widespread use of radically new methods of contraception seems increasingly unlikely in view of lack of major investment by the pharmaceutical industry (Hagenfeldt 1994). The development of a contraceptive vaccine, for instance, seems an increasingly remote possibility. More plausibly, general shifts from less to more effective methods might occur—and indeed are underway in the countries of the former Soviet Union. With the exception of these latter countries, where in the past, access to modern contraception was severely restricted, little evidence exists to support the view that the contraceptive method mix will change over the next 30 years. Contraception in developing countries, unlike Europe or North America, has always been dominated by effective methods: sterilization, intrauterine devices and hormonal methods. In our view, an increased dominance of these methods is possible but not particularly likely. Any such trend would increase relative risks. Offsetting this might be increased uptake of condoms, the only existing contraceptive method that offers protection against HIV and other STIs. So far, no tendency towards greater use of condoms for family planning (and dual protection) within marriage has been recorded (UN 1999). With regard to contraceptive failure for users of specific methods, the assumption of no change is reasonably robust. We are unaware of any evidence of secular trends in the probability of failure.

Infecundity and sexual inactivity among non-users

A final assumption in the estimation of future scenarios was that the projection of infecund and sexually inactive non-users would remain constant. As the general health of adults improves, physiological infecundity is more likely to decline than increase, and prolonged sexual abstinence may become less common for the same reason and because of the erosion of customs of postpartum abstinence in sub-Saharan Africa. Any such trend would probably serve to increase relative risks and increase exposure and the avoidable burden.

Table 15.27 summarizes effects of assumptions on estimates of avoidable risk. A positive symbol (+) indicates that the effect may be to bias risks upwardly and a negative symbol (–) the opposite. A zero (0)

Table 15.27 Possible effects of assumptions on estimates of avoidable risk

Future fertility trends	0
Future fertility desires	0
Translation of desires into contraceptive use	++
Method mix and failure rates	0
Infecundity and sexual inactivity	–

denotes that the direction of the uncertainty, or possible bias, cannot be established.

Acknowledgements

We are obliged to Iqbal Shah and Elisabeth Aahman for access to the unpublished WHO estimates of unsafe abortion at country level. We are grateful to Mohammed Ali for giving access to the unpublished contraceptive failure rates. We have greatly benefited from the helpful comments by the five anonymous reviewers and the regular encouraging feedback from Majid Ezzati.

Note

1 See preface for an explanation of this term.

References

AbouZahr C, Vaughan JP (2000) Assessing the burden of sexual and reproductive ill-health: questions regarding the use of disability-adjusted life years. *Bulletin of the World Health Organization*, **78**:655–666.

Adetunji JA (1998) *Unintended childbearing in developing countries: levels, trends, and determinants*. Macro International, Calverton, MD.

Bankole A, Singh S, Haas T (1999) Characteristics of women who obtain induced abortion: a worldwide review. *International Family Planning Perspectives*, **25**:68–77.

Barrett G, Wellings K (2002) What is a "planned" pregnancy? Empirical data from a British study. *Social Science and Medicine*, **55**:545–557.

Becker S, Begum S (1994) Reliability study of reporting of days since last sexual intercourse in Matlab, Bangladesh. *Journal of Biosocial Science*, **26**:291–299.

Berkley S (1998) Unsafe sex as a risk factor. In: *Health dimensions of sex and reproduction: the global burden of sexually transmitted diseases, HIV, maternal conditions, perinatal disorders, and congenital anomalies*. Global Burden of Disease and Injury, Vol. 3. Murray CJL, Lopez AD, eds. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Biddlecom AE, Fapohunda BM (1998) Covert contraceptive use: prevalence, motivations and consequences. *Studies in Family Planning*, **29**:360–372.

Bongaarts J (1990) The measurement of wanted fertility. *Population Development Review*, **6**:487–506.

Bongaarts J, Butatao RA (2000) *Beyond six billion: forecasting the world's population*. National Academy Press, Washington DC.

Bongaarts J, Potter RG (1983) *Fertility, biology, and behaviour. an analysis of the proximate determinants*. Academic Press, New York.

Bongaarts J, Westoff CF (2000) The potential role of contraception in reducing abortion. *Studies in Family Planning*, **31**:193–202.

Brown MS (2000) Coitus, the proximate determinant of conception: intercountry variance in sub-Saharan Africa. *Journal of Biosocial Science*, **32**:145–159.

Castle S, Konaté MK, Ulin PR, Martin S (1999) Clandestine contraceptive use in urban Mali. *Studies in Family Planning*, **30**:231–248.

Conde Agudelo A, Belizan JM (2000) Maternal morbidity and mortality associated with interpregnancy interval: cross sectional study. *British Medical Journal*, **321**:1255–1259.

Curtis SL, Blanc A (1997) *Determinants of contraceptive failure, switching and discontinuation*: an analysis of DHS. Macro International, Calverton, MD.

Dixon Mueller R, Germain A (1992) Stalking the elusive "unmet need" for family planning. *Studies in Family Planning*, **23**:330–335.

Fortney JA (1987) The importance of family planning in reducing maternal mortality. *Studies in Family Planning*, **18**:109–114.

Freedman R, Freedman DS, Thornton AD (1980) Changes in fertility expectations and preferences between 1962 and 1977: their relation to final parity. *Demography*, **17**:365–378.

Ganatra B, Hirve S, Walawalkar S, Garda L, Rao V (2000) *Induced abortions in rural western Maharashtra: prevalence and patterns*. (Paper presented at Ford Foundation conference: Reproductive health in India: new evidence and issues, Feb 28–March 1, 2000, Pune, India.)

Hagenfeldt K (1994) Current status of contraceptive research and development. In: *Population—the complex reality*. Graham-Smith F, ed. Royal Society, London.

Henshaw SK (1998) Unintended pregnancy in the United States. *Family Planning Perspectives*, **30**:24–29, 46.

Henshaw SK, Singh S, Haas T (1999) Recent trends in abortion rates worldwide. *International Family Planning Perspectives*, **25**:44–48.

Henshaw SK, Singh S, Haas T (1999) The incidence of abortion worldwide. *International Family Planning Perspectives*, **25**:30–38.

James WH (1979) The causes of the decline in fecundability with age. *Social Biology*, **26**:330–334.

Johnson AM, Wadsworth J, Wellings K, Field J (1994) *Sexual attitudes and lifestyles*. Blackwell Scientific Publications, Oxford.

Koenig MA, Simmons GB, Misra BD (1984) Husband-wife inconsistencies in contraceptive use responses. *Population Studies*, **38**:281–298.

Leridon H (1977) *Human fertility: the basic components*. University of Chicago Press, Chicago, IL.

Leridon H (1993) La fréquence des rapports sexuels. Données et analyses de cohérence. *Population*, **5**:1381–1408.

Marston C, Cleland J (2003a) Do unintended pregnancies carried to term lead to adverse outcomes for mother and child? An assessment in five developing countries. *Population Studies*, **57**:77–94.

Marston C, Cleland J (2003b) Relationships between contraception and abortion: a review of the evidence. *International Family Planning Perspectives* **29**:6–13.

Mensch BS, Arends Kuenning M, Jain A, Garate MR (1995) *Meeting reproductive goals: the impact of the quality of family planning services on unintended pregnancy in Peru*. (Research Division Working Paper No. 81). The Population Council, New York.

Montgomery MR, Lloyd CB, Hewett PC, Heuveline P (1997) *The consequences of imperfect fertility control for children's survival, health, and schooling*. Macro International, Calverton, MD.

Potter RG, Millman SR (1986) Fecundability and the frequency of marital intercourse: new models incorporating the ageing of gametes. *Population Studies*, **40**:159–170.

Rahman A, Katzive L, Henshaw SK (1998) A global review of laws on induced abortion, 1985–1997. *International Family Planning Perspectives*, **24**:56–64.

Robey B, Ross J, Bhushan I (1996) Meeting unmet need: new strategies. *Population Reports, Series J, Family Planning Programs*, **43**:1–35.

Ronsmans C, Campbell O (1998) Short birth intervals don't kill women: evidence from Matlab, Bangladesh. *Studies in Family Planning*, **29**:282–290.

Ross JA, Frankenberg E (1993) *Findings from two decades of family planning research*. Population Council, New York.

The Alan Guttmacher Institute (1999) *Sharing responsibility: women, society and abortion worldwide*. The Alan Guttmacher Institute, New York.

Trussell J (1995) Contraceptive efficacy. *Archives of Dermatology*, **131**: 1064–1068.

Trussell J (1998) Contraceptive efficacy. In: *Contraceptive technology,* 17th revised edn. Hatcher R, et al., eds. Ardent Media, New York.

Trussell J, Pebley AR (1984) The potential impact of changes in fertility on infant, child, and maternal mortality. *Studies in Family Planning*, **15**(6/1):267–280.

UN (1999) *Levels and trends of contraceptive use as assessed in 1998*. United Nations Department of Economic and Social Affairs, Population Division, New York.

UN (2002) *United Nations Expert Group meeting on completing the fertility transition*. United Nations Department of Economic and Social Affairs, Population Division, New York.

Weinstein M, Wood J, Chang M (1993) Age patterns of fecundability. In: *Biomedical and demographic determinants of reproduction*. Gray R, Leridon H, Spira A, eds. Clarendon Press, Oxford.

Weinstein M, Wood JW, Stoto MA, Greenfield DD (1990) Components of age-specific fecundability. *Population Studies*, **44**:447–467.

Westoff CF (1981) Unwanted fertility in six developing countries. *International Family Planning Perspectives*, **7**:43–51.

Westoff CF (1990) Reproductive preferences and fertility rates. *International Family Planning Perspectives*, **16**:84–89.

Westoff CF (2000) *The substitution of contraception for abortion in Kazakhstan in the 1990s*. Macro International, Calverton, MD.

Westoff CF, Bankole A (1998) The time dynamics of unmet need: an example from Morocco. *International Family Planning Perspectives*, **24**:12–14.

Westoff CF, Bankole A (2000) Trends in the demand for family limitation in developing countries. *International Family Planning Perspectives*, **26**:56–62.

Westoff CF, Sharmanov AT, Sullivan M, Croft T (1998) *Replacement of abortion by contraception in three Central Asian Republics*. The Policy Project and Macro International, Calverton, MD.

WHO (1998) *Unsafe abortion. Global and regional estimates of incidence of and mortality due to unsafe abortion with a listing of available country data*. (WHO/RHT/MSM/97.16) World Health Organization, Division of Reproductive Health, Geneva.

Winikoff B, Sullivan M (1987) Assessing the role of family planning in reducing maternal mortality. *Studies in Family Planning*, **18**:128–143.

Wood JW, Weinstein M (1988) A model of age-specific fecundability. *Population Studies*, **42**:85–113.

Chapter 16

Unsafe water, sanitation and hygiene

Annette Prüss-Üstün, David Kay, Lorna Fewtrell and Jamie Bartram

Summary

The disease burden from unsafe water, sanitation and hygiene (WSH) is estimated at the global level taking into account various disease outcomes, principally diarrhoeal diseases. The risk factor is defined as including multiple factors, namely the ingestion of unsafe water, lack of water linked to inadequate hygiene, poor personal and domestic hygiene and agricultural practices, contact with unsafe water, and inadequate development and management of water resources or water systems.

For estimating disease burden of infectious diarrhoea, exposure scenarios are established according to water supply and sanitation infrastructure, the level of faecal–oral pathogens in the environment and populations assigned to these scenarios. The total burdens from schistosomiasis, trachoma, ascariasis, trichuriasis and hookworm disease are all wholly attributable to unsafe WSH and have been quantified at global level as an additional exercise.

Unsafe WSH is an important determinant in a number of additional diseases, such as malaria, yellow fever, filariasis, dengue, hepatitis A and hepatitis E, typhoid fever, arsenicosis, fluorosis and legionellosis, some of which present a high disease burden at global level.

For infectious diarrhoea, six exposure levels were defined, with the lowest risk level corresponding to an ideal situation where WSH plays no role in disease transmission. Exposure prevalence, in terms of infrastructure, was determined from the Global Water Supply and Sanitation Assessment 2000. This assessment is a synthesis of major international surveys and national census reports covering 89% of the global population. The parameters considered included access to improved water sources and improved sanitation facilities.

Relative risk estimates were based on reviews and large multi-country studies for areas with high faecal–oral pathogen loads in the environment (i.e. principally in developing countries). The proportion of disease

due to unsafe WSH in regions with low faecal–oral pathogen loads was based on a study analysing the relative importance of etiological agents causing diarrhoeal diseases, supported by evidence from selected studies considered to be of high quality. A low faecal–oral pathogen load in the environment was assumed if sanitation coverage exceeded 98% (which corresponds to the situation in most developed regions).

For the high faecal–oral pathogen exposure group, Esrey's multi-country study (1996) suggests that a mean reduction in diarrhoea of 37.5% is possible following the introduction of improved water supply and sanitation in developing country environments. For the low faecal–oral pathogen exposure group, data from the study by Mead et al. (1999) suggested that the proportion of diarrhoeal illness attributable to food in the United States of America was approximately 35% (excluding those illnesses wholly transmitted by food). We have therefore estimated that approximately 60% was attributable to unsafe WSH. A review by Huttly et al. (1997) of epidemiological studies on hygiene practices in seven nations identified a median reduction of diarrhoea incidence of 35%.

Selected additional studies have suggested ranges of reductions in diarrhoea incidence that could be achieved by reducing the transmission of faecal–oral pathogens through the implementation of interventions, such as point of use treatment and disinfection of stored water (Quick et al. 1999; Semenza et al. 1998). However, this transition has been poorly documented by exposure–risk information, and we considered it appropriate to examine both optimistic and pessimistic estimates in defining the uncertainty around these values.

The disease burden from unsafe WSH was estimated to have been 1.73 million deaths in the year 2000, and 88% of the global burden of diarrhoeal disease due to infectious diarrhoeal diseases. In addition, schistosomiasis, trachoma, ascariasis, trichuriasis and hookworm disease are fully attributable to WSH-related factors. Typically, the fraction of diarrhoeal disease attributed to unsafe WSH in developed countries is approximately 60%, whereas in developing countries as much as 85–90% of diarrhoeal illness can be attributed to unsafe WSH. The major part is borne by children in developing countries.

This estimation of the global disease burden caused by unsafe WSH suggests a significant burden of preventable disease attributable to this cause in developing nations, and a non-negligible burden in developed countries.

1. Introduction

The disease burden caused by the risk factor unsafe WSH was estimated at the global level in 1990 (Murray and Lopez 1996a). This original estimate examined WSH in terms of diarrhoeal and selected parasitic diseases, based on the partial attribution of their disease burden to the

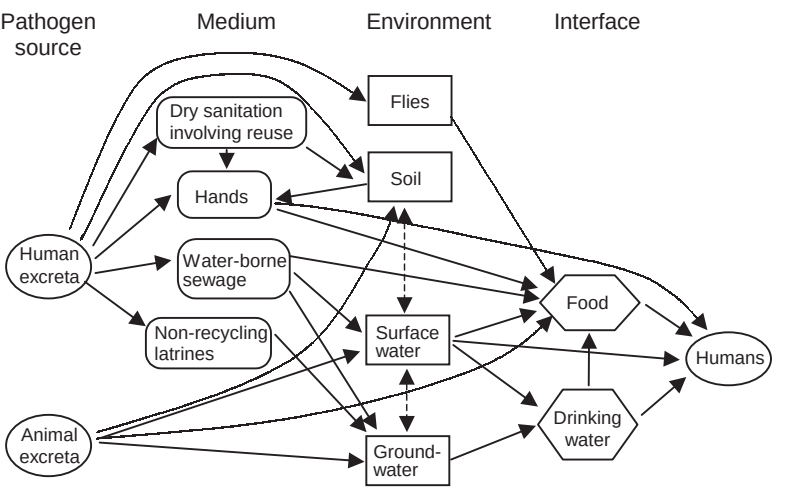
risk factor. It was found that worldwide the risk factor accounted for 5.3% of all deaths and 6.8% of all disability-adjusted life years (DALYs). Other communicable (e.g. hepatitis A and E, malaria) and noncommunicable diseases (arsenicosis, fluorosis, methaemoglobinaemia) were not considered in that assessment.

1.1 Rationale for a composite risk factor

Faecal–oral diseases account for the dominant health outcome of the unsafe WSH risk factor and are the main focus of this chapter. However, not all of them could be included in this estimate (e.g. hepatitis A and E). For infectious diarrhoea, the unsafe WSH risk factor comprises a number of transmission routes mediated by a complex interaction of infrastructure issues, which might affect, for example, microbiological hazards from poor quality drinking water, water availability, microbial risks from inappropriate disposal of faecal wastes and behavioural aspects. The transmission routes interact with the efficiency of interventions such as hygiene within the home, hand-washing and rigorous application of point-of-use treatment within domestic properties. Clearly, in any global assessment, the contribution of each element, together with the plethora of interactions, cannot be precisely quantified in every setting. However, as hazard estimates come from large surveys performed in several countries (Esrey 1996), variations in behaviour and their effects on the transmission of faecal–oral pathogens have, to some extent, been internalized in our estimates.

It is likely that the relationship between faecal–oral pathogen dose and the probability of infection is log-linear for many of the infectious diarrhoeal diseases, reaching a plateau for higher exposures (Briscoe 1984; VanDerslice and Briscoe 1995). Sometimes several component causes (see Figure 16.1) may produce similar infection outcomes. This can mean that the introduction of a single intervention in isolation (e.g. the provision of cleaner water supplies) designed to break an infection pathway may result in a negligible reduction in overall disease burden. This is particularly true in communities where the environmental load of faecal–oral pathogens is high (e.g. a community with low sanitation coverage, faecally contaminated drinking-water supplies, irregular refuse collection and poor hygiene practices). This renders the attribution of a disease fraction to a specific factor particularly difficult and, indeed, potentially misleading to the policy-making community. For this reason, it is necessary to consider WSH as interrelated parts of a single causal web in which cutting one major pathway of transmission may well show no (or minimal) effect on the total disease burden, but may, in other circumstances, provoke a dramatic response. Importantly, however, removing a basic pathway (e.g. by providing safe drinking water or improved sanitation) is likely to be a precondition for the success of subsequent interventions to reduce disease burden.

Figure 16.1 Transmission pathways of faecal–oral disease



Management actions concerning water supply and sanitation often involve water resource management, including the control of insect vectors of disease (such as malaria) and soil-borne helminths (such as ascaris). Similarly, environmental management to control disease vectors impacts directly upon water supply and sanitation. Furthermore, access to improved water sources has a significant impact on exposure to agents of some water-based diseases (such as schistosomiasis) and diseases with water-related insect vectors, and improved sanitation reduces certain vector-borne diseases such as trachoma. These intimate interconnections of exposure pathways and control mechanisms suggest that treating water, including supply and resource management, as an integral part of the risk factor unsafe WSH is rational.

1.2 Definition of risk factors

Unsafe WSH adversely affects health through multiple routes.

1. Transmission through contact with water that contains organisms such as *Schistosoma spp*.
2. Transmission through vectors proliferating in water ecologies related to dams, irrigation schemes and other water resources projects (e.g. malaria, schistosomiasis, lymphatic filariasis). This should be included although it is currently unclear how or whether it can be quantified.

3. Transmission through the ingestion of water as it occurs during drinking and, to some extent, bathing. This category includes diseases from faecal–oral pathogens, dracunculiasis, arsenicosis, fluorosis, from other toxic chemicals and due to excess proliferation of toxic algae.
4. Transmission caused by poor personal, domestic or agricultural practices, including when personal hygiene is affected by lack of water. This includes person-to-person transmission of faecal–oral pathogens, foodborne transmission of faecal–oral pathogens as a result of poor hygiene or use of contaminated water for irrigation or cleaning. Lack of water is in particular linked to diseases such as trachoma and scabies.
5. Transmission through contaminated aerosols from poorly managed water systems (e.g. legionellosis).

Water-related injuries that could be prevented by appropriate water management were not considered in the current estimate because of the different management approaches to their remediation. They were, however, covered by the World Health Organization (WHO 1998), although their disease burden was not quantified. Many social, geographic and behavioural factors, such as hygiene, the domestic storage and potential contamination of potable water, the use of sanitation facilities, etc. are important determinants of health outcome. This set of factors has complex social and behavioural drivers that are highly heterogeneous both within and between nations. In reality, they would modify the effects of the pathways defined in (i) to (v) above. It is beyond the scope of the current assessment to attempt to quantify the unique impacts of this set of factors in each setting.

Diseases relating to unsafe WSH, and their inclusion in the current estimate, are listed in Table 16.1. This first assessment of disease burden should be considered an initial estimate, which will benefit from refinement as additional information becomes available. Table 16.1 is not exhaustive, as the linkages between water and health are extensive and complex. For example, it is likely that the role of inadequate water for food production, and therefore nutrition, will be particularly important, in addition to the direct impact of infectious diarrhoea on nutrition.

1.3 Evidence of causality on infectious diarrhoea

As illustrated in Table 16.1, numerous separate faecal–oral illnesses fall under the "umbrella" of infectious diarrhoea. Their commonality derives from their mode of transmission, in that the source of the pathogen is human (or less commonly, animal) faeces which can cause infection in a new host upon ingestion. The shortest route of transmission is from person-to-person (a hygiene issue), while longer routes include transfer of pathogens to a food crop, as well as to drinking water or recreational

Table 16.1 Diseases related to unsafe water, sanitation and/or hygiene

Disease outcome	*Included in current estimate*
Infectious diarrhoea, including: cholera, salmonellosis, shigellosis, amoebiasis, other bacterial, protozoal and viral intestinal diseases[a]	Yes (acute effects only)
Typhoid and paratyphoid fevers	Partly included in estimate for infectious diarrhoea, but would benefit from separate, more precise consideration
Hepatitis A	No
Hepatitis E	No
Fluorosis	No
Arsenicosis	No
Legionellosis	No
Methaemoglobinaemia	No
Schistosomiasis[a,b]	Yes
Trachoma[a,b]	Yes
Ascariasis[a,b]	Yes
Trichuriasis[a,b]	Yes
Hookworm[a,b]	Yes
Dracunculiasis[b]	No (disease close to eradication)
Scabies	No
Dengue[a]	No
Filariasis[a]	No
Malaria[a]	No
Japanese encephalitis[a]	No
Onchocerciasis[a]	No
Yellow fever	No
Impetigo	No
Drowning[a]	No

[a] Included in this analysis.

[b] Considered to be 100% due to unsafe WSH.

water, as summarized in Figure 16.1. The predominant route will depend upon the survival characteristics of the pathogen as well as local infrastructure and human behaviour. While some of the diseases contained in the group diarrhoeal disease, as defined for the purpose of this project, are relatively mild and self-limited, others may be more severe and cause long-lasting sequelae (Hunter 1997). The disease burden based on these studies has not been taken into account in this estimate.

The fact that faecal–oral pathogens can be spread via the water route is well established (Andersson and Bohan 2001; Esrey et al. 1991; Hunter 1997; Snow 1855). The following sections briefly outline the evidence

for infectious diarrhoea causality in relation to water, sanitation and hygiene. For the most part, studies examining the issue have been intervention studies, which have looked at changes in water supply, excreta disposal or hygiene practices, and assessed the effects on diarrhoea morbidity or mortality rates (generally in young children). Another significant group of investigations comprise case–control studies, particularly following outbreaks suspected to be caused by potable water contamination in developed nations.

Sanitation

Ideally, sanitation (i.e. human excreta management) should result in the isolation or destruction of pathogenic material and, hence, a break in the transmission pathway. In a comprehensive literature review, Esrey et al. (1991) identified 30 studies, from a variety of different countries (including Bangladesh, Brazil, Chile, Guatemala, Kenya, Malaysia and Panama), that examined the impact of sanitation on disease transmission. Twenty-one of those studies reported health improvements (median 22% reduction in diarrhoea morbidity), with a greater median reduction being seen in the rigorous studies (36% reduction). Several studies have isolated various faecal–oral pathogens from the faeces of sick people and the transmission of such pathogens isolated from infected faeces to human hosts has been shown in numerous studies (e.g. for *Shigella* [Dupont et al. 1989]). Clearly, the relationship is both plausible and coherent.

Water

The number of outbreaks of infectious diarrhoea caused by faecal–oral pathogens in developed countries attests to the efficiency of this mode of transmission. In the United States, for example, 14 outbreaks of infectious etiology associated with drinking water were reported for the two-year period 1997–1998 (Barwick et al. 2000).

In developing countries, it is not only water contaminated at source or during distribution that is an issue, but water stored within the home which may also become contaminated (arguably a hygiene issue). For example, in a literature review, VanDerslice and Briscoe (1993) found 11 observational studies showing that mean coliform levels (an indicator of contamination) were considerably higher in household water containers than in the original source waters.

Numerous epidemiological studies and outbreak investigations have found an association between poor water quality and infectious diarrhoea. In France, water that did not meet microbiological standards was associated with an increased risk of gastroenteritis (RR 1.36, CI 1.24–1.49) (Ferley et al. 1986). In the Philippines, Moe et al. (1991) reported an odds ratio (OR) of 1.92 (CI 1.27–2.91) for diarrhoea following consumption of water contaminated with high levels of *Escherichia coli* (a faecal indicator bacteria). Mahalanabis et al. (1991)

reported that children with prolonged diarrhoeal illness (more than 14 days) were more likely to drink water from an unprotected water source (OR 1.56, CI 1.18–2.06). Birmingham et al. (1997) conducted an epidemiological investigation to identify sources of infection and risk factors for cholera in Burundi during an epidemic in 1992. Water from Lake Tanganyika was implicated, as a case–control study found that both bathing in the lake (OR 1.6, CI 1.1–2.1) and drinking its water (OR 2.78, CI 1.0–7.5) were independently related to illness; additionally *Vibrio cholerae O1* was isolated from the lake water.

As seen above, the causal relationship between ingesting water of poor sanitary quality and diarrhoeal illness has been observed worldwide, using a variety of techniques and assessing quality in a number of different ways. The biological gradient can be illustrated by increases in infectious diarrhoea morbidity as contamination levels increase, and also as consumption of water from a single contaminated source increases. For example, Njemanze et al. (1999) examined the annual diarrhoeal incidence rate (per 1000 population) in 39 communities in Imo State, Nigeria, in relation to the characteristics (including pollution) of their drinking water source. Sources were classified from A to C with A representing the most desirable sources (with favourable geology, sparse population and clean and unpolluted water). Diarrhoeal incidence rate was found to show a statistically significant increase with a mean of 1.61 for category A, a mean of 6.25 for category B, and a mean of 15.6 for category C.

The relationship between infectious diarrhoea and transmission of pathogens through water is both plausible and coherent. Isolation and enumeration of specific pathogens in water are often not feasible or very imprecise; thus a more common measure of faecal contamination is derived from the use of indicator bacteria. There have been many studies using such indicator species that have demonstrated the faecal contamination of drinking water sources in both developed and developing countries (e.g. Ampofo 1997).

Hygiene

A number of studies have attempted to examine the role of personal and domestic hygiene, although in many cases some of the "hygiene" measures or interventions could also impact on sanitation, and hygiene interventions may also interact with water quality.

Six studies examined by Esrey et al. (1991) identified reductions in diarrhoea morbidity associated with the uptake of hygiene interventions. These ranged from 14% to 48%, with a median reduction of 33%. In a more recent review, Huttly et al. (1997) identified a further four studies addressing the impact of improved hygiene. All four studies showed a decrease in diarrhoea, as did a subsequent study of Curtis et al. (2000). These studies were conducted in diverse locations including Bangladesh, Burma, Guatemala and the United States.

The temporal adoption of hygiene measures can be illustrated by the study by Ahmed et al. (1993). This group compared cleanliness and diarrhoea levels in villages with and without hygiene education interventions. Higher adoption rates of the intervention were associated with a better cleanliness state, which was paralleled by a decrease in diarrhoea and malnutrition rates. These differences were found to increase over time as more villagers adopted the intervention.

Alam et al. (1989) studied the effect of four different hygiene measures (source of washing water; presence of faeces in the yard; hand-washing before serving food; and hand-washing after defecation). They showed decreasing diarrhoea incidence as the number of adopted hygienic practices increased (4.9 cases per child-year for one practice to 2.6 cases for all four; $P<0.01$).

A review by Feachem (1984) documented the presence of pathogens on the hands following toilet activities. In the same review, Feachem also noted a number of studies on hand-washing which demonstrated the almost complete removal (98–100%) of seeded bacteria.

1.4 Evidence of causality on other outcomes

Schistosomiasis

Schistosomiasis is caused by infection with trematodes of the *Schistosoma* species. Transmission of the disease occurs when people come into contact with water containing cercariae (the mobile larval stage of the life cycle), which penetrate the skin. Water is contaminated by infected humans who excrete the schistosome eggs in their faeces or urine (depending upon the *Schistosoma* species). The final link in the chain of infection is provided by an intermediate snail host, which the parasite needs in order to complete its life cycle. Current knowledge on disease transmission indicates that the disease is fully attributable to unsafe WSH.

Esrey et al. (1991) identified 12 studies that related water and sanitation facilities to the rates of schistosomiasis. Reported decreases in infection rates varied between 59% and 87%, with the median value of the rigorous studies being a 77% reduction. Numerous studies, in addition to those identified above, have noted the relationship between contact with contaminated water and high levels of infection with schistosomiasis (Hunter 1997). These have been conducted in various countries and have examined different *Schistosoma* species. Lima e Costa et al. (1991) found that individuals reporting water contact less than once a week had a smaller excess risk of schistosomiasis than those reporting water contact at least weekly (OR 3.0, CI 1.3–6.6 in comparison to OR 4.3, CI 2.6–7.0).

A number of studies have examined reinfection with schistosomiasis following an intervention programme (such as treatment of infected individuals). In China, Zhaowu et al. (1993) found that reinfection was

associated with the frequency of water contact, the type of water contact and the proximity of residence to snail-infected water. In Brazil, discontinuation of a control programme led to an increased prevalence of schistosomiasis (Coura-Filho et al. 1994). Risk factors for the disease included any form of water contact (OR 2.79, CI 1.19–6.85).

The relationship is plausible and the results of numerous studies are coherent and do not conflict with what is known about the disease. Interventions centring on water and sanitation provision designed to either decrease water contamination or decrease contact with contaminated water have proved to be effective in reducing the rates of schistosomiasis (e.g. Barbosa et al. 1971; Jordan 1972).

Trachoma

Trachoma is a chronic contagious eye disease, which can result in blindness, caused by *Chlamydia trachomatis*. Transmission occurs by several routes (Dolin et al. 1997), all of which are hygiene related (e.g. direct infection by flies, person-to-person from clothing used to wipe children's faces and by hand-to-face contact). Risk factors for the disease include lack of facial cleanliness, poor access to water supplies, lack of latrines and a high number of flies.

A total of 16 studies were identified by Esrey et al. (1991) which examined the role of WSH on the level of trachoma. The median reduction in trachoma was 50% (0–91) from all the studies and 27% (0–9) when considering the rigorous studies. More recently Prüss and Mariotti (2000) identified 39 studies which examined the level of trachoma in relation to environmental causation; they report that relative risks ranged between 1 and 4. Thirteen of the 16 studies identified by Esrey et al. (1991) reported positive effects, i.e. a water, sanitation or hygiene intervention resulted in lower levels of trachoma. The studies were conducted in a variety of locations including Australia, China, India, Mexico, Mozambique, the Sudan and Tunisia.

Prüss and Mariotti (2000) reported that the biological gradient was verified in most of the studies in which it was investigated, although they also noted that few studies examined this issue. Preventative measures through hygiene education and interventions aimed at reducing fly numbers have both resulted in decreases in trachoma (Emerson et al. 1999; Sutter and Ballard 1983).

Ascariasis

Ascariasis is caused by the large roundworm *Ascaris lumbricoides*. Eggs are passed in the faeces of an infected person and in poor sanitation conditions may contaminate the soil. Ingestion of infective eggs, from contaminated soil or from uncooked products contaminated with soil or wastewater containing infective eggs, cause the disease. Transmission does not occur from person to person. The knowledge on transmission pathway indicates that the disease is fully attributable to unsafe WSH.

The eggs can survive for months or years in favourable conditions and can, thus, pose an infective hazard for a considerable period of time.

A total of 14 studies examining the level of ascariasis and water and sanitation provision were identified by Esrey et al. (1991). These studies reported reductions between 0–83%, with a median reduction from all the studies of 28%. More recently, Cifuentes (1998) reported big differences in infection between children exposed to untreated wastewater and those exposed to either partially treated wastewater or rainwater irrigation (OR 5.71–13.18, depending upon the age group under consideration). Similar results were reported by Habbari et al. (2000), who showed that *Ascaris* infection was five times higher in children in the wastewater impacted regions compared to control regions. In Indonesia, Toma et al. (1999) reported a 64% reduction in *Ascaris* infection in people who used a latrine compared with those who did not.

A biological gradient is suggested from the results of the four rigorous studies identified by Esrey et al. (1991) where the rate of morbidity reduction was dependent upon the level of sanitation facility. The work of Cifuentes (1998) also indicates a dose–response relationship with children exposed to increasingly contaminated water having increased rates of infection.

The relationship is plausible and the study results are coherent. Eggs have been isolated from faecal samples, soil samples, water samples and hand-washing samples (Jonnalagadda and Bhat 1995). Additional experimental evidence is provided by the studies that have examined the increased use of latrines and noted the parallel decrease in both egg counts in soil and levels of infection (e.g. Arfaa et al. 1977).

Trichuriasis

Trichuriasis is caused by ingestion of the human infectious eggs of the whipworm *Trichuris trichiura*. The infection is not directly transmissible from person to person. As with other faecal–oral transmitted diseases, the mode of transmission indicates that the disease is fully attributable to unsafe WSH, although the risk factors for trichuriasis in relation to WSH do not seem to have been as well researched as the other illnesses covered here. Studies of prevalence often show an association between *Ascaris* and *Trichuris* infection (Anderson et al. 1993; Saldiva et al. 1999; Smith et al. 2001), suggesting similar modes of transmission.

Of the studies that were identified, Henry (1981) found that *Trichuris* infections decreased by 50% after water supplies and latrines were installed in a rural area of Saint Lucia. Rajeswari et al. (1994) noted that the prevalence of infection was associated with a number of factors, including socioeconomic status, water supply, sanitary disposal of faeces and family size. Similarly, Narain et al. (2000) found that open field defecation and large family size were independently associated with *Trichuris* infection.

Hookworm disease

Hookworm infection is caused by *Ancylostoma duodenale* or *Necator americanu*s, and results from the ingestion or skin penetration of the hookworm larvae that live in the soil. Larvae develop in the soil through the deposit of faeces containing eggs from infected persons. The disease is therefore caused by poor sanitation and hygiene practices. The disease is not transmitted from person to person.

Eleven studies were identified by Esrey et al. (1991) which examined water, sanitation and hookworm infection. From the nine that could be used to calculate a reduction in morbidity, the range was 0–100%, although only one of these was considered to be rigorous. Sorensen et al. (1994) found that the severity of hookworm infection was lower in children coming from communities with good sanitary facilities.

Norhayati et al. (1995) studied the reinfection of children in a hookworm endemic area. In the absence of any interventions the reinfection rate at 4-months post-treatment was 30%. The authors suggested that long-term strategies incorporating education on personal hygiene, provision of toilets and safe water supply were required to control the rapid reinfection. Humphries et al. (1997) reported that hookworm egg counts were significantly higher in Vietnamese women who used fresh human faeces as a fertilizer in comparison to those who used either treated human faeces or did not use human faeces as a fertilizer.

2. Methods

The approach builds on methods presented in Prüss et al. (2002), which are further developed in this estimate.

There is strong evidence that, even in developed nations, there is a considerable burden of disease associated with poor-quality potable water or inappropriate sewage disposal and sanitary control. This was demonstrated by disease outbreaks such as the cryptosporidiosis and *Escherichia coli O157* epidemics, which affected Canada, the United States and the United Kingdom of Great Britain and Northern Ireland (Andersson and Bohan 2001; Bouchier 1998; Bruce-Grey-Owen Sound Health Unit 2000). In addition, there is a background of sporadic cases in which unsafe WSH has been implicated (Fewtrell and Delahunty 1995). Hence significant health gain is achievable through further improvement in developed nation WSH conditions. This improved condition represents the theoretical minimum exposure in which no disease transmission would occur through unsafe WSH.

As the five transmission pathways of the various outcomes caused by unsafe WSH are quite different (see section 1.2 of this chapter), two approaches for estimating the disease burden were chosen according to the outcome. The estimates of the burden of infectious diarrhoeal disease caused by unsafe WSH are based on exposure information. The burden of other diseases is entirely due to unsafe WSH.

2.1 Estimating exposure for diarrhoeal diseases

For estimating the burden of diarrhoeal disease caused by unsafe WSH, we used a scenario-based approach to define exposure categories. In this approach the risk of diarrhoeal disease is conditioned by a typical exposure or a representative combination of risk factors at commonly encountered levels. Six scenarios[1] (Table 16.2) were defined on the basis of the following:

Table 16.2 Exposure scenarios

Level	*Description*	*Environmental faecal–oral pathogen load*
VI	Population not served with improved water supply and no improved sanitation in countries which are not extensively covered by those services (less than 98% coverage), and where water supply is not likely to be routinely controlled	Very high
Vb[a]	Population having access to improved water supply but not served with improved sanitation in countries which are not extensively covered by those services, and where water supply is not likely to be routinely controlled (less than 98% coverage)	Very high
Va[a]	Population having access to improved sanitation but no improved water supply in countries where less than 98% of the population is served by water supply and sanitation services, and where water supply is likely not to be routinely controlled	High
IV	Population having access to improved water supply and improved sanitation in countries where less than 98% of the population is served by water supply and sanitation services, and where water supply is likely not to be routinely controlled	High
III[b]	IV and improved access/quality to drinking water; or IV and improved personal hygiene; or IV and drinking water disinfected at point of use, etc.	High
II	Population having access to improved water supply and sanitation services in countries where more than 98% of the population is served by those services; generally corresponds to regulated water supply and full sanitation coverage, with partial treatment for sewage, and is typical in developed countries	Medium to low
I	Ideal situation, corresponding to the absence of transmission of diarrhoeal disease through WSH	Very low

[a] Transitions between exposure levels Va and Vb do not generally occur.

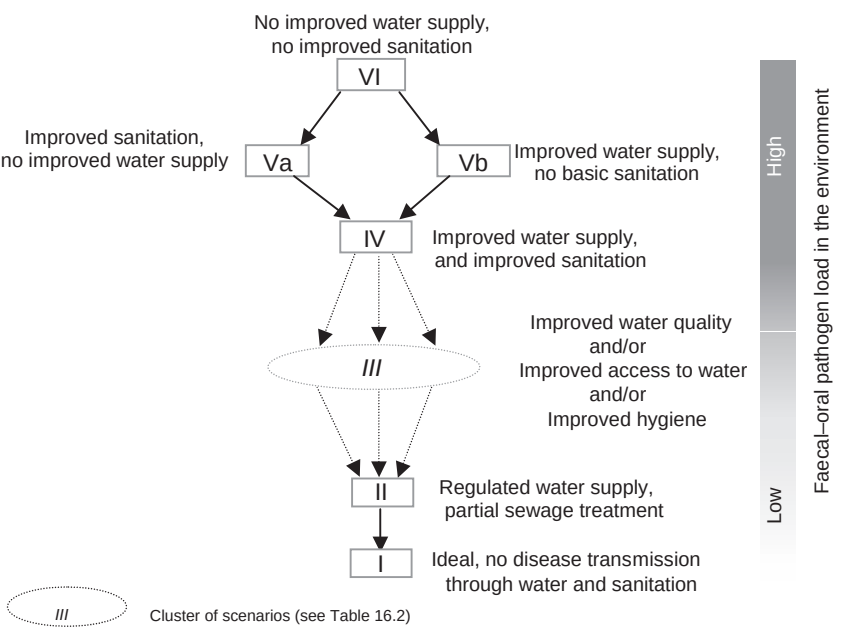
[b] Cluster of possible improvements over scenario IV, but not reaching scenario II.

- the type of water and sanitation infrastructure; and
- the load of faecal–oral pathogens in the environment based on qualitative assessment of sources and disease circulation in the community.

This choice was based on the absence of comprehensive exposure information at individual level and also the lack of relative risk information relating to individual exposure. Risk information was gathered from the literature to match each of the scenarios.

Scenario I represents the minimum theoretical risk and II the situation typically encountered in developed countries. These two scenarios have very low to medium loads of faecal–oral pathogens, characterized by more than 98% coverage in improved water supply and sanitation and a regional incidence of diarrhoea of less than 0.3 per person per year (Anonymus 2000; Murray and Lopez 1996b). Scenarios IV–VI are in a high faecal–oral pathogen environment, typical for developing countries. Scenario III represents any intervention that improves on scenario IV, and does not currently occur widely. As such, various transitions can be proposed for scenario III and so it is represented as a cluster of possibilities rather than a specific scenario (see Figure 16.2).

Figure 16.2 Scenarios determining risk of diarrhoeal disease from unsafe WSH



Data sources and quantification of exposure

The exposure scenarios were selected according to available information on exposure–risk relationships and exposure information from the *Global water supply and sanitation assessment 2000* (WHO/UNICEF/WSSCC 2000). The data on water supply and sanitation coverage provided in this assessment are a compilation of two main sources: household surveys, and to a lesser degree assessment questionaires. Relevant information from available household surveys performed on a large scale was accessed, including:

- Demographic Health Surveys (DHS) performed by Macro International and funded by the United States Agency for International Development;
- United Nations Children's Fund's (UNICEF) Multiple Indicator Cluster Surveys (MICS);
- national census reports; and
- other national sample household surveys.

DHS and MICS are national cluster sample surveys, covering several thousand households in each country. The samples are stratified to ensure they are representative of urban and rural areas of each country. In household surveys, consumers are asked to identify the type of water facility they use from a list of technologies. In estimating coverage for the Global Water Supply and Sanitation Assessment 2000, the types of access to services were categorized into "improved" (e.g. borehole, protected dug well, simple pit latrine) and "not improved" (e.g. unprotected well, vendor-provided water, bucket latrines). In addition, national assessment questionnaires were completed by the relevant national agencies in cooperation with WHO and UNICEF country staff. The resulting country estimate for coverage was then based on linear regressions prepared according to available survey data. In the rare cases where household survey data were not available, the coverage figures adopted were those estimated by a local expert committee, based on national assessments and information provided by the country's water authorities.

The Assessment 2000 provides data for water supply and sanitation for almost every country, with information typically available for more than 90% of the population in every region. It is the only comprehensive assessment of this kind and is, therefore, the single source used for assessing exposure in this analysis. Overall, the Assessment 2000 represents more than 89% of the global population. Only the European and Western Pacific Regions contain large data gaps, with information on water supply and sanitation coverage lacking for some large countries. Subregions[2] with low information coverage include EUR-A (25% information coverage), EUR-B (65%), EUR-C (11%) and WPR-A (15%). In

each case (based on those countries responding), we considered that the available figures on coverage were likely to be representative of the whole subregion. Countries without information were ascribed subregional coverage rates.

Using data from the Assessment 2000, it is not possible to assess whether those served with improved water corresponded to those with improved sanitation, as only coverage was reported. Reports suggest a strong societal and individual preference for improved water supply over improved sanitation, and this is further supported by the higher levels achieved worldwide for improved water supply when compared to improved sanitation. In apportioning populations among exposure scenarios IV and Vb, we therefore assumed that people with improved water supplies were likely to have access to improved sanitation.

The population of each country was thus assigned to the various scenarios based on the Assessment 2000 as described above, and population-weighted regional means were calculated. The resulting exposure distribution is represented in Table 16.3. In 2000, the percentage of people served with some form of improved water supply worldwide reached 82% (4.9 billion), and 60% (3.6 billion) had access to improved sanitation facilities. In 2000, one sixth (1.1 billion people) of the world's population was still without access to improved water supply and two fifths (2.4 billion people) lacked access to improved sanitation.

Table 16.3 Distribution of the population in exposure scenarios, 2000

Subregion	*II (%)*	*IV (%)*	*Va (%)*	*Vb (%)*	*VI (%)*
AFR-D	0	54	5	6	35
AFR-E	0	42	10	9	38
AMR-A	99.8	0	0	0	0.2
AMR-B	0	76	1	9	14
AMR-D	0	68	0	7	25
EMR-B	0	83	5	8	4
EMR-D	0	66	0	16	18
EUR-A	100	0	0	0	0
EUR-B[a]	0	79	8	1	12
EUR-C[a]	0	94	5	0	1
SEAR-B	0	70	3	7	19
SEAR-D	0	35	0	53	12
WPR-A	100	0	0	0	0
WPR-B	0	42	1	33	24

[a] Low data coverage.

Source: Based on data from the *Global water supply and sanitation assessment 2000* (WHO/UNICEF/WSSCC 2000), assuming that improved water supplies are most likely to have sanitation coverage.

Scenario I does not occur on a large scale and, in global terms, is probably negligible, hence its omission from Table 16.3. Scenario III is a poorly characterized series of transition states between IV and II and is not separately accounted for. Such scenarios are nevertheless important concepts in policy development and are therefore retained in the model described in Figure 16.2.

2.2 Risk factor–disease relationships for diarrhoeal diseases

Approach

We selected major reviews, multi-country studies or studies of superior design to quantify the transition between two or more chosen exposure scenarios. This included the review and multi-country study by Esrey (Esrey 1996; Esrey et al. 1991), the reviews by Huttly et al. (1997) and Mead et al. (1999), in conjunction with key literature and high quality studies published since the review papers (Payment et al. 1991, 1997; Quick et al. 1999; Semenza et al. 1998). The majority of this literature was based on intervention studies and surveillance information. The final selection of used studies depended largely on the degree to which the study exposure data could be matched with the chosen exposure scenarios and also the sample size and quality of studies. Brief details on the chosen studies are outlined in Table 16.4.

Relative risk for exposure scenario II

The ideal situation (scenario I) is the theoretical minimum (RR = 1). In scenario II, the pathogen load is mostly transferred from land to water (e.g. in discharge of normally treated sewage, such as biological secondary treatment, to surface water). Such pathogens can potentially pass through potable water treatment systems, which can not guarantee 100% pathogen elimination in even the most advanced plants used in developed nations. Water contaminated with such pathogens is also used for other purposes such as recreation and irrigation. Hygiene behaviour is still imperfect in scenario II, and small population groups may still be served with poorly regulated community supplied water. In scenario I, the ideal scenario, all this would not occur.

Relative risk for scenario II was based on the review by Mead et al. (1999). Mead et al. assessed the level of all infectious foodborne illness in the United States, using data from a large number of surveys and other sources (including FoodNet, the National Notifiable Disease Surveillance System, the Public Laboratory Information System, the Foodborne Disease Outbreak Surveillance System, the National Hospital Discharge Survey, the National Vital Statistics System and a number of published studies). Based on the literature, they also estimated the percentage of each disease caused by foodborne transmission. This is a very comprehensive study based on more than 400 000 diagnosed cases, bringing

Table 16.4 Key studies and reviews

Reference	*Study population*	*Sample size*	*Outcome measured/ reported*	*Reductions*	*Comments*
Esrey (1996)	Representative populations from Bolivia, Burundi, Ghana, Guatemala, Morocco, Sri Lanka, Togo, Uganda	16 880	Diarrhoea morbidity nutritional status, child development	20.8–37.5% according to type of infrastructure	Detailed examination of effects of incremental improvements in water and sanitation based on survey data
Huttly et al. (1997)	Bangladesh, Burma, India, Indonesia, USA for hand-washing; Bangladesh, Guatemala, Thailand, Zaire for various other forms of behaviour	NA	Diarrhoea morbidity	Median reduction 35% for hand-washing; median 26% for other hygiene behaviours	Review paper/5 intervention studies on hand-washing and 5 on other hygiene behaviours
Mead et al. (1999)	Gastrointestinal illness in the USA population	More than 400 000 diagnosed patients	Foodborne illness	Approx. 60% of gastrointestinal illness due to unsafe WSH[a]	Surveillance data
Payment et al. (1991)	606 households in Montreal, Canada	2 408	Diarrhoea morbidity	35%	Water quality intervention
Payment et al. (1997)	1 400 families in Montreal, Canada	5 253	Diarrhoea morbidity	14–40%	Water quality intervention
Quick et al. (1999)	Two Bolivian communities	791	Diarrhoea morbidity	45% for all age groups	Water quality intervention
Semenza et al. (1998)	Householders in Nukus, Uzbekistan	1 583	Diarrhoea morbidity	62–85%	Water quality intervention

NA Not applicable.

[a] Extrapolated from study results for the purpose of this analysis.

together numerous different data sources and some assumptions relating to likely levels of underreporting. According to this study, about 35% of intestinal illness in the United States is foodborne. The level of faecal–oral illness due to unsafe WSH was estimated as 100% of the cases of infectious diarrhoea, less the percentage due to foodborne transmission. This is probably an underestimate as it is likely that unsafe WSH play a role in some foodborne transmission (e.g. through irrigation of food products with pathogen-contaminated water or via an infected food handler). After deduction of the portion of foodborne transmission and accounting for likely ratios of person-to-person transmission through aerosols of certain viruses (estimated as up to 25% for rotavirus and astrovirus), the remaining fraction attributable to unsafe WSH is about 60%. This order of magnitude is supported by intervention studies acting on point-of-use treatment of drinking water in Canada (Payment et al. 1991, 1997) and hand-washing in the United States (Black et al. 1981), reporting reductions of 40%, 35% and 48%, respectively. A 60% reduction in disease corresponds to a relative risk of 2.5 (RR = 1/(1 – 0.6)) for exposure scenario II.

Risk transition between scenarios II and IV

Scenarios II and IV represent high and low environmental pathogen loads. Intervention studies were not available, as it is not possible to transform environments high in pathogen load into environments low in pathogen load; doing so would imply completing the coverage in improved water supply and sanitation in a reasonable time frame and without simultaneous change in other major determinants of health. Therefore, relative risks for scenarios between II and IV were estimated using selected studies.

- *Scenario IV and improved drinking-water quality:* Quick et al. (1999) examined the level of diarrhoea prevention that could be achieved through point-of-use water treatment along with safe water storage. This study was selected as the intervention strongly reduces the pathway of transmission through drinking water, and "simulates" the reduction that could be achieved by improved drinking water quality and its handling inside the house. The study randomized 791 participants into two groups. The intervention group received a special storage container (preventing hand contact with the stored water) and a supply of disinfectant. The control group not receiving the intervention was similar in terms of demographic characteristics, sanitary conditions and baseline water quality. During the baseline investigations only 5% of household samples were free of *E. coli*. During the study period this varied between 0% and 13% of the control group (with the median level being between 5000 and 85 000 of *E. Coli*/100 ml), while the intervention group exceeded 50% of households at all times, rising to almost 80% on one occasion (median *E.*

coli counts were zero, throughout). Overall diarrhoea reductions of 44.7% in the total population and 54.5% in children have been reported by Quick et al. (RR=1.81 and 2.20). The reduction of 44.7% was selected as a component in the transition between II and IV in this analysis.

In a randomized intervention study in 240 households (120 with and 120 without access to municipal piped water) with a total population of 1583 in Uzbekistan (Semenza et al. 1998), approximately half of the households without piped water were trained to chlorinate their drinking water within the home and store it in a safe manner. Diarrhoea morbidity was markedly lower in the home-chlorination group (28.8/1000 subjects per month), compared to 75.5/1000 in the piped water group and 179.2/1000 in the no piped water group (i.e. a 62% reduction in diarrhoea rates for an intervention with home chlorination of drinking water, as compared to those living in areas with access to piped water [RR = 2.6]; in individuals without a piped supply, the same intervention achieved a 85% reduction in disease [RR = 6.7]). The authors considered that home chlorination of water was unlikely to affect disease transmission via other routes, and suggested that a large fraction of the diarrhoeal pathogens in this area were spread through water.

- *Scenario IV and improved personal hygiene:* reductions in diarrhoea morbidity have been reviewed by Huttly et al. (1997), and handwashing resulted in a median 35% reduction in diarrhoea incidence (RR = 1.5). The results of this review outlined possible achievements due to a reduction in the transmission pathway of hygiene, which in itself is conditioned by the pathogen load in the environment.

Risk transition between scenarios IV and VI

The multi-country study conducted by Esrey (1996) provided data to allow calculation of relative risks between scenarios IV, Va, Vb and VI. This study examined whether incremental health effects relating to diarrhoea and nutritional status resulted from incremental improvements in water and sanitation conditions and was based on DHS from eight countries from five different regions (Bolivia, Burundi, Ghana, Guatemala, Morocco, Sri Lanka, Togo, Uganda). DHS included information on diarrhoea prevalence, child weight, child height, child age, source of drinking water and type of sanitation facility. In addition, the survey data were supplemented by field studies that determined current levels of diarrhoea prevalence in children aged 3–36 months. According to this study, a reduction of 20.8% in diarrhoeal disease rates (RR=1.26) could be observed when progressing from scenario VI to Vb (i.e. when providing an improved water supply), and 37.5% (RR=1.6) when progressing from VI to Va (i.e. when providing improved sanitation facilities). When progressing from VI to IV (i.e. when providing both an improved water

Table 16.5 Relative risks

	Exposure categories or transition between scenarios							
	I	*II*	*III*	*IV*	*Va (to IV)*	*Vb (to IV)*[a]	*VI (to Vb)*	*VI (to Va and IV)*
Risk reduction[b]	NA	60%	Various[c]	45% and 35%	0%	—	20.8%	37.5%
Partial relative risk[a]	NA	2.5	Various[c]	1.81 and 1.54	1.0	1.60/1.26 = 1.27	1.26	1.60
Absolute relative risks (compared to scenario I)	I	2.5	Various[c]	6.9	6.9	8.7	11.0	11.0

NA Not applicable.

— No data.

[a] Obtained by calculating the remaining risk differences between VI to Vb as compared to VI and IV.

[b] Relative to the scenario below.

[c] See text.

supply and improved sanitation facilities), a reduction of 37.5% was also achieved. This implies that no further reduction in diarrhoeal disease is achieved when implementing an improved water supply, when improved sanitation is already available. These data are supported by the review of Esrey et al. (1991), which provides similar results for the same types of interventions.

The resulting relative risks are obtained by multiplying the relative risks between each scenario, summarized in Table 16.5.

According to our model, the risks of diarrhoea incidence in developing countries are 2.8 to 4.4 times higher (Table 16.5) than current risks in developed countries. The same order of magnitude of difference in diarrhoea rates was reported by various compilations of health statistics or studies (Esrey 1996; Murray and Lopez 1996b).

2.3 Estimating risk factor–disease relationships for diseases other than diarrhoeal diseases

The *World health report 2001* (WHO 2001) provided estimates of the burden of additional diseases that are exclusively (or virtually exclusively) caused by unsafe WSH (Table 16.6).

2.4 Sources of uncertainty

Method

The method is based on typical scenarios, characterized by a combination of sub-risk factors, which should represent most of the world's situations. Certain population groups may not be captured by any one of these scenarios, but the number of groups is probably small, which may

Table 16.6 Global disease burden caused by selected water-related diseases other than infectious diarrhoea in 2000

Disease	*Deaths (000s)*	*DALYs (000s)*
Schistosomiasis	11	1713
Trachoma	0	1161
Ascariasis	6	1252
Trichuriasis	2	1640
Hookworm disease	6	1829
Total	25	7595

Source: WHO (2001).

be partly internalized in the risk estimates. For example, differences are likely to exist in the specific WSH practices in the various households in the same exposure scenario. The effect of these differences should, however, largely be captured in the large samples on which this study is based. The current study is therefore based on average risks for large population groups within which a variety of individual practices and situations are represented.

Exposure estimates

The Water Supply and Sanitation Assessment 2000, which reports individual country data, exhibited variable precision between respondents, particularly in relation to rural and tribal populations. A more precise exposure estimate would require actual assessments, such as the water quality of the supply. Such measures are impractical on a large level. The Assessment 2000, however, captures exposure information for a majority of the world's countries and represents a solid source of information. Uncertainty in water supply and sanitation coverage has therefore not introduced major uncertainty into our analysis.

Risk estimates

This analysis used large surveys and multi-country studies where available. It is therefore based on risk averages, i.e. the average of risk related to the described scenarios across the world and across an array of situations. While this method may not be suitable for specific local settings, it should provide a reasonable estimate for large regions.

Where no large surveys, reviews or multi-country studies were available (i.e. in part the transition between scenarios IV and II), the use of sentinel studies for "global" application may constitute a significant source of error. Therefore, this analysis has been selective on the basis of study quality and coverage, to ensure maximum transferability.

As much of the described imprecision will remain largely unquantifiable, upper and lower uncertainty boundaries are based on varying the

Table 16.7 Low and high relative risk estimates

	Exposure scenario						
	I	*II*	*III*	*IV*	*Va*	*Vb*	*VI*
Lower estimate	1	2.5	Variable	3.8	3.8	4.9	6.1
Best estimate	1	2.5	Variable	6.9	6.9	8.7	11.0
Upper estimate	1	2.5	Variable	10.0	10.0	12.6	16.0

estimates of the potentially greatest source of uncertainty—the transition between scenarios IV and II (i.e. from an environment with a high faecal–oral pathogen load to one with a low faecal–oral pathogen load). The lower estimate was based solely on the improvement that can be achieved by implementing personal hygiene measures (35% risk reduction or a RR of 1.54). This is comparable with the best estimate, which is based on a combination of improved water quality and improved hygiene (see Table 16.5). For the upper estimate, the additional risk reduction relating to the provision of continuous piped water supply (i.e. improved access to water) was considered in addition to hygiene improvements. This is represented by a relative risk of 2.6 (from a 62% risk reduction from the study by Semenza et al. 1998) in addition to that resulting from hygiene improvements. The resulting relative risks for each of the estimates are summarized in Table 16.7.

The same relative risk is assumed for all age groups. As most of these rates have been assessed for children and the largest disease burden also occurs in that age group, the error of applying the same relative risks to adults is probably small. Also, several studies that have assessed both relative risks for children and adults have shown that figures do not generally differ dramatically, although the impact on young children tends to be higher (e.g. Quick et al. 1999).

It should be noted that faecal–oral disease transmission is partly conditioned by the prevalence of the risk factor at community level. For example, protection of drinking water depends on the effective implementation of an intervention by all members of the community, whereas studies have often been performed at individual level, generally resulting in underestimation of the benefits of community-wide interventions.

Sequelae and delayed effects

Estimation of the burden of disease due to infectious diarrhoea is based upon the acute diarrhoeal episode and associated mortality. Several of the agents of infectious diarrhoea are associated with other health effects, often delayed. These may add significantly to the burden of disease, as is the case of campylobacteriosis, for example. Inadequate evidence was available to reliably estimate the additional burden of disease.

3. Results

The attributable fractions, deaths and number of DALYs are listed in the annex tables (see CD-ROM accompanying this book), for the 14 subregions, males, females and eight age groups.

Globally, in the year 2000, almost 1.73 million deaths due to diarrhoeal diseases were attributable to unsafe WSH as defined in the exposure variable used in this work; 68% of them are children. Most of these deaths, >99%, occur in developing countries. The attributable fractions of diarrhoeal disease vary between 60% in developed countries to 85–90% in developing countries. The difference in disease burden between developed and developing subregions, despite the relative similarities in attributable fractions, is largely due to the lower incidence and case fatality rates of diarrhoeal disease in developed nations. The African subregions alone, together with SEAR-D and EMR-D, bear 88% of the death burden. The disease burden in males and females is similar. The disease burden from the five other diseases that have been quantified separately is 25 000 deaths and 7.6 million DALYs, also concentrated in developing countries. This chapter highlights and confirms the concentration of the burden of disease due to the risk factor unsafe WSH in poor countries and on children—99.7% of DALYs and 99.8% of deaths occur in developing countries, with 80% of DALYs among children. Globally, 3.1% of all deaths and 3.7% of DALYs were attributable to water, sanitation and hygiene, caused by the diseases we could include in this analysis. In the age group 0 to 4 years, these percentages amounted to 11% of all deaths and 9% of all DALYs, which shows the importance of this risk factor.

4. Projections of future exposure

As the methods for estimating disease burden rely heavily on water supply and sanitation coverage, these are the main parameters that need to be projected for estimating future burden. Progress with water supply and sanitation coverage is affected by factors such as demographic change, income, policies and investments, education, technology, types and management of infrastructure, and involvement of the community and the public and private sectors. In practice, these vary widely within and between countries, making future projections difficult and complex.

To some extent these factors respond to major national and international policy initiatives. The International Drinking-water Supply and Sanitation Decade (1981–1990) established momentum that certainly produced an acceleration of investments from 1981 to 1990 and beyond this period. The Millennium Declaration established the targets of halving the proportion of the population not served with safe water supply by 2015 and improving sanitation for the urban poor. The impact of these historic and future activities on either overall progress with service levels or upon the factors outlined above is difficult to assess. The

proposed coverage forecast method and respective coverage figures generated are presented below.

While efforts are ongoing to develop a model for forecasting improved water supply and sanitation coverage based upon understanding of the factors outlined above, the lack of sufficient data has limited the value of this in preparing future projections. Water supply and sanitation coverage may be predicted by certain distal causes such as income and education; however, in the given time frame a prediction based on past evolution and future demographic changes was preferred. Prediction on the basis of the Human Development Index provided similar results at global level. Global data sets on service coverage generated by WHO in the 1980s were primarily based on country reporting and provided results with limited comparability. More recently, WHO and UNICEF have assessed water supply and sanitation coverage in 1990 and 2000 (WHO/UNICEF/WSSCC 2000), based on household survey data and data by service providers (water agencies, ministries) in the absence of survey data. This shift in methodology provided more reliable and comparable data. The prediction was thus based on the following points.

- It was assumed that the same number of people that acquired coverage between 1990 and 2000 would acquire coverage per decade during the next three decades.
- Population projections from the United Nations Statistics Division (UN 2001) were used.
- For EUR-B and EUR-C, progress in the decade 1990–2000 shows declining trends and does not provide a reasonable basis for projection. Zero change in absolute numbers served was assumed.

It is important to note that this projection assumes that local, national and international efforts as undertaken in the last decades, will continue. The method further assumes no development of approaches or technologies that will enable a shift for part of the population into exposure scenario I, the ideal scenario.

Coverage was projected separately for each subregion. A summary of projected water supply and sanitation coverage is provided in Table 16.8, and a detailed projection per subregion, according to the exposure scenarios used in this analysis, is presented in Table 16.9.

The data presented suggest that the water supply goal and target[3] adopted in the Millennium Declaration are likely to be achieved globally if a similar effort as compared to that undertaken in the last decade is continued until the year 2015. For certain subregions, however, the target may not be achieved, namely those in the African continent, as well as EMR-B (where coverage is already high and where half of the population are likely to experience an important risk reduction), EUR-B and EUR-C, under the assumption that past trends will continue.

Table 16.8 Global projection of water supply and sanitation coverage

		Access to improved water sources			Access to improved sanitation		
Year	Total population (millions)	Population served (millions)	% served	Population not served (millions)	Population with access (millions)	% having access	Population without access (millions)
1990	5255	4072	77	1183	2582	49	2673
2000	6057	4976	82	1081	3646	60	2411
2010	6826	5894	86	932	4739	69	2087
2015	7207	6353	88	854	5285	73	1922
2020	7579	6802	90	777	5831	77	1748
2030	8270	7681	93	589	6902	83	1368

It is not always possible to see clear trends within scenarios. This is likely to be due to highly variable rates of population growth and movement of populations between scenarios. Greatest health gains are likely to be associated with movement of populations from scenario VI to better circumstances. AFR-D and AFR-E contain the highest proportion of population in scenario VI. The forecasts indicate that the situation is not likely to change dramatically over the next 30 years if the trends of the last decade continue.

SEAR-D and WPR-B present large proportions of their population with fairly good levels of coverage but relatively low proportions of people served with sanitation facilities (scenario Vb). AMR-A, EUR-A, and WPR-A have reached or will soon reach 100% coverage (scenario II). SEAR-B and AMR-B are projected to make good progress, tending from exposure scenarios Vb and VI to IV and II. AMR-D, in addition to developing a trend similar to the trend above is likely to make considerable progress towards scenario II. EUR-C should experience a large shift into scenario II (two large countries that were close to full coverage in 1990 will reach such status by 2010).

Sources of error/sensitivity analysis

Factors such as water scarcity, competition for water resources and the cumulative effects of pollution of water resources are likely to both increase the cost of interventions and reduce their sustainability. This is due to pressures on both the quality and availability of water resources and would suggest that the projections might be optimistic estimates.

Acknowledgements

We gratefully acknowledge the comments and suggestions provided by Steve Luby and Carlos Corvalan, the participants in the review meeting of the WHO comparative risk assessment in Auckland in December

Table 16.9 Projection of distribution of exposure by scenarios, 2000 to 2030[a]

	II (%)				*IV (%)*				*Va (%)*				*Vb (%)*				*VI (%)*			
	Year				*Year*				*Year*				*Year*				*Year*			
Subregion	*2000*	*2010*	*2020*	*2030*	*2000*	*2010*	*2020*	*2030*	*2000*	*2010*	*2020*	*2030*	*2000*	*2010*	*2020*	*2030*	*2000*	*2010*	*2020*	*2030*
AFR-D[b]	0	1	1	0	54	55	54	54	5	2	1	1	6	11	12	14	35	32	32	31
AFR-E	0	16	12	12	42	44	42	48	10	7	4	3	9	7	11	6	38	27	31	31
AMR-A	99.8	100	100	100	0.2	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0
AMR-B	0	0	4	4	76	81	81	85	1	0	0	1	9	10	10	9	14	8	4	2
AMR-D	0	0	17	16	68	77	65	68	0	1	1	1	7	6	5	3	25	15	13	11
EMR-B	0	3	7	53	83	83	79	34	5	5	5	5	8	4	3	1	4	5	6	7
EMR-D	0	23	22	21	66	53	59	65	0	0	0	0	16	10	6	3	18	13	13	11
EUR-A	100	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EUR-B	0	7	7	6	79	72	73	74	8	6	6	5	1	1	1	1	12	13	14	14
EUR-C	0	74	72	70	94	23	24	25	5	2	2	2	0	0	0	0	1	1	2	3
SEAR-B	0	27	20	19	70	45	56	61	3	0	1	0	7	17	17	17	19	11	7	3
SEAR-D	0	0	0	0	35	39	47	54	0	0	0	0	53	58	53	46	12	2	0	0
WPR-A	100	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WPR-B	0	0	0	0	42	49	60	67	1	0	1	1	33	31	29	26	24	19	14	8

[a] Results in percentage of regional population.

[b] Rounding of percentages may lead to sums slightly different from 100%.

Source: José Hueb, personal communication.

2000, and the five peer reviewers. We also wish to acknowledge the valuable support and contribution of the U.S. Environmental Protection Agency. This chapter has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency.

Notes

1 The scenarios are equivalent to exposure categories used in other chapters in this book, in the sense that there is increasing risk across scenarios defined based on faecal–oral load. The term scenario is used here, as the shift from one level of faecal–oral load to another may occur due to changes in any of the multiple dimensions of exposure (water, sanitation and hygiene).

2 See the preface for an explanation of this term.

3 To halve the proportion of people not having access to water supply services by 2015 compared to 1990.

References

Ahmed NU, Zeitlin MF, Beiser AS, Super CM, Gershoff SN (1993) A longitudinal study of the impact of behavioural change intervention on cleanliness, diarrhoeal morbidity and growth of children in rural Bangladesh. *Social Science and Medicine*, **37**:159–171.

Alam N, Wojtyniak B, Henry FJ, Rahaman MM (1989) Mothers' personal and domestic hygiene and diarrhoea incidence in young children in rural Bangladesh. *International Journal of Epidemiology*, **18**:242–247.

Ampofo JA (1997) A survey of microbial pollution of rural domestic water supply in Ghana. *International Journal of Environmental Health Research*, **7**:121–130.

Anderson TJ, Zizza CA, Leche GM, Scott ME, Solomons NW (1993) The distribution of intestinal helminth infections in a rural village in Guatemala. *Memorias do Instituto Oswaldo Cruz*, **88**:53–65.

Andersson Y, Bohan P (2001) Disease surveillance and waterborne outbreaks. In: *Water quality: guidelines, standards and health. Assessment of risk and risk management for water-related infectious disease*. Fewtrell L, Bartram J, eds. World Health Organization, Geneva.

Arfaa F, Sahba GH, Farahmandian I, Jalali H (1977) Evaluation of the effect of different methods of control of soil-transmitted helminths in Khuzestan, south-west Iran. *American Journal of Tropical Medicine and Hygiene*, **26**: 230–233.

Barbosa FS, Pinto R, Souza OA (1971) Control of schistosomiasis mansoni in a small north east Brazilian community. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **65**:206–213.

Barwick RS, Levy DA, Craun GF, Beach MJ, Calderon RL (2000) Surveillance for waterborne-disease outbreaks—United States, 1997–1998. *Morbidity and Mortality Weekly Reports*, **49**:S1–36.

Birmingham ME, Lee LA, Ndayimirije N et al. (1997) Epidemic cholera in Burundi: patterns of transmission in the Great Rift Valley Lake region. *The Lancet*, **349**:981–985.

Black RE, Dykes AC, Anderson AC et al. (1981) Handwashing to prevent diarrhoea in day-care centers. *American Journal of Epidemiology*, **113**:445–451.

Bouchier IAD (1998) Report of the Group of Experts on Cryptosporidium in water supplies. Department of the Environment, Transport and the Regions. London.

Bruce-Grey-Owen Sound Health Unit (2000) The investigative report on the Walkerton outbreak of waterborne gastroenteritis. Available at http://www.publichealthbrucegrey.oc.ca/private/Report/SPReport.htm.

Briscoe J (1984) Intervention studies and the definition of dominant transmission routes. *American Journal of Epidemiology*, **120**:449–455.

Cifuentes E (1998) The epidemiology of enteric infections in agricultural communities exposed to wastewater irrigation: perspectives for risk control. *International Journal of Environmental Health Research*, **8**:203–213.

Coura-Filho P, Rocha RS, Farah MW, da Silva GC, Katz N (1994) Identification of factors and groups at risk of infection with *Schistosoma mansoni*: a strategy for the implementation of control measures? *Revista Instituto de Medicina Tropical de São Paulo*, **36**:245–253.

Curtis V, Cairncross S, Yonli R (2000) Domestic hygiene and diarrhoea—pinpointing the problem. *Tropical Medicine and International Health*, **5**:22–32.

Dolin PJ, Faal H, Johnson GJ, Minassian D et al. (1997) Reduction of trachoma in a sub-Saharan village in absence of a disease control programme. *The Lancet*, **349**:1511–1512.

Dupont HL, Levine MM, Hornick RB, Formal SB (1989) Inoculum size in shigellosis and implications for expected mode of transmission. *Journal of Infectious Diseases*, **159**(6):1126–1128.

Emerson PM, Lindsay SW, Walraven GE et al. (1999) Effect of fly control on trachoma and diarrhoea. *The Lancet*, **353**:1401–1403.

Esrey SA (1996) Water, waste, and well-being: a multicountry study. *American Journal of Epidemiology*, **143**:608–623.

Esrey SA, Potash JB, Roberts L, Shiff C (1991) Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization*, **69**:609–621.

Feachem RG (1984) Interventions for the control of diarrhoeal diseases among young children: promotion of personal and domestic hygiene. *Bulletin of the World Health Organization*, **62**:467–476.

Ferley JP, Zmirou D, Collin JF, Charrel M (1986) Etude longitudinale des risques liés à la consommation d'eaux non conformes aux normes bactériologiques. *Revue d' Epidémiologie et de Santé Publique*, **34**:89–99.

Fewtrell L, Delahunty A (1995) The incidence of cryptosporidiosis in comparison with other gastrointestinal illnesses in Blackpool, Wyre and Fylde. *Journal of the Chartered Institution of Water and Environmental Management*, **9**:598–601.

Habbari K, Tifnouti A, Bitton G, Mandil A (2000) Geohelminthic infections associated with raw wastewater reuse for agricultural purposes in Beni-Mellal, Morocco. *Parasitology International*, **48**:249–254.

Henry FJ (1981) Environmental sanitation infection and nutritional status in infants in rural St. Lucia, West Indies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **75**:507–513.

Humphries DL, Stephenson LS, Pearce EJ, The PH, Dan HT, Khanh LT (1997) The use of human faeces for fertilizer is associated with increased intensity of hookworm infection in Vietnamese women. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **91**:518–520.

Hunter PR (1997) *Waterborne disease: epidemiology and ecology*. John Wiley & Sons Ltd., Chichester, England.

Huttly SRA, Morris SS, Pisani V (1997) Prevention of diarrhoea in young children in developing countries. *Bulletin of the World Health Organization*, **75**:163–174.

Jonnalagadda PR, Bhat RV (1995) Parasitic contamination of stored water used for drinking/cooking in Hyderabad. *Southest Asian Journal of Tropical Medicine and Public Health*, **26**:789–794.

Jordan P (1972) Epidemiology and control of schistosomiasis. *British Medical Bulletin*, **28**:55–59.

Lima e Costa MFF, Rocha RS, Leite MLC et al. (1991) A multivariate analysis of socio-demographic factors, water contact patterns and *Schistosoma mansoni* infection in an endemic area in Brazil. *Revista Instituto de Medicina Tropical de São Paulo*, **33**:58–63.

Mahalanabis D, Alam AN, Rahman N, Hasnat A (1991) Prognostic indicators and risk factors for increased duration of acute diarrhoea and for persistent diarrhoea in children. *International Journal of Epidemiology*, **20**:1064–1072.

Mead PS, Abdel-Moneim M, al-Erian RA, al-Amari OM (1999) Food-related illness and death in the United States. *Emerging Infectious Diseases*, **5**:607–625.

Moe CL, Sobsey MD, Samsa GP, Mesolo V (1991) Bacterial indicators of risk of diarrhoeal disease from drinking-water in the Philippines. *Bulletin of the World Health Organization*, **69**:305–317.

Murray CJL, Lopez AD, eds. (1996a) *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Global Burden of Disease and Injury, Vol 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Murray CJL, Lopez AD, eds. (1996b) *Global health statistics*: *a compendium of incidence, prevalence and mortality estimates for over 200 conditions*. Global Burden of Disease and Injury, Vol 2. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Narain K, Rajguru SK, Mahanta J (2000) Prevalence of *Trichuris trichuria* in relation to socio-economic and behavioural determinants of exposure to infection in rural Assam. *Indian Journal of Medical Research*, **112**:140–146.

Njemanze PC, Anozie J, Ihenacho JO, Russell MJ, Uwaeziozi AB (1999) Application of risk analysis and geographic information system technologies to the prevention of diarrheal diseases in Nigeria. *American Journal of Tropical Medicine and Hygiene*, **61**:356–360.

Norhayati M, Oothuman P, Fatmah MS, Muzain Minudin Y, Zainuddin B (1995) Hookworm infection and reinfection following treatment among Orang Asli children. *Medical Journal of Malaysia*, **50**:314–319.

Payment P, Richardson L, Siemiatycki J, Dewar R, Edwardes M, Franco E (1991) A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. *American Journal of Public Health*, **81**:703–708.

Payment P, Siemiatycki J, Richardson L, Renaud G, Franco E, Prévost M (1997) A prospective epidemiological study of gastrointestinal health effects due to the consumption of drinking water. *International Journal of Environmental Health Research*, **7**:5–31.

Prüss A, Mariotti SP (2000) Preventing trachoma through environmental sanitation: a review of the evidence base. *Bulletin of the World Health Organization*, **78**:258–266.

Prüss A, Kay D, Fewtrell L, Bartram J (2002) Estimating the burden of disease due to water, sanitation and hygiene at global level. *Environmental Health Perspectives* **110**:537–542.

Quick RE, Venczel LV, Mintz ED et al. (1999) Diarrhoea prevention in Bolivia through point-of-use water treatment and safe storage: a promising new strategy. *Epidemiology and Infection*, **122**:83–90.

Rajeswari B, Sinniah B, Hussein H (1994) Socio-economic factors associated with intestinal parasites among children living in Gombak, Malaysia. *Asia Pacific Journal of Public Health*, **7**:21–25.

Saldiva SR, Silveira AS, Philippi ST et al. (1999) *Ascaris-Trichuris* association and malnutrition in Brazilian children. *Paediatric and Perinatal Epidemiology*, **13**:89–98.

Semenza JC, Roberts L, Henderson A, Bogan J, Rubin CH (1998) Water distribution system and diarrheal disease transmission: a case study in Uzbekistan. *American Journal of Tropical Medicine and Hygiene*, **59**:941–946.

Smith H, Dekaminsky R, Niwas S, Soto R, Jolly P (2001) Prevalence and intensity of infections of *Ascaris lumbricoides* and *Trichuris trichiura* and associated socio-demographic variables in four rural Honduran communities. *Memorias do Instituto Oswaldo Cruz*, **96**:303–314.

Snow J (1855) *On the mode of cholera communication*. John Churchill, London.

Sorensen E, Ismail M, Amarasinghe DK, Hettiarachchi I, Dassenaieke TS (1994) The effect of the availability of latrines on soil-transmitted nematode infections in the plantation sector in Sri Lanka. *American Journal of Tropical Medicine and Hygiene*, **51**:36–39.

Sutter EE, Ballard RC (1983) Community participation in the control of trachoma in Gazankulu. *Social Science and Medicine*, **17**:1813–1817.

Toma T, Miyagi I, Kamimura K et al. (1999) Questionnaire survey and prevalence of intestinal helminthic infections in Barru, Sulawesi, Indonesia. *Southeast Asian Journal of Tropical Medicine and Public Health*, **30**:68–77.

UN (2001) *World urbanization prospects. The 2001 revision. Urban and rural areas*. (POP/OB/WVP/Rev.2001/1). United Nations, New York.

UN (2000) *United Nations Millennium Declaration*. (Resolution 55/2, adopted at the Millennium Summit, September 6–8.) United Nations, New York.

VanDerslice J, Briscoe J (1993) All coliforms are not created equal: a comparison of the effects of water source and in-house water contamination on infantile diarrheal disease. *Water Resources Research*, **29**:1983–1995.

VanDerslice J, Briscoe J (1995) Environmental interventions in developing countries: interactions and their implications. *American Journal of Epidemiology*, **141**:135–144.

WHO (1998*) Guidelines for safe recreational-water environments: coastal and fresh-waters*. Draft for consultation. World Health Organization, Geneva.

WHO (2001) *World health report 2001*. World Health Organization, Geneva.

WHO/UNICEF/WSSCC (2000) *Global water supply and sanitation assessment 2000 report*. World Health Organization/United Nations Children's Fund Water Supply and Sanitation Collaborative Council, Geneva.

Zhaowu W, Kaiming B, Liping Y, Guifeng Y, Jinhua Z, Qili L (1993) Factors contributing to reinfection with *Schistosomiasis japonica* after treatment in the lake region of China. *Acta Tropica*, **54**:83–88.

Chapter 17

Urban air pollution

Aaron J. Cohen, H. Ross Anderson, Bart Ostro, Kiran Dev Pandey, Michal Krzyzanowski, Nino Künzli, Kersten Gutschmidt, C. Arden Pope III, Isabelle Romieu, Jonathan M. Samet and Kirk R. Smith

Summary

Current scientific evidence, derived largely from studies in North America and Western Europe (NAWE), indicates that urban air pollution,[1] which is derived largely from combustion sources, causes a spectrum of health effects ranging from eye irritation to death. Recent assessments suggest that the impacts on public health may be considerable. This evidence has increasingly been used by national and international agencies to inform environmental policies, and quantification of the impact of air pollution on public health has gradually become a critical component in policy discussions as governments weigh options for the control of pollution.

Quantifying the magnitude of these health impacts in cities worldwide, however, presents considerable challenges owing to the limited availability of information on both effects on health and on exposures to air pollution in many parts of the world. Man-made urban air pollution is a complex mixture with many toxic components. We have chosen to index this mixture in terms of particulate matter (PM), a component that has been linked consistently with serious health effects, and, importantly, levels of which can be estimated worldwide. Exposure to PM has been associated with a wide range of effects on health, but effects on mortality are arguably the most important, and are also most amenable to global assessment. Our estimates, therefore, consider only mortality. Currently, most epidemiological evidence and data on air quality that could be used for such estimates comes from developed countries. We have had, therefore, to make assumptions concerning factors such as the transferability of risk functions, exposure of the population and their underlying vulnerability to air pollution, while trying to ensure that these assumptions are transparent and that the uncertainty associated with them is assessed through appropriate sensitivity analyses.

In order to provide estimates for all 14 subregions,[2] models developed by the World Bank were used to estimate ambient concentrations of inhalable particles (particulate matter with an aerodynamic diameter of <10 μm, PM_{10}) for PM in 3211 national capitals and cities with populations of >100 000 using economic, meteorological and demographic data and the available measurements. To allow the most appropriate epidemiological studies to be used for the estimation of the burden of disease, the estimates for PM_{10} were converted to estimates of fine particles (particulate matter with an aerodynamic diameter of <2.5 μm, $PM_{2.5}$) using available information on geographic variation in the ratio of $PM_{2.5}$ to PM_{10}. Population-weighted subregional annual average concentrations of $PM_{2.5}$ and PM_{10} were obtained using the population of the cities in the year 2000.

Our estimates of the burden of disease were based on the contributions of three health outcomes: mortality from cardiopulmonary disease in adults, mortality from lung cancer, and mortality from acute respiratory infections (ARI) in children aged 0–4 years. Numbers of attributable deaths and years of life lost (YLL) for adults and children (aged 0–4 years) were estimated using risk coefficients from a large cohort study of adults in the United States of America (Pope et al. 2002) and a meta-analytical summary of five time-series studies of mortality in children, respectively. Base-case estimates were calculated assuming that the risk of death increases linearly over a range of annual average concentrations of $PM_{2.5}$, between a counterfactual (or referent) concentration of 7.5 μg/m^3 and a maximum of 50 μg/m^3.

The results indicate that the impact of urban air pollution on the burden of disease in the cities of the world is large, but this is likely to be an underestimate of the actual burden, on the basis of an assessment of sources of uncertainty. There is also considerable variation in our estimates among the 14 subregions, with the greatest burden occurring in the more polluted and rapidly growing cities of developing countries. We estimated that air pollution in urban areas worldwide, in terms of concentrations of PM, causes about 3% of mortality attributable to cardiopulmonary disease in adults, about 5% of mortality attributable to cancers of the trachea, bronchus and lung, and about 1% of mortality attributable to ARI in children. This amounts to about 0.80 million premature deaths (1.4% of the global total) and 6.4 million YLL (0.7% of the global total). This burden occurs predominantly in developing countries, with 39% of attributable YLL occurring in WPR-B and 20% in SEAR-D. The highest proportions of the total burden occurred in WPR-B and EUR-B, where urban air pollution caused 0.7–1.0% of the burden of disease.

We quantified the statistical uncertainty of our base-case estimates by estimating the joint uncertainty in the estimates of annual average concentration of PM and the estimates of the relative risks. Estimates worldwide and for most subregions vary by less than two-fold (50%

uncertainty interval). Model uncertainty due to assumptions about the shape of the concentration–response function, the choice of counterfactual level for PM, and other factors were assessed in sensitivity analyses. For the most part, the worldwide estimates in each sensitivity case are within the 50% uncertainty intervals for the base-case estimates. The sensitivity analyses indicate that our base-case estimates were most sensitive to our choice of concentration–response function and theoretical level of minimum exposure.

1. Introduction

The potential for serious consequences of exposure to high levels of ambient air pollution was made clear in the mid-20th century, when cities in Europe and the United States experienced episodes of air pollution, such as the infamous London Fog of 1952 and Donora Smog of 1948, that resulted in large numbers of excess deaths and hospital admissions. Subsequent clean air legislation and other regulatory actions led to the reduction of ambient air pollution in many regions of the world, and particularly in the wealthy developed countries of North America and Europe. New epidemiological studies, however, conducted over the last decade, using sensitive designs and methods of analysis, have identified adverse health effects caused by combustion-derived air pollution even at the low ambient concentrations that now generally prevail in cities in North America and western Europe (Health Effects Institute 2001). At the same time, the populations of the rapidly expanding mega-cities of Asia, Africa and Latin America are increasingly exposed to levels of ambient combustion-related pollution that rival and often exceed the levels experienced in developed countries in the first half of the 20th century. Current scientific evidence, derived largely from studies in North America and western Europe, indicates that urban air pollution causes a spectrum of effects on health, ranging from eye irritation to death (Anonymous 1996a, 1996b). Recent assessments suggest that the impacts on public health may be considerable (Brunekreef 1997; Cifuentes et al. 2001; COMEAP 2001; Künzli et al. 2000; Ostro and Chestnut 1998). This evidence has increasingly been used by national and international agencies to inform environmental policies, and quantification of the impact of air pollution on public health has gradually become a critical component in policy discussions as governments weigh options for the control of pollution.

Quantifying the magnitude of the impact of air pollution in cities worldwide, however, presents considerable challenges owing to the limited availability of information on both effects on health and on exposures to air pollution in many parts of the world. Measurements of urban air pollution, when available, are available largely for a non-representative sample of urban areas. Many areas of the world lack measurements of any kind, and these must then be estimated using

statistical models (see below). On the basis of these considerations, we defined the target population for this risk assessment exercise as the residents in the year 2000 of national capital cities and of cities worldwide with populations of >100 000.

Man-made urban air pollution, which is derived largely from combustion processes, is a complex mixture containing many toxic components. We indexed this mixture in terms of PM, a component that has been consistently linked with serious effects on health, and, importantly, the levels of which can be estimated worldwide. Exposure to PM has been associated with a wide range of effects on health, but its effects on mortality are arguably the most important, and are also most amenable to global assessment. Our estimates, therefore, consider only mortality. Currently, most epidemiological evidence and data on air quality that could be used for such estimates come from developed countries. We have had, therefore, to make assumptions concerning factors such as the transferability of risk functions, exposure of the population and their underlying vulnerability to air pollution, while trying to ensure that these assumptions are transparent and that the uncertainty associated with them is assessed through appropriate sensitivity analyses.

The general framework for estimating the global burden of disease attributable to specific risk factors is described in chapters 1 and 25. Briefly, the approach involves estimating an attributable fraction(s) for each risk factor in each of the 14 subregions of the world. Estimating the attributable fraction for urban air pollution requires several steps. First, the exposure to urban air pollution of the population of each subregion must be estimated. Second, a theoretical minimum level of exposure must be specified. The attributable fraction quantifies the impact of exposure above this theoretical minimum level. Finally, deriving the attributable fraction requires the estimation of the gradient of risk between the theoretical minimum level and the estimated subregional exposure. These risk functions are derived from epidemiological studies for the purposes of estimating the global burden of disease. As discussed below, epidemiological studies generally estimate exposure to air pollution in terms of ambient concentrations, thus, we use the term "concentration–response" (rather than "exposure–response") to describe the risk function.

This chapter describes our approach to estimating the attributable fraction and presents our estimates of the attributable burden of disease caused by urban air pollution. First, we briefly review background information on exposure to air pollution and then describe our choice of the theoretical minimum level and the approach to estimating the exposure to PM of the populations of the world's cities. Next, we review the current information on the effects of air pollution on health and describe our approach to deriving the concentration–response function(s). Finally, we present and discuss our estimates of the attributable burden and their uncertainties.

2. Exposure to urban air pollution from combustion sources

Combustion of fossil fuels for transportation, power generation, and other human activities produces a complex mixture of pollutants comprising literally thousands of chemical constituents (Derwent 1999; Holman 1999). Exposure to such mixtures is a ubiquitous feature of urban life. The precise characteristics of the mixture in a given locale depend on the relative contributions of the different sources of pollution, such as vehicular traffic and power generation, and on the effects of the local geoclimatic factors. The relative contribution of different combustion sources is a function of economic, social and technological factors, but all mixtures contain certain primary gaseous pollutants, such as sulfur dioxide (SO_2), nitrogen oxides (NO_X) and carbon monoxide (CO), that are emitted directly from combustion sources, as well as secondary pollutants, such as ozone (O_3), that are formed in the atmosphere from directly-emitted pollutants. The pollutant mixture also contains carcinogens such as benzo(α)pyrene, benzene and 1,3-butadiene. When petrol contains lead (Pb), as is still the case in many developing countries, this element is a common constituent of the pollution mix, assessed in a separate chapter in this volume (chapter 19).

All combustion processes produce particles, most of which are small enough to be inhaled into the lung either as primary emissions (such as diesel soot), or as secondary particles via atmospheric transformation (such as sulfate particles formed from the burning of fuel containing sulfur). Their concentrations (in micrograms per cubic metre, or $\mu g/m^3$) are generally measured as inhalable and fine particles, PM_{10} and $PM_{2.5}$, respectively.[3] However, the total suspended particle mass (TSP) is still the only particle measurement available in many developing countries (Krzyzanowski and Schwela 1999).

Pollution from the combustion of fossil fuels is largely emitted into the outdoor air, but human exposure occurs both indoors and outdoors (Ozkaynak 1999). An individual's exposure to ambient urban air pollution depends on the relative amounts of time spent indoors and outdoors, the proximity to sources of ambient air pollution, and on the indoor concentration of outdoor pollutants. The indoor concentrations depend on factors such as the circulation of the indoor air and the degree to which constituents of the outdoor combustion mixture penetrate and persist in the indoor environment. Studies conducted largely in Europe and North America have shown that the fine particles generated from combustion outdoors both effectively penetrate and persist in many indoor environments. Gases, such as sulfur dioxide and ozone, may penetrate the indoor environment, but generally do not persist because of their reactivity. In some rural areas of developing countries, indoor cooking on unvented coal- or biomass-burning stoves is the most significant exposure to pollution from combustion sources. The burden of disease caused by

such exposure is addressed in chapter 18. The actual dose delivered to the lung or other organs will further depend on the type of pollutant, the breathing pattern and physical characteristics of the individual that determine the extent and site of deposition.

Governments in many parts of the world monitor ambient concentrations of air pollution as part of regulatory programmes designed to protect public health and the environment (Grant et al. 1999). The most extensive monitoring systems are in the United States and western Europe, where regular monitoring of ambient air quality has been in place since the mid-1970s. The most frequently and routinely monitored air pollutants include sulfur dioxide (SO_2), nitrogen oxides (NO_X, including NO and NO_2), carbon monoxide (CO), ozone (O_3), lead (Pb), black smoke (BS) or soot, and PM. National monitoring systems also exist in other parts of the world, but access to the data collected by these systems and international standardization of the monitoring methods are limited. The World Health Organization (WHO) Air Management Information System (AMIS) (WHO 2001c) collects the available information, but the reporting from many regions is poor, and for some regions there are no data in the WHO database. The various designs of the networks, differences in monitoring objectives and limited availability of the collected data for the outside users limit access to the information on population exposure in the greater proportion of the world's cities. In some parts of the world (e.g. in most of the countries of the former Soviet Union), the monitoring systems exist but do not provide the data necessary for assessment of the impact on health (Krzyzanowski and Schwela 1999). More details about the data available for this analysis are provided in further sections of this chapter.

These monitoring systems currently provide much of the data on exposure to urban air pollution that have been used in epidemiological research, although some studies establish their own monitoring networks when routinely-collected data are either unavailable or of poor quality, or to measure specific air pollution constituents, such as specific known carcinogens. Typically, monitoring sites are located in the city centre or throughout a given metropolitan area, in order to more accurately reflect the average residential exposure of the population. The data from monitors sited so as to measure emissions from specific sources, such as a local industry or heavy vehicular traffic, are frequently excluded from the data sets, as they may significantly deviate from the average levels of exposure experienced by the population.

Exposure estimates that rely exclusively on data from one or more stationary monitoring sites may provide inaccurate estimates of the short- and/or long-term average personal exposures of study populations (Navidi and Lurmann 1995; Zeger et al. 2000). The direction and magnitude of the errors that will be induced in estimates of the relative risk attributable to exposure to air pollution depend on the precision of the air quality monitoring data (or models used to generate the estimates of

the concentration of pollution), the applicability of one estimate to the entire target population and the correlation of the errors with the health outcome. Generally, such errors will be smaller for pollutants that tend to be uniformly distributed over large urban areas, and that penetrate efficiently indoors, both of these features being the case for fine PM produced by combustion. If the errors in the estimates of exposure are uncorrelated with the risk of the health outcome, then the estimates of relative risk attributable to air pollution will, in most cases, be too low (i.e. biased to the null) (Navidi and Lurmann 1995).

2.1 Definition of the air pollution metric for exposure variable

We selected PM_{10} and $PM_{2.5}$ as the indicators of exposure to urban air pollution from combustion sources. As noted above, PM is a ubiquitous component of the mixtures emitted into, and formed in, the ambient environment by combustion processes, and indicates the presence of these mixtures in outdoor air. Most importantly, these measures of particulate air pollution have been used in many epidemiological studies from around the world, of both mortality and morbidity of air pollution, and so provide the best overall indicator of exposure for our purposes (see section 3). Although other components of ambient air pollution from combustion sources are associated with these and other effects on health (Anonymous 1996a, 1996b), particulate air pollution has been found to be consistently and independently related to the most serious effects of air pollution, including daily and longer-term average mortality (California Air Resources Board 2002; Health Effects Institute 2001; U.S. Environmental Protection Agency 2002; WHO 2000a, 2003). There is some evidence, although much less than that for PM, linking ozone to premature mortality, particularly during the summer months (Abbey et al. 1999; Health Effects Institute 2000b). However, despite recent progress in developing models to estimate tropospheric (ground-level) ozone on a global scale, it was not currently feasible to derive the subregional estimates that would have been required for this project. In many developing countries, exposure to lead in the ambient air may also be of great consequence, having effects on mortality perhaps via effects on blood pressure. The impacts of lead in outdoor air are dealt with in chapter 19.

PM has been linked to serious effects on health after both short-term exposure (days to weeks), and more prolonged exposure (years), although there remains some uncertainty as to the distribution of induction times with regard to mortality (see below). We chose the annual average concentration(s) of PM as the exposure metric(s) because it corresponds to the time-scales of *a priori* interest for estimates of attributable and avoidable burden in the Global Burden of Disease (GBD) project, and because it was used to estimate the effects of exposure to PM in the key epidemiological study that provides our estimates of the concentration–response function.

2.2 Estimation of annual average concentrations of particulate matter

Air pollution measurements used in estimating annual average concentrations

The availability of measurements of ambient concentrations of PM varies widely across the globe, making estimation of annual average concentrations a considerable challenge (Krzyzanowski and Schwela 1999). To estimate ambient PM concentrations for all 14 subregions, we used a model (Global Model of Ambient Particulates [GMAPS]) recently developed at the World Bank to estimate concentrations of PM_{10} in cities, on the basis of available measurements of PM at population-oriented monitoring sites (Pandey et al. forthcoming). The model incorporates information on factors such as fuel mix, level of economic development, demographics and weather, in order to predict ambient concentrations of PM_{10} in urban residential areas. These estimates of PM_{10} were converted to $PM_{2.5}$ using available information on geographic variation in the ratio of $PM_{2.5}$ to PM_{10}. For each PM metric, the population-weighted subregional annual average was derived using the population of each city within each subregion in the year 2000.

The GMAPS model developed at the World Bank can be used to generate estimates of concentrations of PM_{10} in all world cities with populations of >100 000, and in national capitals. The estimation model is based on available measurements of PM_{10} and TSP from population-oriented monitoring stations in cities worldwide for the period 1985 to 1999, retrieved in October 2001. In all cases, data from a monitoring site were included if and only if it was clearly identified as a residential or mixed residential site (see section 2.3 for definition). For instance, city averages reported for many Chinese cities (National Environmental Protection Agency of China 2000) were not included in the model estimation because the location of these sites could not be ascertained.

In principle, the monitoring data used for calculation of annual averages should be collected throughout the year, since seasonal patterns in the data are fairly common. More than 85% of cities in Europe and the United States collect measurements of PM throughout the year. The representativeness of the data for cities in other parts of the world could not be confirmed. In addition, in many countries where PM was measured throughout the year, it was only measured on every sixth day. The methods for measuring concentrations of PM also varied, both gravimetric and automatic methods (tapered element oscillating microbalance monitors [TEOMS] or beta gauge monitors) being included.

Most of the data on annual average ambient concentrations used in the model come from AMIS (WHO 2001c). This information is submitted to WHO by national environmental agencies and air quality control authorities, which perform these measurements using nationally

approved methods and standards of data quality. The data set contains the annual mean concentration of selected air pollutants, including PM, by monitoring site. Additional data, such as 95th percentiles of daily means, are also available for some sites. Although WHO requests that all Member States provide data for compilation in the AMIS database, the reported data are still limited because many countries do not have air quality monitoring networks. Additionally, some countries with monitoring networks may not report the data because of poor data quality or limited ability to process and report the data.

The data from AMIS were supplemented with other sources of data on TSP and PM_{10} from monitoring sites. These included data for European cities collected by WHO/European Centre for Environment and Health (ECEH) for the Health Impact Assessment of Air Pollution (HIAAP) project in 1999 from both national and local environmental agencies (WHO 2001a), data for Canadian cities provided by Environment Canada (www.ec.gc.ca) and statistics Canada (http://www.statcan.ca/english/ads/cansimII/index.htm), and data for cities in the United States from the U.S. Environmental Protection Agency AIRS database (Aerometric Information Retrieval System 2001). Data for Chinese cities were also obtained from the Environmental Quality Reports from China (National Environmental Protection Agency of China 2000), and Mexican cities from the Instituto Nacional de Ecología (INE), SEMARNAP, Mexico (Instituto Nacional de Ecología 2000). Additional data were also obtained from the World Bank URBAIR studies of air pollution in Jakarta and Kathmandu (Grønskei et al. 1997a, 1997b). To limit undue influence of the data from cities in the United States, data used from the United States AIRS database were limited to the years 1996–1999.[4]

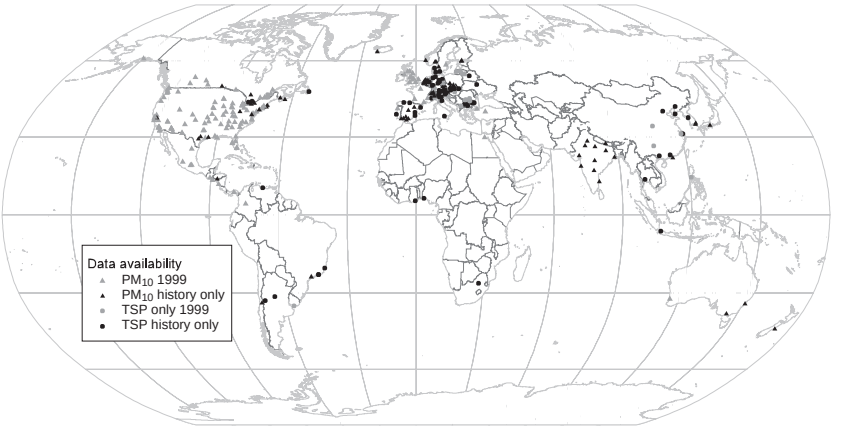
Measured annual average concentrations of PM_{10} and TSP data from monitoring sites were available for 512 unique locations in 304 cities in 55 countries over the period 1985–1999, and provided 1997 time–location data points. For some sites and years, data on both TSP and PM_{10} were available, yielding a total of 2344 individual observations.[5] The number of cities with measured data on PM from monitoring sites in each subregion and for each year by PM measure is shown in Table 17.1. A total of 304 cities reported either the annual average concentrations of PM_{10} or TSP for at least 1 year between 1985 and 1999. Of these, 51 cities reported both PM_{10} and TSP while 165 cities, mostly in North America and western Europe, reported PM_{10} only, and the remaining 88 cities reported data for TSP only.

Coverage of cities and populations with data from monitoring sites varies significantly across different subregions (Figure 17.1). For instance, data from monitoring were available for fewer than two cities for six of the subregions, AFR-D, AFR-E, AMR-D, EMR-B, EMR-D and SEAR-B. In contrast, data from monitoring sites were available for 218 cities in NAWE, of which 174 report data on PM_{10}. The 304 world cities

Table 17.1 Number of cities for which data on particulate matter are available from monitoring sites, by subregion, year and type of particulate matter

	PM_{10} or TSP	PM_{10}	TSP
Subregion			
AFR-D	2	0	2
AFR-E	1	0	1
AMR-A	123	118	25
AMR-B	19	12	12
AMR-D	2	2	2
EMR-B	0	0	0
EMR-D	1	1	0
EUR-A	95	56	43
EUR-B	22	7	17
EUR-C	7	1	7
SEAR-B	2	0	2
SEAR-D	11	11	10
WPR-A	5	5	4
WPR-B	14	3	14
World	304	216	139
Year			
1985	28	7	28
1986	52	15	50
1987	53	9	52
1988	47	16	45
1989	53	17	51
1990	64	20	60
1991	63	30	60
1992	70	34	67
1993	73	41	68
1994	78	40	73
1995	73	42	68
1996	156	132	54
1997	144	127	40
1998	211	150	81
1999	166	143	40
1985–1998	267	187	127

Figure 17.1 Cities from which data on exposure to PM_{10} or TSP during 1985–1999 are available from monitoring cites



Source: K. D. Pandey, Personal Communication.

with data from monitoring account for 9% of the total number of cities with a population of >100 000 worldwide and have a combined population in the year 2000 of around 559 million, or about 28% of the global urban population (Table 17.2).

Global Model of Ambient Particulates (GMAPS)

The GMAPS model econometrically estimates a fixed-effect model of the concentrations of urban ambient PM using the latest available data from WHO and other sources, as outlined above. The estimating Equation 1 focuses on the anthropogenic sources of pollution and the capacity of the natural environment to generate, disperse and dissipate pollutants.[6] Its determinants include the scale and composition of economic activity, the energy mix, the strength of local regulation of pollution, and geographic and atmospheric conditions that affect the transport of pollutants.

$$\begin{aligned} C_{ijkt} = & \sum_{k=1}^{K} \beta_k Z_k + \sum_{f=1}^{F} \beta_{Ef} E_{fkt} + \sum_{g=1}^{G2} \beta_{Mg} M_{gik} + \beta_R R_{kt} + \beta_N N_{jkt} + \beta_D D_{jk} \\ & + \beta_{Scale} Scale_{jkt} + \beta_Y Y_{kt} + \beta_T Trend_{ijkt} + \beta_{YT} Y_{kt} Trend_{ijkt} \\ & + \theta_S S_{ijkt} + \theta_{Scale} S_{ijkt} Scale_{jkt} + \theta_Y S_{ijkt} Y_{kt} + \theta_T S_{ijkt} Trend_{ijkt} \\ & + \theta_{YT} S_{ijkt} Y_{kt} Trend_{ijkt} + \sum_{g=1}^{G1} \theta_{Mg} S_{ijkt} M_{gjk} + \varepsilon_{ijkt} \end{aligned} \quad (1)$$

Table 17.2 Cities for which measurements of particulate matter are available from monitoring sites, by subregion

	Number of cities			*Urban population (000s) in 2000*[a]		
Subregion	*Cities in subregion*	*Cities with monitoring sites*	*% with monitoring sites*	*Cities in subregion*	*Cities with monitoring sites*	*% with monitoring sites*
AFR-D	107	2	2	66 960	14 914	22
AFR-E	105	1	1	68 367	2 388	3
AMR-A	267	123	46	232 439	178 240	77
AMR-B	399	19	5	217 159	64 121	30
AMR-D	47	2	4	29 512	3 291	11
EMR-B	89	0	0	56 621	0	0
EMR-D	126	1	1	99 397	8 124	8
EUR-A	429	95	22	161 808	79 160	49
EUR-B	182	22	12	81 756	21 494	26
EUR-C	275	7	3	109 178	7 670	7
SEAR-B	68	2	3	53 708	18 793	35
SEAR-D	356	11	3	214 175	67 081	31
WPR-A	242	5	2	100 079	11 459	11
WPR-B	519	14	3	528 318	81 817	15
World	3 211	304	9	2 019 479	558 553	28

[a] The total urban population is for 3211 cities with populations >100 000 and national capitals.
The total urban population in 2000, including cities of all sizes, is 2.8 billion.

where

C_{ijkt} = log of concentration of PM in monitoring station i, city j, country k, at time t

Z_k = binary variable for country k

E_{fkt} = log of per capita energy consumption of energy source type f for country k at time t (f=1 . . . F)

M_{gjk} = log of meteorological/geographic factor g for city j, country k (factors g=1 . . . $G1$ affect PM_{10} concentration in a different way than TSP concentration. Factors g=$G1$+1 . . . $G2$ do not make a distinction between PM_{10} and TSP)

R_{kt} = log of population density of country k at time t

N_{jkt} = log of population of city j, country k, at time t

D_{jk} = log of local population density in the vicinity of city j in country k

$Scale_{jkt}$ = log of scale of economy (intensity of economic activity) for city j, country k at time t

Y_{kt} = log of income per capita (1-year lagged 3-year moving average) of country k at time t

$Trend_{ijkt}$ = time trend (1985=1, 1986=2, . . . 1999=15)

S_{ijkt} = binary variable for PM type measured at monitoring station i, city j, country k, at time t, (1=TSP, 0=PM_{10}), and

the β_S and θ_S are the parameters that are estimated by the model.

Equation 1 jointly determines the concentrations of total suspended particulate matter (TSP) and inhalable particulates (PM_{10}) in residential areas. Most cities in developing countries only monitor TSP and not PM_{10}. Adoption of the pooled specification permits use of all available data and provides better information about the concentrations of PM, especially for cities in developing countries. Limiting the estimation sample to PM_{10} observations is sensible only if knowledge of the concentration of TSP in a city makes no contribution to predicting PM_{10}. Since PM_{10} comprises the smaller size particles within TSP, this assumption is clearly unreasonable. The pooled specification allows for separate estimation of concentrations of PM_{10} and TSP for each city by setting the binary variable, S_{ijkt}, equal to zero or one, as shown in Equations 2 and 3.

$$\log[PM10_{ijkt}] = \sum_{k=1}^{K} \beta_k Z_k + \sum_{f=1}^{F} \beta_{Ef} E_{fkt} + \sum_{g=1}^{G2} \beta_{Mg} M_{gjk} + \beta_R R_{kt} + \beta_N N_{jkt} + \beta_D D_{jk} + \beta_{Scale} Scale_{jkt} + \beta_Y Y_{kt} + \beta_T Trend_{ijkt} + \beta_{YT} Y_{kt} Trend_{ijkt} \quad (2)$$

$$\begin{aligned}\log[TSP_{ijkt}] = &\sum_{k=1}^{K}\beta_k Z_k + \sum_{f=1}^{F}\beta_{Ef}E_{fkt} + \sum_{g=1}^{G2}\beta_{Mg}M_{gjk} + \beta_R R_{kt} + \beta_N N_{jkt} \\ &+ \beta_D D_{jk} + \beta_{Scale}Scale_{jkt} + \beta_Y Y_{kt} + \beta_T Trend_{ijkt} + \beta_{YT}Y_{kt}Trend_{ijkt} \\ &+ \theta_S + \theta_{Scale}Scale_{jkt} + \theta_Y Y_{kt} + \theta_T Trend_{ijkt} + \theta_{YT}Y_{kt}Trend_{ijkt} \\ &+ \sum_{g=1}^{G1}\theta_{Mg}M_{gjk}\end{aligned} \tag{3}$$

To reduce undue influence from extreme values, all of the continuous variables in the model were specified in log form and each exogenous variable in the estimation sample was truncated to the middle 98% range observed in the estimation sample.

The estimation Equation 1 includes country-specific binary variables, Z_k, to control for economic, social and natural factors that are not captured by the other explanatory variables. These include differences in the quality of the data on ambient concentration and in collection methods across countries, the degree of regulatory heterogeneity within a country, the relative importance of intercity transport, proximity of and pollution levels in neighbouring cities and the composition of economic activity. The country-specific binary variables measure the average concentration of PM in each country during the 15-year period 1986–1999, controlling for variations within the country caused by factors accounted for in the remainder of the estimating Equation 1. In contrast, the rest of the estimation model (1) explains the marginal contribution of the included factors to deviations in the ambient concentration in the city from this average.

The primary determinants of the observed variations in the ambient concentrations of PM within a country in the estimation model are:

Energy consumption. The model includes six separate per capita energy consumption categories—coal, oil, natural gas, nuclear, hydroelectric, combustible renewables and wastes—that account for all energy consumed in each country for which data are available from the International Energy Agency's (IEA) Annual Energy Balance database (International Energy Agency 2001a, 2001b). The separate inclusion of each type of energy source accounts for differences in emission factors, variations in economic activity and intensity of fuel use across countries. In addition, the model also includes per capita consumption of petrol and diesel used in the transportation sector, also available from IEA's database, to capture additional detail about one of the most significant contributors to ambient concentrations of PM.

Meteorological and geographic factors. The model includes 22 atmospheric and geographic factors for each city to account for both the dissipative/dispersive capacity of the natural environment and natural sources of particulates, such as desert dust storms, forest fires and sea spray. These include a suite of 18 climatic variables representing the long-

term average climatic conditions related to local atmospheric conditions and transport of PM, consisting of the annual average (average of the monthly data) and seasonal changes (measured as the standard deviation of the monthly data) for the following nine factors: mean temperature, diurnal temperature, mean precipitation, barometric pressure, wind speed, percentage cloud cover and frequency of wet, sunny and frosty days (New et al. 1999).[7] In addition, two meteorological variables related to energy demand (heating and cooling degree-days) are estimated for each city from the mean monthly temperature. Two topographical variables related to atmospheric transport—distance from the city centre to the nearest point on the coastline, calculated using the geographic information system (GIS), and elevation of the city, derived from a global digital elevation model (USGS 1996)—are also included in the model.

City and national population and national population density. These variables provide measures of the scale and intensity of the pollution problem in each city. The data on population comes from the *Demographic yearbook* published by the United Nations (UN 2001).

Local population density. The local population density in the vicinity of each city provides a measure of the intensity of pollution. It is estimated from the Gridded Population of the World (version 2), available from the Consortium for International Earth Science Information Network (CIESIN 2000). This data set provides the best available population data for about 120 000 administrative units, converted to a regular grid of population counts at a resolution of about 5 km. The local population density in the vicinity of each city is the average population density for all grid cells within a 20-km radius of the city centre.

Local intensity of economic activity. Most cities do not collect data on the amount or composition of economic activity. Instead, the local gross domestic product (GDP) per square kilometre computed as the product of the national per capita GDP and the local population density in the vicinity of each city is used as a proxy for the intensity of economic activity within each city (World Bank 2002).

National income per capita. This variable is used to capture the following national indicators: valuation of the quality of the environment, strength of environmental policy and regulation, the institutional capacity to enforce environmental policies, and the potential use of cleaner fuels along the fuel-use chain as countries develop. It is measured as a 1-year lag of the average of the previous 3 years (World Bank 2002).

Time trends. The model includes two time-trend variables (with 1985=1 . . . 1999=15) to allow for differential time trends for PM_{10} and TSP particulate pollution. Both of these variables are in turn interacted with lagged national per capita income to allow trends to vary across countries on the basis of differential valuation and improvements in environmental quality across countries as measured by the level of economic development. These trends measure changes in concentrations of PM

that are caused by factors not already captured in the model, such as technological changes, improvements in knowledge and structural shifts in the composition of economic activity. They do not represent the unconditional aggregate trends in concentrations of PM.

Binary variable to differentiate PM_{10} and TSP. The model includes a binary variable indicating whether PM is measured as TSP or PM_{10}. This binary variable is also allowed to interact in the model with other variables to allow for size class differences in the composition of particulates across cities and countries. It provides a better representation of intercity differences across the world, rather than assuming a uniform relationship across all cities. The log of the ratio of PM_{10} to TSP in each city can be estimated by subtracting Equation 3 from 2, as shown in Equation 4. The key determinants of this ratio are the scale of economic activity, differential trends across countries, level of economic development and strength of environmental policy, and the subset of meteorological variables that are directly related to particle size (annual mean and seasonal variations in wind speed, precipitation and frequency of wet days).

$$\log[PM10_{ijkt}] - \log[TSP_{ijkt}] = -\theta_S - \theta_{Scale} Scale_{jkt} - \theta_Y Y_{kt} - \theta_T Trend_{ijkt} - \theta_{YT} Y_{kt} Trend_{ijkt} - \sum_{g=1}^{G1} \theta_{Mg} M_{gjk} \quad (4)$$

In order to facilitate predictions for countries not included in the estimation, a secondary model shown in Equation 5 is estimated to explain the average level of ambient PM concentration in each country.

$$\hat{\beta}_k = \sum_{f=1}^{F} \gamma_{Ef} \overline{E}_{fk} + \gamma_R \overline{R}_k + \gamma_R \overline{Y}_k + u_k \quad (5)$$

where

$\hat{\beta}_k$ = country-specific binary variable coefficient estimated in Equation 1

$\overline{E}_{fk}$ = log of average per capita energy consumption of energy type f for country k during 1985–1999 ($f = 1 \ldots F$)

$\overline{R}_k$ = log of average population density of country k during 1985–1999

$\overline{Y}_k$ = log of average national per capita income of country k during 1985–1999 (1-year lagged average of previous 3 years)

This secondary model (5) explains the average level of pollution under reference conditions for a country, on the basis of the scale of the economy, the composition of economic activity as measured by the energy mix, and the strength of local pollution regulations and the institutional capacity for implementing these regulations.

2.3 Model outputs

The GMAPS model is designed to obtain the best city-level prediction of concentrations of PM for a wide range of cities on the basis of the limited amount of data from monitoring available, so it focuses on increasing the fit of the model. It is not designed to provide a causal model for ambient concentrations of air pollution. The estimation model (1) explains 88% of the variation in the observed data from monitoring, indicating a good fit (Pandey et al. forthcoming). The overall correlation between the measured and the predicted data is around 0.9 for both PM_{10} and TSP observations (see Table 17.3), and is >0.80 for all years and for both observations of PM_{10} and TSP, with the exception of PM_{10} in 1985. The correlation by subregion is smaller than that over time, ranging between 0.2 and 0.9 for subregions with more than 10 data points. The correlations for subregions with fewer data points are smaller than 0.2 and are less precisely estimated. A negative correlation for EUR-B is driven by a single erroneous observation for Bucharest, Romania, where the observed concentration of PM_{10} is higher than that of TSP. These results originated from two different monitoring locations; had the model been re-estimated without this particular PM_{10} observation, the correlation for the subregion would have been 0.32.

Subregion- and PM type-specific scatter plots of model predictions compared to actual data also show a clustering of points around the solid line drawn at a 45° angle, indicating that the actual values are close to the predicted values. As would be expected, the predicted values are less extreme than the actual values at both tails, owing to the truncation of all explanatory variables to the middle 98% range of the estimation sample. F-tests revealed that all of the eight aggregate factors in the model added significant explanatory power to the regression.

The secondary estimation model (5) explains 85% of the variation in the estimated average level of pollution in a country, indicating that this model provides a good fit. The explanatory power of the secondary model is not as robust to changes in the estimation sample owing to significant uncertainties in the estimated dependent variable.

Out-of-sample predictions were used to validate the model using both statistical and heuristic criteria. The model was re-estimated using subsamples of the data on the basis of different cut-off points for per capita income, to examine the appropriateness of extrapolating from a model primarily based on industrialized cities in North America and western Europe to cities in developing countries. The resulting estimates from the model were used to predict concentrations of PM_{10} in residential areas in the out-of-sample cities located in developing countries. A second set of estimates was also made comprising income-based subsamples using only the available data on PM_{10} from monitoring in residential sites. These validation estimates consistently showed that out-of-sample correlations were higher when data on TSP were included in the estimations. Furthermore, the out-of-sample correlations on aggregate ranged

Table 17.3 Correlation between observed concentrations of particulate matter at monitoring sites and predictions by subregion, year and type of particulate matter

	PM_{10} or TSP		*PM_{10}*		*TSP*	
	No. of observations	*Correlation*	*No. of observations*	*Correlation*	*No. of observations*	*Correlation*
Subregion						
AFR-D	6	0.86	0	NA	6	0.86
AFR-E	2	–1.00	0	NA	2	–1.00
AMR-A	1 273	0.79	938	0.59	335	0.67
AMR-B	361	0.80	215	0.52	146	0.75
AMR-D	34	0.88	18	0.31	16	0.72
EMR-D	1	NA	1	NA	0	NA
EUR-A	182	0.85	75	0.82	107	0.73
EUR-B	63	0.83	16	–0.29	47	0.78
EUR-C	54	0.84	1	NA	53	0.83
SEAR-B	9	0.14	0	NA	9	0.14
SEAR-D	158	0.81	65	0.69	93	0.80
WPR-A	69	0.85	36	0.86	33	0.20
WPR-B	132	0.92	21	0.49	111	0.89
World	2 344	0.94	1 386	0.89	958	0.92
Year						
1985	35	0.95	7	0.11	28	0.85
1986	68	0.93	17	0.81	51	0.92
1987	65	0.93	11	0.94	54	0.92
1988	70	0.93	20	0.91	50	0.92
1989	76	0.94	21	0.90	55	0.94
1990	91	0.94	24	0.90	67	0.94
1991	101	0.96	34	0.94	67	0.95
1992	116	0.94	38	0.95	78	0.93
1993	130	0.94	49	0.95	81	0.92
1994	138	0.94	46	0.94	92	0.93
1995	144	0.94	59	0.94	85	0.93
1996	330	0.92	259	0.88	71	0.90
1997	298	0.89	253	0.79	45	0.92
1998	377	0.88	289	0.84	88	0.84
1999	305	0.90	259	0.83	46	0.87
All years except 1999	2 039	0.94	1 127	0.90	912	0.92

NA Not applicable.

between 0.40 and 0.59, based on the income cut-off used, and lend support to the modelling approach.

Since cities with data from monitoring are not representative of all cities and account for a small fraction of urban residents in developing countries, the following heuristic criteria were also used to evaluate the predictions of the model.

- *Comparison of the relative variation of the predictions within countries and between countries relative to the actual data*: The model predictions exhibited significant variations both across countries and across cities within a country. The predicted variations within a country were about 60% of those between countries and were comparable to the corresponding variations in the actual data.
- *Number of cities for which predictions were outside the range of the estimation sample*: The predictions for PM_{10} were within the range observed in the actual data. They continued to be within bounds when the same fractions of values are removed from the tails of the estimated and measured data.
- *Magnitude of predictions outside the range observed in the estimation sample*: Of the 304 cities with data from monitoring, concentrations of PM_{10} exceeded 200 $\mu g/m^3$ in three cities and concentrations of TSP exceeded 400 $\mu g/m^3$ in 10 cities. The predicted concentrations of PM_{10} exceeded 200 $\mu g/m^3$ in only four out of 3226 cities.
- *Range of the PM_{10}:TSP ratio*: The PM_{10}:TSP ratio predicted by the model is between 0.24 and 0.98 and spans the middle 95% of the range observed in the actual data. The mean ratio predicted by the model is 0.49; the ratio for half of the cities lies between 0.39 and 0.56.
- *Comparison of the uncertainty in estimates for cities, relative to the amount of available information for neighbouring cities*: Bootstrap error estimates of the prediction error for the city showed that the confidence intervals were wider for cities with no data from monitoring and are largest in the countries with no data from monitoring.

The robustness of the model was tested using alternative specifications of the model based on the goodness-of-fit of the model and the heuristic criteria outlined above. The alternative models that were considered were:

- *Linear model*. The linear model provides undue weight to the extreme values in the explanatory variables, resulting in predictions that are orders of magnitude larger than those for cities with data from monitoring.

- *Explanatory variable truncation.* The model was re-estimated with four different levels of truncation for the explanatory variables: no truncation, truncation to the actual range for the cities with data from monitoring, truncation to the middle 98% range of all explanatory variables for these cities, and truncation to the middle 90% range for these cities. Estimates based on the first two of these were sensitive to some of the extreme data points in the estimation sample, resulting in large variations in the predictions. Estimates from the last truncation were rejected because more than one quarter of the observations were truncated, leading to a poorer model fit.
- *Energy consumption variables.* The model was re-estimated with three alternative measures for the energy consumption variables: energy consumption per area, total per capita energy consumption and share of each energy type in the total energy mix, and the product of national per capita energy consumption by energy type and city population density. The specification per area resulted in predictions that were unstable and orders of magnitude larger than those observed in any city because of truncation of extreme values in countries with missing data on fuels. The second and third specifications resulted in poorer fits with over-predictions for >100 cities with values outside of the observed range of concentrations of PM_{10}.
- *Income.* The model was estimated using income-squared and income-cubed terms to measure the impact of national per capita income. Higher order terms were unstable and resulted in predictions that were orders of magnitude larger than those observed in any city.
- *PM_{10}:TSP ratio.* A number of different models were estimated from full interactivity of the binary variable S_{ijkt} with all of the continuous variables, to no interactivity with the continuous variables. The full interactivity model was rejected because it predicted physically implausible PM_{10}:TSP ratios of 2 for a significant number of cities. The limited model with no interactivity was rejected because it over-predicted the results for many cities in the Middle East and North Africa that contain a larger fraction of wind-blown coarser particles. Other models were estimated that incrementally added groups of variables, such as energy type and the other climate variables. These were all rejected using the heuristic criteria outlined above.
- *Location of monitoring sites.* The sensitivity of the model predictions to the inclusion of mixed residential sites was examined by re-estimating the model using only pure residential sites.[8] Although estimates for some individual cities change in significant ways, predictions at the subregional level, and for most countries, are not statistically different as compared to when mixed sites are included.

- *Inclusion of non-residential sites*. A more inclusive model, which jointly estimates concentrations of PM in residential and non-residential sites indicated that most model parameters were relatively stable and that the model predictions for subregional residential concentrations of PM_{10} were not significantly different for most subregions.
- *Additional monitoring data*. The sensitivity of the model was tested to the inclusion of additional monitoring data that became available between October 2001 and July 2002. The aggregate PM predictions were not statistically different for all subregions, except for EMR-B where concentrations of PM increase by nearly 50%. This is primarily owing to the inclusion of data for Kuwait City, which is the only city for which data from monitoring sites are available in this subregion.
- *Influential data point*. The sensitivity of the model predictions to influential data points was examined using bootstrap error techniques. Variations in the predictions based on different subsamples of the data were used to estimate the degree of uncertainty in the model estimates.

Estimating ambient concentrations of particulate matter in cities

The average subregional ambient concentrations used in this work are estimated from the city-level model predictions for 1999, the latest year for which all of the explanatory variables were available. The estimates of concentrations of PM_{10} in each city for 1999 were generated using a three-step approach. First, for all cities located in countries with at least one population-oriented monitoring site, the concentration of PM_{10} was estimated using the GMAPS model, as specified in Equation 1. The concentration of PM_{10} cannot be estimated using Equation 1 alone for cities in every country, because the average level of pollution in the country as measured by the country binary variable was not available for countries without monitors. Therefore, in the second step, the secondary Equation 5 was used to predict the country coefficient for countries without monitors. These predictions were combined with estimates from Equation 1 that explain variations around the average level to generate 1999 concentrations of PM_{10} for these cities in these countries.

Finally, for cities with actual data from monitoring, a best estimate of concentration of PM_{10} in the city in 1999 was generated by incorporating information on concentrations from previous years. Specifically, an average residual for each city was determined by comparing each year-specific predicted value generated by the estimation model (1) with the actual monitored value for that year and city. This served to adjust the model predictions for local factors that are known but unmeasured in the model, such as the composition of local economic activity. Given the

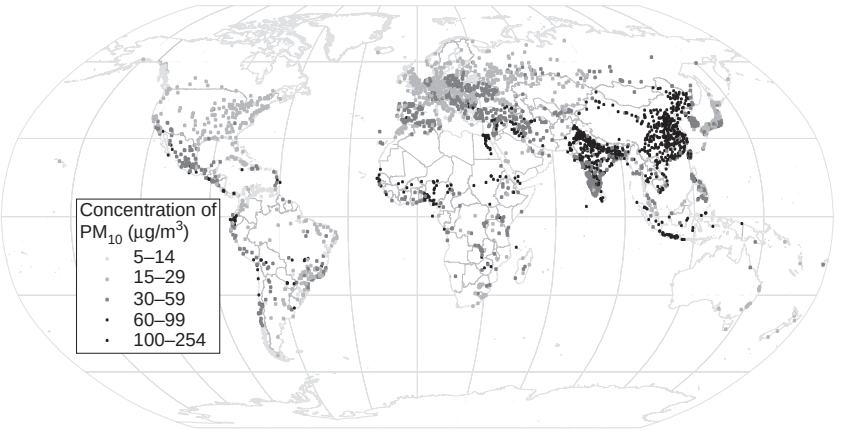
large year-to-year variations in the available measured data even at the same monitoring station, correcting for the average residual provides a better representation of long-term average factors affecting concentrations of PM in a city than using the actual monitored value for the last year of data from monitoring alone.

ESTIMATING SUBREGIONAL AMBIENT CONCENTRATIONS OF PARTICULATE MATTER

To avoid extrapolating outside the sample frame, all exogenous variables were truncated to the range used in the estimation sample. When necessary, missing explanatory variables for the country were filled in with the median values for economically similar countries located in the same geographic area. For most subregions, data were available for at least 95% of the cities, accounting for at least 95% of the population in each subregion. In contrast, data on either fuel, GDP or gross national product (GNP) were missing for 20–30% of the cities, accounting for 20–30% of the population for each of the four subregions AFR-D, AFR-E, EMR-B and EMR-D. In all, data on either fuel, GDP or GNP were completed in this way for 176 cities worldwide, accounting for 5% of the total world urban population.

The estimated annual average concentrations of PM_{10} in urban areas for world cities with populations of >100 000 and national capitals are shown in Figure 17.2. Each circle on the map represents a city and is shaded according to the estimated concentrations of PM_{10} in that city. Standards currently in place in North America and western Europe lie

Figure 17.2 Estimated annual average concentrations of PM_{10} in cities with populations of >100 000 and in national capitals



Source: Map provided by Kiran Dev Pandey, World Bank.

Table 17.4 Population-weighted predicted PM_{10} and TSP and percentiles of the distribution of estimated concentrations of PM_{10}

	Predicted point estimate ($\mu g/m^3$)			*Percentiles of the distribution of estimated PM_{10} ($\mu g/m^3$)*				
Subregion	*PM_{10}*	*TSP*	*PM_{10}/TSP*	*5%*	*25%*	*50%*	*75%*	*95%*
AFR-D	68	195	0.350	32	43	61	72	84
AFR-E	39	104	0.372	30	35	39	44	58
AMR-A	25	39	0.642	24	25	25	25	25
AMR-B	37	79	0.470	35	36	38	39	42
AMR-D	51	146	0.349	37	43	48	53	58
EMR-B	40	118	0.341	23	30	34	39	48
EMR-D	110	276	0.397	62	78	99	110	127
EUR-A	26	49	0.531	25	26	26	27	28
EUR-B	48	118	0.406	41	44	46	48	50
EUR-C	31	90	0.340	21	25	29	33	38
SEAR-B	108	245	0.439	39	86	105	129	151
SEAR-D	84	206	0.409	73	80	84	88	96
WPR-A	32	50	0.646	27	30	32	34	37
WPR-B	89	221	0.403	73	83	89	96	104
World	60	144	0.417	51	56	58	62	65

between 30–60 $\mu g/m^3$. Therefore, we defined a middle group with concentrations of PM_{10} in the range of 30–60 $\mu g/m^3$. Cities with values that fell outside this range were sorted into two groups of cities with higher concentrations and two groups with lower concentrations (thus forming a total of five groups). Worldwide, about 30% of the urban population live in the less polluted cities while 40% live in the more polluted cities. The remaining 30% of people live in cities with concentrations of PM_{10} in the middle range. However, there are significant regional differences. More than 70% of the people living in NAWE live in cities with concentrations of less than 30 $\mu g/m^3$, meeting the most stringent standards. In contrast, more than 70% of the populations in SEAR-D, WPR-B, EMR-D and SEAR-D live in cities where concentrations exceed even the most lenient standards.

This difference can also be seen in the estimated population-weighted mean concentrations of PM_{10} for each subregion, which are presented in Table 17.4. These are computed from 1999 estimates of concentrations of PM_{10} in cities, using the populations of each city in 2000 as weights. We have not directly used data for cities with data from monitoring in computing the subregional averages, to avoid incorporating short-term transitional variations into our exposure estimates.[9] The mean exposures in the most polluted subregions (EMR-D, SEAR-B, SEAR-D and

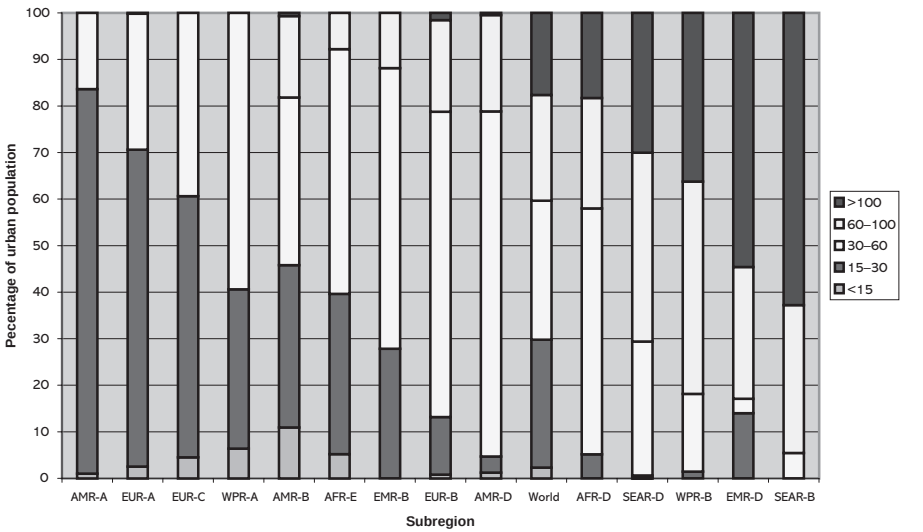
WPR-B) are about three times higher than those in the least polluted subregions (AMR-A and EUR-A). The table also shows predictions of population-weighted average concentrations of TSP and the size composition of PM for each subregion. Finer particles account for a larger share of PM in the highly industrialized countries of AMR-A and EUR-A compared to the other less industrialized subregions.

We quantified the uncertainty in our estimates of the subregion-specific mean concentrations of PM using a bootstrap technique. In this method, the model is re-estimated many times (200 trials) using a randomly repeated sample of the observations used in estimating the model. For each trial, city and population-weighted subregional predictions of PM are generated using the methods described above. The predictions from all trials are sorted from highest to lowest to obtain the percentile distribution of concentrations of PM_{10} for each subregion and are also shown in Table 17.4. The degree of certainty in the point estimates of concentration of PM_{10} for each subregion is directly related to the number of observations available from monitoring of PM. For example, the two subregions (AMR-A and EUR-A) with the most frequently monitored cities also have the smallest confidence intervals for PM_{10} values. In contrast, the five subregions with two or fewer cities that are monitored (AFR-E, AFR-D, EMR-B, EMR-D and SEAR-B) have larger confidence intervals than the other subregions. The width of the confidence intervals for these subregions depends on the geographic and climatic similarity of their cities with monitored data. For example, confidence intervals for AFR-E and EMR-B are about half of those for AFR-D and for EMR-D.

The estimates also show that substantial differences exist in the average concentration of PM within each subregion. The share of the urban population in cities with populations >100 000 and in national capitals according to estimated concentrations of PM_{10} is given in Figure 17.3. All cities in AMR-A, EUR-A, EUR-C and WPR-A are estimated to have concentrations of PM_{10} of $<60\,\mu g/m^3$. In contrast, 95% of the urban population in SEAR-B and about 82% of the urban population in WPR-B and EMR-D are exposed to $>60\,\mu g/m^3$ PM_{10}. We also estimate that a high proportion of the urban population in SEAR-D is exposed to high annual average concentrations of PM_{10}.

Since some of the health outcomes are based on $PM_{2.5}$, rather than PM_{10}, concentrations for this pollutant had to be estimated. City-specific concentration of $PM_{2.5}$ was estimated as a fixed proportion of PM_{10}. Available measurements indicate that the ratio of $PM_{2.5}$ to PM_{10} ranges from 0.5 to 0.8 in many urban areas in developed countries, (California Air Resources Board 2002; U.S. Environmental Protection Agency 2002). Limited evidence suggests that a similar ratio may exist in large cities in other subregions. For example, a recent study from China reports the $PM_{2.5}$:PM_{10} ratio to be in the range of 0.51 to 0.72 in four urban locations (Quian et al. 2001). However, in areas impacted

Figure 17.3 Distribution of the urban population according to estimated concentrations of PM_{10} in cities with populations of >100 and in national capitals, by subregion



by more crustal particles (e.g. arid areas or cities with a significant number of unpaved roads or windy days), the ratios are likely to be much lower. These areas will have a greater proportion of PM_{10} in the coarse size range of 2.5–10μm. For example, evidence from the Coachella Valley (i.e. the Palm Springs area), an arid range of southern California suggests that the $PM_{2.5}$:PM_{10} ratio is 0.35 (Ostro et al. 1999b). Therefore, we assumed a ratio of 0.5 for our base case and have examined the sensitivity of our results to this assumption. Specifically, for our sensitivity analysis, for cities in AMR-A, EUR-A, EUR-B, EUR-C and WPR-A (including the United States, Canada, all Europe, Japan, Singapore, Australia and New Zealand), a higher scaling factor of 0.65 was used, assuming relatively more combustion-related particles, while a lower scaling factor of 0.35 was used for cities in all other subregions.

Estimates of the annual average population-weighted concentration of $PM_{2.5}$ for each subregion were calculated in a similar manner to that for PM_{10}, using the estimated concentration of $PM_{2.5}$ for the city in 1999 and the population for each city in 2000.

2.4 Choice of the theoretical-minimum-risk exposure

Studies of mortality associated with both short- and long-term exposure to PM, discussed below, have been unable to detect a threshold below

which there is no effect of exposure. For most results presented below, we estimated the burden of disease with respect to a counterfactual concentration of 7.5 $\mu g/m^3$ $PM_{2.5}$ (or 15 $\mu g/m^3$ PM_{10}). This value is close to the lowest concentration observed in the epidemiological study (Pope et al. 2002) from which we derived the concentration–response functions used for the majority of our estimates. This choice avoids extrapolating the concentration–response function(s) below the concentrations actually observed in the epidemiologic studies from which they were derived, although health benefits may well accrue from reductions below those concentrations.

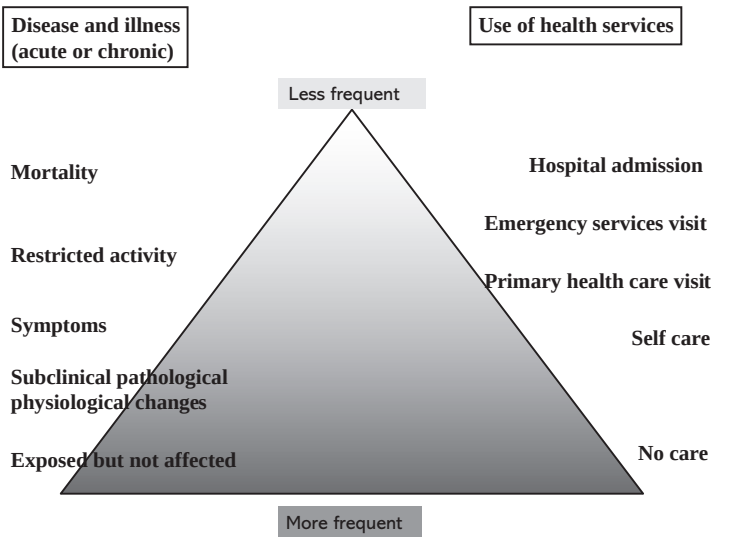
We were aware, however, that for some cities the estimated (and observed) concentrations of PM are lower, e.g. in AMR-A (United States and Canada), and that achieving such concentrations more widely would be not only desirable, but also feasible in some settings (U.S. Environmental Protection Agency 2002). Moreover, previous impact estimates have been sensitive to where this value was set (Künzli et al. 2000). Therefore, we also conducted sensitivity analyses in which the theoretical minimum concentration was halved and doubled (see below).

3. Health effects of exposure to urban air pollution

The past 10–15 years have seen a rapid increase in research on the health effects of air pollution, and it is now widely accepted that exposure to urban air pollution is associated with a broad range of acute and chronic health effects, ranging from minor physiological disturbances to death from respiratory and cardiovascular disease (Anonymous 1996a, 1996b; Figure 17.4). Recently, a committee of the American Thoracic Society identified effects on respiratory health associated with air pollution, which should be considered adverse, spanning outcomes from death from respiratory diseases to reduced quality of life, and included some irreversible changes in physiological function (American Thoracic Society 2000). In general, the frequency of occurrence of the health outcome is inversely related to its severity, with the consequence that assessing total health impact solely in terms of the most severe, but less common, outcomes, such as mortality, will underestimate the total health burden of air pollution (WHO 2001b).

A large body of epidemiological research, discussed in more detail below, provides evidence that exposure to air pollution is associated with increased mortality and morbidity. The respiratory and cardiovascular systems appear to be the most affected. A growing body of toxicological and clinical evidence currently offers some limited insight into the mechanisms through which exposure to air pollution may produce the effects on respiratory and cardiovascular outcomes observed in epidemiological studies (Anonymous 1996a, 1996b; Health Effects Institute 2002). These mechanisms may involve decrements in pulmonary func-

Figure 17.4 The relative frequencies of health events associated with exposure to air pollution



tion, effects on heart rate variability and inflammatory response. Long-term bioassays and other studies of toxicity provide evidence for the mutagenicity and/or carcinogenicity of some components of urban air pollution, such as emissions from diesel-powered vehicles (Cohen and Nikula 1999; Diesel Working Group 1995).

Air pollution may elicit both acute and chronic biological responses. Acute responses to air pollution in otherwise healthy persons may be confined to reversible physiological adaptations resulting from natural defence mechanisms (e.g. watery eyes, cough or a transient fall in lung function). Acute responses may, however, also increase the severity or duration of an already established respiratory infection or of diseases such as asthma or chronic obstructive lung disease that have already placed the individual in a vulnerable position, and increase the risk of hospital admission or even death. If such vulnerability were temporary, for example, a severe infection of the lower respiratory tract, the individual might have recovered and lived for some time, had it not been for the added factor of exposure to air pollution at the time the individual was most vulnerable because of the infection. On the other hand, if the individual had a terminal chronic condition, such as severe chronic obstructive pulmonary disease or chronic congestive heart failure, exposure to air pollution might advance death by only a short time, this being imminent in any case. There is limited epidemiological evidence to suggest that ambient air pollution may contribute to the development

of diseases such as chronic obstructive pulmonary disease, for which smoking and, in developing countries, indoor air pollution, are also risk factors (Abbey et al. 1999; Pope and Dockery 1999; Tager et al. 1998). Distributions of short-and long-term vulnerability, reflecting the prevalence of acute and chronic cardiorespiratory disease, may well differ across populations worldwide. This will have implications for the transferability of risk functions from studies in populations in NAWE to populations where differences in genetic factors, diet, tobacco smoking, extent of urbanization, distribution of wealth and other factors related to social class, have resulted in different patterns of disease.

For example, recent analyses of two cohorts in the United States (Krewski et al. 2000) showed clearly that the effects of long-term exposure to air pollution on mortality depend on attained educational level, with the largest relative effects observed among the least educated. Recent studies in developing countries have also reported such gradients in the relative risks of mortality (O'Neill et al. 2003). It is not clear what factor(s) might be responsible for these observations (e.g. aspects of occupation or diet), but it is reasonable to expect that they might vary across the globe. Differences in vulnerability to air pollution introduce a source of uncertainty in our estimates that can currently be only partially quantified.

Epidemiological evidence about exposure–response relationships is most directly applicable to the risk assessment of air pollution, because it comes from the direct observation of human populations under relevant conditions (Samet 1999). Epidemiological studies have limitations that are largely a result of their observational nature. These relate to the accurate measurement of exposure, definition of outcomes and interpretation of associations that are observed. Assessing the causality of such associations requires a process of scientific reasoning that considers all evidence, including that from experimental studies (WHO 2000b). While there remain many gaps in our knowledge about the explanations for epidemiological associations, they can provide the best evidence to guide action to reduce the exposure of the population to air pollution, and to undertake health impact assessments, provided the uncertainties are recognized.

The epidemiology of air pollution takes advantage of the fact that concentrations of urban air pollution, and thus human exposure, vary in both time and space. For the most part, current epidemiological research has focused on either one or the other dimension, but infrequently on both within the same population(s). Short-term temporal variation in concentrations of air pollution over days and weeks has been used to estimate effects on daily mortality and morbidity. Spatial variation in long-term average concentrations of air pollution has provided the basis for cross-sectional and cohort studies of long-term exposure.

3.1 Studies of short-term exposure

The effects of short-term exposure to air pollution have been extensively studied in time-series studies in which daily rates of health events (e.g. deaths or hospital admissions) in one or more locales are analysed in relation to contemporaneous series of daily concentrations of air pollutants, and other risk factors (e.g. weather) that vary over time periods of months or years. Regression techniques are used to estimate a coefficient that represents the relationship between exposure to pollution and the outcome variable. The usual method of regression models the logarithm of the outcome, and thus arrives at an estimate of the relative risk, a proportional change in the outcome per increment of ambient concentration. There has been a rapid increase in the number of these studies as computing and statistical techniques have improved and as data on outcomes and air pollution have become more extensive and easily accessible from routine sources. It is a strength of these studies that individual cofactors, such as smoking, nutrition, behaviour, genetic factors, etc., are unlikely confounders because they are not generally associated, on a day-to-day basis, with the daily concentration of air pollution. Studies of time series have found associations between concentrations of PM in the air and a large range of outcomes. These have been reviewed by Pope and Dockery (1999) and include daily mortality (all causes, respiratory, cardiovascular), hospitalization for respiratory diseases (all causes, chronic obstructive pulmonary disease, asthma, pneumonia) and for cardiovascular diseases (acute myocardial infarction, congestive cardiac failure). Since this review, associations have also been reported for primary health care visits for disease of the lower respiratory tract, and diseases of the upper respiratory tract of both infective and allergic origin (Hajat et al. 2001, 2002). However, recent methodological studies and re-examination of earlier work indicate that the magnitude of the estimates of relative risk from time-series studies of daily mortality depends on the approach used to model both the temporal pattern of exposure (Braga et al. 2001) and potential confounders that vary with time, such as season and weather (Health Effects Institute 2003).

The acute effects of air pollution have also been studied longitudinally in panel studies, which can provide evidence of physiological effects at an individual level. Small groups, or panels, of individuals are followed over short time intervals, and health outcomes, exposure to air pollution and potential confounders are ascertained for each subject on one or more occasions. Panel studies have generally reported associations of exposure to urban air pollution with increased prevalence of symptoms involving the upper and lower respiratory tract, and increased rates of asthma attacks and medication use. Associations with short-term reduction in lung function and the prevalence of cough symptoms have been reported in studies in the United States (Pope and Dockery 1999), but are not consistently supported by studies in Europe (Roemer et al. 1999).

Time-series studies in adults across the world

Studies of time series concerning daily mortality and, to a lesser extent, daily hospital admissions, have been conducted in cities throughout the world. A recent meta-analysis summarized the evidence from >100 studies of daily mortality (Stieb et al. 2003). In addition, large studies have now been conducted using uniform methods for assembling and analysing data from multiple cities: APHEA 2 (Air Pollution and Health: A European Approach) (Katsouyanni et al. 1996, 2001) and NMMAPS (National Mortality and Morbidity Air Pollution Study) (Health Effects Institute 2000a, 2000b) in the United States. These multi-city studies have confirmed the findings of earlier studies of individual cities in finding positive associations between daily mortality and hospital admissions and concentrations of PM, and have also attempted to explain the heterogeneity among cities in the relative risks associated with exposure to air pollution. For example, in the APHEA 2 Study, it was found that the effects of PM on mortality were modified by mean concentrations of nitrogen dioxide (Katsouyanni et al. 2001), and in the NMMAPS Study, daily mortality was modified by the long-term average concentrations of PM_{10}. Levy et al. (2000) reported that the effects of PM_{10} were greater in cities where $PM_{2.5}$ comprised a higher proportion of PM_{10}.

Most studies of time series are from countries in NAWE, where air pollution is low and decreasing and populations are characterized by western lifestyles and patterns of disease. To examine the epidemiological evidence for other non-NAWE countries, we searched a database of studies of time series and panel studies compiled at St George's Hospital Medical School, for which researchers had systematically ascertained, reviewed, and abstracted results from studies published in the peer-reviewed scientific literature (WHO 2003). All studies meeting pre-specified quality criteria related to adequacy of confounder control, and which provided estimates of the concentration–response relationship and its statistical precision were included. We classified them by the subregion in which the study was performed and tabulated the results for six outcomes: all-cause mortality, respiratory mortality, cardiovascular mortality, infant and child mortality, reduction in peak expiratory flow rate and cough symptom.

The distribution of time-series and panel studies by outcome and subregion is shown in Tables 17.5(a) and (b). Up to mid-November 2001, the number of studies from AMR-A (71) and EUR-A (75) far exceeded the total for the remaining 12 subregions (42). The next largest contributor was AMR-B (Central and South America), with 18 studies. The table shows the numbers of panel studies presenting usable numerical estimates. Only 14 studies from non-NAWE subregions were identified, compared with 64 from the NAWE subregions. Some of the non-NAWE countries had lifestyles and patterns of disease similar to those in NAWE—those in Australasia, for example.

Table 17.5 Distribution of studies by outcome status and subregion

(a) Selected time-series studies

	Outcome status					
	Cause of mortality			*Hospital admissions/ emergency room visits*	*Other time-series studies*	*Total (from subregion)*
Subregion	*All causes*	*Respiratory disease*	*Cardiovascular disease*			
AFR-D	0	0	0	0	0	0
AFR-E	0	0	0	0	0	0
AMR-A	34	11	13	39	2	71
AMR-B	10	9	6	5	0	18
AMR-D	0	0	0	0	0	0
EMR-B	0	0	0	0	0	0
EMR-D	0	0	0	0	0	0
EUR-A	44	28	24	28	3	75
EUR-B	4	3	4	0	0	5
EUR-C	0	0	0	0	0	0
SEAR-B	1	1	1	0	0	1
SEAR-D	0	0	0	0	0	0
WPR-A	4	4	4	2	0	6
WPR-B	8	3	4	4	0	12
World (from outcome status)		113		78	5	187

(b) Selected panel studies

	Outcome status for valid studies		*Total for selected studies (from subregion)*
Subregion	*Lung function*	*Symptoms*	
AFR-D	0	0	0
AFR-E	0	0	0
AMR-A	19	17	28
AMR-B	3	3	3
AMR-D	0	0	0
EMR-B	0	0	0
EMR-D	0	0	0
EUR-A	32	31	38
EUR-B	3	2	3
EUR-C	2	1	2
SEAR-B	0	0	0
SEAR-D	0	1	1
WPR-A	4	1	4
WPR-B	0	1	1
World (from outcome status)	62	57	79

Figures 17.5 and 17.6 show the results for daily mortality and PM, by subregion, for adults and children, respectively. These estimates were scaled to PM_{10} ($PM_{2.5} = 0.6 \times PM_{10}$, $BS = 0.5 \times PM_{10}$, $TSP = 2 \times PM_{10}$). These scaling factors were decided after examining a number of co-located measures, but are likely to be variable across the individual cities. Estimates of random effects and fixed effects are shown because there is heterogeneity.

The pooled estimates are shown in Table 17.6. Most of the studies of mortality showed relative risks of >1.0 (i.e. a change of >0.0%), with lower 95% confidence intervals also >1.0. However, the studies showing the largest confidence intervals also tended to have the largest effects, this indicating the possibility of publication bias. There was considerable heterogeneity in the estimates. For this reason, the summary estimate for random effects is more appropriate because it takes into account the greater uncertainty. The estimate for random effects was increased and statistically significant for mortality from all causes, respiratory and cardiovascular diseases. It is remarkable that for daily mortality, the pooled estimate for the non-NAWE subregions of 0.5% increase in daily mortality per 10 $\mu g/m^3$ increase in PM was very similar to the estimates produced by the APHEA 2 and NMMAPS studies of 0.6 and 0.5, respectively. A recent meta-analysis of 109 published studies from around the world reports similar estimates (Stieb et al. 2003).

These results indicate that daily mortality is positively associated with short-term exposure to urban air pollution at time-scales in the order of days, in all subregions where this association has been measured. They also suggest that the relative effect of exposure may also be of similar magnitude in different parts of the world.

Air pollution and reproductive and child health

Six time-series studies of daily mortality report associations between particulate pollution and adverse effects in children, and all of them are from non-NAWE countries. Their estimates are shown in Figure 17.6. Four were from São Paulo (Conceiao et al. 2001; Gouveia and Fletcher 2000; Pereira et al. 1998; Saldiva et al. 1994), one from Mexico City (Loomis et al. 1999) and one from Bangkok (Ostro et al. 1999a). The Bangkok study was of PM_{10} and daily mortality from all causes in children aged <6 years. The study conducted in Mexico City evaluated the impact of daily changes in concentrations of $PM_{2.5}$ and total mortality in children aged <1 year. Three studies in São Paulo (Conceiao et al. 2001; Gouveia and Fletcher 2000; Saldiva et al. 1994), conducted during different periods of time, all reported an association between PM and mortality from respiratory disease in children aged <5 years. The study conducted in São Paulo by Pereira et al. (1998) investigated the association of exposure to urban air pollution with intrauterine mortality. Some of the relative risks reported in these studies were >1.0, but only the estimate from Mexico City was statistically significant at the 95% level.

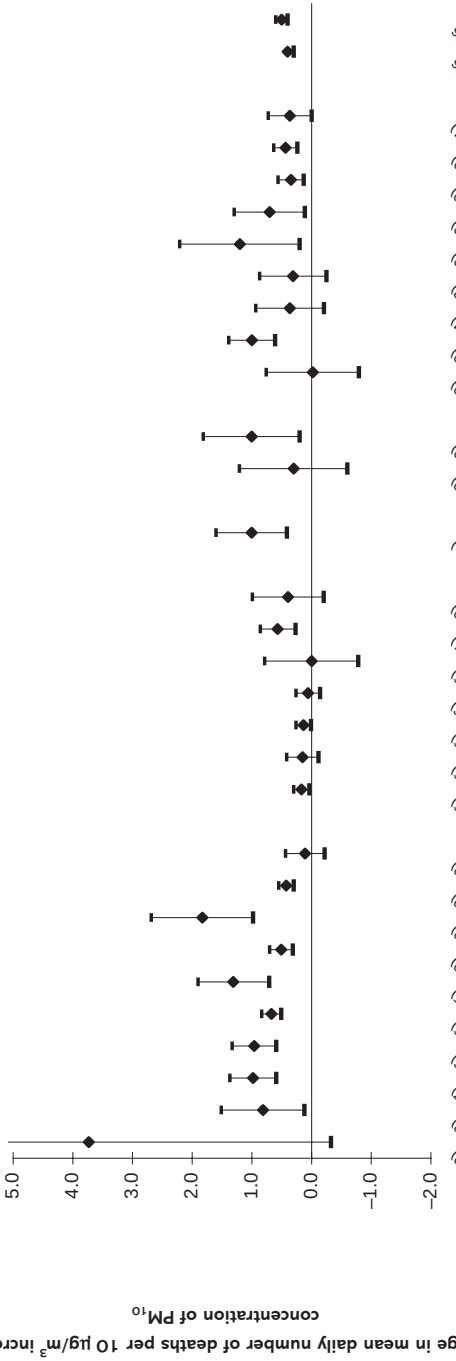
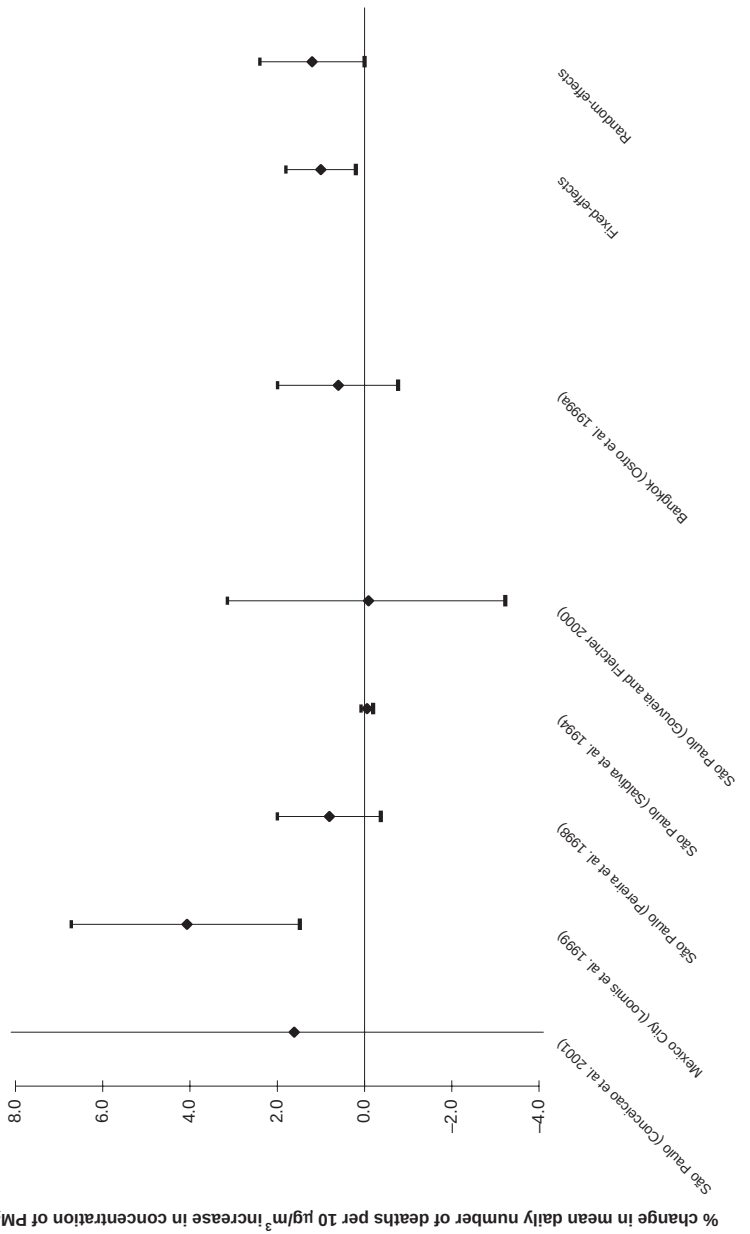


Figure 17.5 Percentage change in mean daily number of non-accidental deaths, in adults, per 10µg/m^3 increase in PM_{10}, by city

Note: Scaling factors used: $PM_{2.5} = 0.6 \times PM_{10}$, $BS = 0.5 \times PM_{10}$, $TSP = 2 \times PM_{10}$.

Figure 17.6 Percentage change in mean daily number of non-accidental deaths, in children, per 10μg/m^3 increase in PM_{10}, by study



Note: Scaling factors used: $PM_{2.5} = 0.6 \times PM_{10}$, $BS = 0.5 \times PM_{10}$, $TSP = 2 \times PM_{10}$.

Table 17.6 Pooled estimates of daily mortality from all causes and concentrations of PM_{10}, from time-series studies, excluding North America and western Europe

	All-cause mortality (% change)	*Mortality from respiratory disease (% change)*	*Mortality from cardiovascular disease (% change)*	*Infant and child mortality (% change)*	*Lung function (regression coefficients)*	*Symptoms (odds ratios)*
No. of studies	29	18	18	5[a]	11	10[b]
Heterogeneity	Q=121.62 on 28 *df* (P<0.001)	Q=68.86 on 17 *df* (P<0.001)	Q=51.40 on 17 *df* (P<0.001)	Q=7.72 on 4 *df* (P=0.102)	Q=16.16 on 10 *df* (P=0.095)	Q=26.68 on 9 *df* (P=0.002)
RE (95% CI)	0.5 (0.4–0.6)	0.6 (0.2–1.1)	0.3 (0.1–0.5)	1.0 (–0.9–3.1)	–0.024 (–0.146–0.098)	1.006 (0.990–1.022)
FE (95% CI)	0.4 (0.3–0.4)	0.4 (0.2–0.6)	0.1 (0.1–0.2)	1.0 (–0.1–2.1)	–0.007 (–0.095–0.081)	1.006 (0.998–1.013)
NMMAPS study 20 USA cities (Health Effects Institute 2000b)	0.46 (0.27–0.65)	—	—	—	—	—
APHEA 2 study 29 European cities (Katsouyanni et al. 2001)	0.62 (0.4–0.8)	—	—	—	—	—

Key: RE, random-effects estimate; FE, fixed-effects estimate; Q, χ^2 statistical test for heterogeneity; *df*, degrees of freedom.

— No data.

[a] 1 study (Pereira et al. 1998) not included in meta-analysis due to uncommon outcome.

[b] 1 study (Awasthi et al. 1996) not included in meta-analysis due to uncommon particle measurement.

Infant mortality from respiratory disease and other adverse perinatal events, such as low birth weight and malformations, have also been associated with more prolonged exposure to air pollution (Bobak and Leon 1999; Wilhelm and Ritz 2003; Woodruff et al. 1997). Woodruff et al. followed a large birth cohort in the United States for one year and estimated the relative risk of mortality associated with residential exposure to PM_{10} in the first two post-natal months, conditional on a variety of potential confounders, including maternal smoking. They reported an increase in total mortality of 4% per 10 $\mu g/m^3$, and 20% for mortality from respiratory causes. Two studies recently evaluated changes in infant mortality associated with reductions in industrial emissions caused by a recession and mandated reductions in pollution resulting from the United States Clean Air Act Amendments of 1970 (Chay and Greenstone 1999, 2001). Using county-level data, the authors estimated that 4–8 infant deaths per 100 000 live births were prevented for each 1 $\mu g/m^3$ reduction in TSP.

Studies of acute morbidity

Far fewer studies have been conducted of the association of exposure to urban air pollution with acute morbidity, especially in non-NAWE countries. There were insufficient studies in any one outcome group to allow formal meta-analysis of non-NAWE studies, but most reports suggested a positive association with urban air pollution, consistent with that observed in NAWE countries, especially for hospital admissions (Atkinson et al. 2001; Burnett et al. 1999; Health Effects Institute 2000b). A recent study compared directly the effects of air pollution on hospital admissions in China, Hong Kong Special Administrative Region (Hong Kong SAR) and London. Similar associations were observed for PM_{10} and gaseous pollutants and hospital admissions for ischaemic heart disease in both locations, and the associations were strongest during seasons of low humidity in both cities, but no association with admissions for cardiac disease was observed in Hong Kong SAR (Wong et al. 2002).

3.2 Studies of long-term exposure

Cohort studies of mortality from chronic respiratory and cardiovascular disease

Cohort studies take advantage of spatial heterogeneity in concentrations of air pollution to compare the incidence of disease and death in populations exposed in the long term to differing levels of pollution. By following large populations for a number of years, cohort studies provide estimates of both attributable numbers of deaths and, more importantly, average reductions in life span attributable to air pollution.

The evidence from cohort studies of populations in Europe and the United States indicates that long-term exposure to urban air pollution

is associated with an increase in total and cardiopulmonary mortality in adults (Dockery et al. 1993; Hoek et al. 2002; Lipfert et al. 2003; McDonnell et al. 2000; Pope et al. 2002). In each of these studies, the effects of potential confounders such as cigarette smoking, occupation and prior medical history were adjusted for in regression analyses. Most studies find the strongest and most consistent associations with exposure to PM, and $PM_{2.5}$ appears to be more closely associated with mortality than PM_{10} or TSP (Dockery et al. 1993; Pope et al. 2002). The recently published results of a study conducted in the Netherlands confirm the impacts of long-term exposure to air pollution, and in particular that related to road traffic, on mortality (Hoek et al. 2002).

Unfortunately, the cohort studies provide little information on when exposure to air pollution acts to increase the risk of mortality (i.e. the induction time for mortality attributable to exposure to long-term exposure to air pollution), making it difficult to estimate when the effects of reduction of air pollution might be observed.

Comparable cohort studies have not yet been carried out in developing countries. However, the imposition of restrictions on the sulfur content of fuel for power generation and transportation in Hong Kong SAR, instituted over short time intervals in 1990, provided opportunities for researchers to measure directly the impact of reducing air pollution on long-term average mortality (Hedley et al. 2002). Hedley et al. (2002) documented both changes in ambient air quality subsequent to the imposition of the restrictions, and declines in long-term average rates of mortality from cardiovascular and respiratory diseases associated with those changes. Comparison of changes in mortality in more and less polluted areas of Hong Kong SAR provided limited ability to account for secular changes in other risk factors for mortality that could have produced the observed decrease in mortality following the change in the sulfur content of fuel. A similar study was also published recently by Clancy et al. (2002), who measured decreased long-term average mortality in Dublin after the banning of the sale of bituminous coal in Dublin in 1990.

The American Cancer Society study

The American Cancer Society (ACS) study (Pope et al. 2002) in the United States is by far the largest cohort study of air pollution and long-term average mortality to date. The ACS study of air pollution and mortality is based in the ACS Cancer Prevention II Study, an on-going prospective cohort of approximately 1.2 million adults from all 50 states (Calle et al. 2002). Friends and neighbours recruited cohort members on behalf of the ACS. Participants were enrolled in 1982, when they were aged ≥30 years, and their mortality has been ascertained through to 1998. Data on a wide range of risk factors for cancer and other chronic diseases were obtained from each participant. The ACS study links the data for approximately 500 000 cohort members with data on air pol-

lution from metropolitan areas throughout the United States. The first study of air pollution and mortality in this cohort (Pope et al. 1995) was based on follow-up through to 1990. That study reported increases in mortality from cardiopulmonary disease for 19.9 $\mu g/m^3$ fine particulate sulfate (relative risk of 1.26, 95% CI 1.16–1.37), and from lung cancer (relative risk of 1.36, 95% CI 1.11–1.66). These findings were subsequently corroborated in an independent re-analysis (Krewski et al. 2000). A more recent analysis of this cohort extended follow-up through to 1998, and ascertained 40 706 deaths from cardiopulmonary disease, and 10 749 from lung cancer. Data were analysed using Cox proportional hazards regression models that incorporated both random effects and non-parametric spatial smoothing to adjust for unmeasured factors correlated spatially with air pollution and mortality across the United States. The models also adjusted for age, sex, race, education, marital status, body mass, diet, alcohol consumption, occupational exposures and the duration and intensity of cigarette smoking, all measured via questionnaire at enrolment.

Concentrations of ambient air pollution had, in general, declined across the United States between 1982 and 1998. Measurements of ambient concentrations of fine particulate air pollution ($PM_{2.5}$) in the cities where subjects resided at enrolment were available for periods both briefly preceding enrolment (1979–83) and immediately after follow-up (1999–2000). In separate regression analyses, cohort members were assigned estimates of exposure corresponding to their city-of-residence-specific value for each of those periods, as well as for the average value across the two periods. For a change of 10 $\mu g/m^3$ in the ambient concentration of $PM_{2.5}$, the smallest relative increases were observed for the mean concentration of the time period 1979–1983. This estimate was based on data from 61 cities, with a mean concentration of $PM_{2.5}$ of 21.1 $\mu g/m^3$, and a range of 10–30 $\mu g/m^3$. The relative risks for a 10 $\mu g/m^3$ change in the concentration of ambient $PM_{2.5}$ were larger when exposure was specified as the average of the ambient concentrations of the two time periods. This may be explained by the fact that the estimates from the earliest periods are more subject to random (and non-differential) error. However, it also suggests that more recent exposures may be exerting the strongest effects on mortality, an interpretation also offered in the recent re-analysis of the earlier follow-up of the ACS cohort (Krewski et al. 2000). Unfortunately, it was not possible to derive individual time-varying estimates of exposure from the available data (e.g. detailed residence histories were unavailable), precluding direct evaluation of the induction time for mortality attributable to exposure to air pollution.

Long-term exposure and the incidence of chronic disease

Little evidence is available concerning exposure to air pollution and the incidence of chronic cardiovascular or respiratory disease. One study in

the United States reported an association of long-term exposure to PM_{10} with the incidence of self-reported physician-diagnosed chronic bronchitis (Abbey et al. 1999). A recent case–control study reported an association between short-term exposure and the incidence of non-fatal myocardial infarction (Peters et al. 2001). Cross-sectional studies have found associations with reduced lung function and increased respiratory symptoms in both adults and children, which might in part represent chronic disease as the result of long-term exposure. Several recent cross-sectional studies in large Chinese cities have reported increased prevalence of respiratory symptoms in adults (Qian et al. 2001; Zhang et al. 1999) and elementary school children (Qian et al. 2000; Zhang et al. 2002) exposed to urban air pollution. A cross-sectional study in Delhi observed reductions in pulmonary function in residents of highly polluted areas, but little evidence of increased prevalence of symptoms (Chhabra et al. 2001).

Air pollution and lung cancer

Epidemiological studies over the last 40 years have observed that general ambient air pollution, chiefly composed of the by-products of the incomplete combustion of fossil fuels, is associated with small relative increases in the incidence of lung cancer. The evidence derives from studies of trends in the incidence of lung cancer, studies of occupational groups, comparisons of urban and rural populations, and case–control and cohort studies using diverse exposure metrics. Recent prospective cohort and case–control studies which have controlled for the effects of cigarette smoking, occupation and other risk factors have consistently observed small increases in the relative risk of lung cancer in relation to exposure to particulate air pollution (Abbey et al. 1999; Dockery et al. 1993; Krewski et al. 2000; Pope et al. 2002; Samet and Cohen 1999). A recent Swedish case–control study reported that excess lung cancer was related specifically to exposure to mobile sources of air pollution, with the largest effects observed for exposure occurring 20 years prior to diagnosis (Nyberg et al. 2000).

3.3 Choice of outcomes and hazards

Studies used for hazard estimates

The use of results from time series to estimate the disease burden attributable to urban air pollution is problematic for various reasons. First, data on rates of occurrence, such as hospital admissions, primary health care consultations or asthma exacerbation are not collected in many countries. Thus there is no baseline upon which to develop an estimate of health impact. Mortality is an exception in that data are available from death registration or indirect demographic methods in all subregions.

The application of the time-series concentration–response functions to the assessment of mortality impact, however, is limited. Specifically, it is not possible to use the results of studies of time series to estimate YLL in adults. This is because time-series studies of daily mortality do not in themselves allow the estimation of lost life time, but rather only allow estimation of the number of deaths that have been brought forward by an unspecified amount of time. Recent research has made clear that the time-series estimates reflect deaths that may have been brought forward by as much as several months, rather than simply advancing the time of death in frail people by a few days (Schwartz 2000; Zeger et al. 1999). The design of the time-series study of daily mortality, which requires the control of long-term variation in air pollution, precludes estimation of greater losses (Künzli et al. 2001; Leksell and Rabl 2001; McMichael et al. 1998). Thus, the time-series studies only provide an estimate of the daily number of deaths brought-forward.

Cohort studies include not only people whose deaths were advanced by recent exposure to air pollution, but also those who died from chronic disease caused by long-term exposure (COMEAP 1998; Künzli et al. 2001), thus they provide a more comprehensive estimate of the effects on mortality. Furthermore, because their relative risks can be applied to population life tables, the effects of air pollution on life span can be estimated (Brunekreef 1997; COMEAP 2001; Hurley et al. 2000; Sommer et al. 2000)

The situation may be different for children. In developing countries, the major causes of death, such as acute respiratory disease, are very likely to result from a severe acute infection, which represents a brief window of vulnerability. If the child survives, they might be expected to fully recover and enjoy a full life expectancy. If we assume that death was not otherwise imminent, then these deaths, on average, represent the loss of considerable life years. Under such an assumption, one could use time-series estimates to estimate YLL in children aged 0–4 years.

In making these estimates, several further considerations should be kept in mind. The first is that the effects of cumulative exposure over several weeks are several times greater than those obtained by using a single day lag and thus underestimate the impacts on health. The second is that other air pollutants in the mixture may be exerting additional effects, may interact with particles or may be confounding the associations of particles. Ozone, for example, is also toxic and while its effects tend to be independent of PM, it also seems to modify the effect of particles on number of hospital admissions (Atkinson et al. 2001).

Definition and specification of health outcomes

We estimated the burden of disease imposed by mortality from cardiopulmonary disease and lung cancer in adults, and from ARI in children aged 0–4 years. We made this choice despite the fact, discussed above, that other serious health effects of air pollution are well-

documented, and that still others appear from more limited evidence to be of potential concern. Our decision to focus on mortality outcomes was made on the basis of the following considerations.

- *Strength of evidence*. A large body of research from many parts of the world indicates that ambient air pollution causes increased daily mortality from cardiovascular and respiratory disease. There appear to be comparable effects in the cities of developed and developing countries, on the basis of the limited evidence available. Although no cohort studies of mortality have as yet been conducted in developing countries, the possibility that associations comparable to those observed in the United States would be observed is strengthened by the results of the studies of daily mortality, and the limited results from studies of morbidity.
- *Consistent definition of the end-point*. Mortality *per se* is a well-defined event that is registered in most countries. For this reason, epidemiologists have frequently measured the effect of air pollution on total mortality from all natural causes, ascertained from death certificates or other sources of vital statistics. Other outcomes, such as bronchitis and the symptom of wheeze, are subject to very large variations in severity, and without such qualification their impact on health is difficult to assess. The definitions of other possible health outcomes, such as restricted activity days, use of primary health care services, diagnoses and school absences, are likely to vary with national culture and among health care systems.

  The cause of death is more problematic because it is not certified medically in many countries, and there are considerable differences between and within countries in terms of diagnostic practice. Nevertheless, we propose to base our estimates primarily on cardiopulmonary, rather than total, mortality. There is strong evidence from both time-series and cohort studies that ambient air pollution specifically increases the risk of death from these causes. Moreover, variations in the relative contribution of non-cardiopulmonary mortality among countries could increase the error in the burden assessment, particularly in countries with lower cardiopulmonary death rates, potentially leading to overestimates of impact. Since considerable cross-coding is likely, we have chosen to use the combined cardiopulmonary group consisting of GBD infectious and chronic respiratory diseases and selected cardiovascular outcomes for adults. In children, death from cardiovascular diseases is rare and the pulmonary group is adequate.
- *Availability of baseline occurrence rates*. Data on age-specific mortality are collected or estimated using consistent methods for all subregions. This is not the case for some important potential measures of morbidity, such as the frequency of asthma attacks, or measures of

the utilization of health care services outside of Europe and North America.

- *Importance of the end-point in terms of health impact*. Although the impacts of air pollution on other health end-points must certainly contribute to the global burden of disease, mortality, quantified in terms of either numbers of deaths or reduced survival time, currently plays the most prominent role in impact assessments. We chose these three specific mortality outcomes (mortality from cardiopulmonary causes and lung cancer in adults aged ≥30 years, and mortality from ARI in children aged 0–4 years) because they allow us to estimate YLL, as discussed above.
- *Feasibility within the time constraints of the current work*. Given additional time and resources, future efforts might possibly consider, for example, using the evidence from the International Study of Asthma and Allergies in Childhood (ISSAC) (Anonymous 1998), which has data on prevalence from 60 countries as a baseline upon which to estimate the effect of particles on the exacerbation of asthma. Another possibility might be the effect on hospital admissions or primary health care visits.

3.4 Developing the concentration–response functions

We derived concentration–response functions for three end-points to produce the estimates of global burden of disease reported in this work: mortality from cardiopulmonary causes and lung cancer in adults aged ≥30 years, and mortality from ARI in children aged 0–4 years. As discussed earlier, we made no estimates of the impacts of PM on the incidence of disease, so the disability-adjusted life years (DALYs) quantify only YLL.

We assumed a log-linear risk model which leads to the following formula for the relative risk (RR) in a population whose exposure is estimated by an average concentration of pollution C relative to the reference level C_0:

$$RR = \exp[\beta(C - C_0)]; \tag{6}$$

where, C_0, the reference, or theoretical minimum level of exposure, is defined as above and β is the estimated effect of PM on the health outcome of interest. We calculated a subregion-specific relative risk for each of the 14 subregions using a population-weighted mean of concentrations for all cities in the subregion calculated as follows:

The subregion-specific relative risk for outcome i in subregion k related to $PM_{2.5}$, $RR_{2.5ik}$, is:

$$RR_{2.5ik} = \exp[\beta_{2.5i} \times (Ck_{2.5} - 7.5)] \tag{7}$$

where $C_{k2.5}$ is the subregion-specific population-weighted mean concentration of $PM_{2.5}$ (calculated from the estimated concentration of PM_{10} in Table 17.4, as described above), and $\beta_{2.5i}$ is the slope of the concentration–response function for $PM_{2.5}$ (Table 17.7).

The subregion-specific relative risk for outcomes quantified in terms of PM_{10}, RR_{10ik}, is:

Table 17.7 Estimates of relative risk of mortality, coefficients of concentration–response functions and study types

Health outcome	*Data source*	*PM exposure metric*	*Relative risk per 10 μg/m³ (95% CI), from data source*	*Concentration–response slope[a] per μg/m³ (standard error)*
Mortality from cardiopulmonary disease—adults	ACS study (Pope et al. 2002)	$PM_{2.5}$	1.059 (1.015–1.105)	Linear[b] 79–83 0.00575 (0.002160) Linear average[c] 0.008933 (0.002907) Log-linear[d] 79–83 0.11605 (0.044790) Log-linear average 0.155148 (0.050460)
Mortality from lung cancer	ACS study (Pope et al. 2002)	$PM_{2.5}$	1.082 (1.011–1.158)	Linear 79–83 0.00789 (0.003447) Linear average 0.012673 (0.00426) Log-linear 79–83 0.17114 (0.071968) Log-linear average 0.232179 (0.074770)
Mortality from acute respiratory infection—children aged 0–4 years	St George's Hospital meta-analysis of five time-series studies of daily mortality	PM_{10}	1.010 (0.991–1.031)	0.0010 (0.0010)
Deaths-brought-forward—all ages	St George's Hospital meta-analysis of 165 time-series studies of daily mortality	PM_{10}	1.006 (1.005–1.007)	0.0006 (0.00005)

[a] Slope of the concentration–response (CR) function for air pollution and mortality.

[b] Results form regression models in which annual average concentrations measured from 1979–1983 were used as estimates of exposure (Pope et al. 2002).

[c] Results form regression models in which the average of annual average concentrations measured from 1979–1983 and 1999–2000 were used as estimates of exposure (Pope et al. 2002).

[d] Results from regression models where exposure (i.e. annual average $PM_{2.5}$) is specified on the log scale.

$$RR_{10ik} = \exp[\beta_{10i} \times (C_{k10} - 15)] \quad (8)$$

where C_{k10} is the subregion-specific population-weighted mean concentration of PM_{10} (Table 17.4), and β_{10i} is the slope of the concentration–response function for PM_{10}. The city-specific concentrations of PM_{10} were truncated at 15 and 100 μg/m^3 for calculation of subregion-specific population-weighted mean concentrations of PM_{10}.

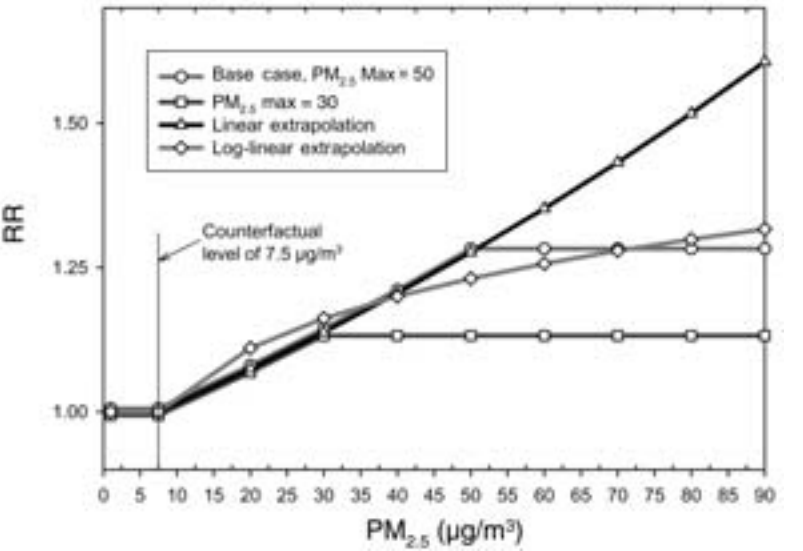
Mortality from cardiopulmonary disease and lung cancer in adults

We used the results of the ACS study of urban air pollution and mortality (Pope et al. 2002) to estimate attributable deaths and YLL from cardiopulmonary diseases and lung cancer in adults aged ≥30 years. In our base-case analyses we used the estimates of the concentration–response functions based on the ambient concentrations in 1979–1983, which correspond to increases of 5.9% and 8.2% in mortality from cardiopulmonary disease and lung cancer, respectively, for each 10 μg/m^3 change in the ambient concentration of $PM_{2.5}$ (Table 17.7).

Deaths from cardiopulmonary disease and lung cancer in the ACS cohort were defined as persons whose underlying cause of death was coded according to the International Statistical Classification of Diseases, ninth revision, on their death certificates as ICD-9 401–440 and 460–519, and 162, respectively. When calculating the attributable fraction, ACS concentration–response functions for cardiopulmonary disease defined in this way were applied to baseline cause-specific rates of mortality in the GBD project. For lung cancer, this corresponded exactly to the definition used in the ACS study. For cardiopulmonary deaths, the GBD groupings of cause of death (39, 40, 106–109, 111) did not include several ICD codes (406–409, 415–417, 423–424, 426–429, 440) that were included in the ACS definition. These codes represent diverse cardiac diseases, including conduction disorders, cardiac dysrhythmias, heart failure and ill-defined cardiac causes. Together they comprise approximately 18% of all cardiopulmonary deaths in the ACS study (R. Burnett, personal communication).

The ACS study estimated concentration–response functions for $PM_{2.5}$ over a range that extends from annual average concentrations of $PM_{2.5}$ of about 5–30 μg/m^3 (Pope et al. 2002). The shape of the concentration–response function for fine particulate air pollution outside that range is currently unknown, as noted above, and estimated annual average concentrations of $PM_{2.5}$ in some subregions are outside that range (Table 17.1). In our base-case analyses we limited the risk of mortality in any city to be no greater than that attained at a concentration of $PM_{2.5}$ of 50 μg/m^3 (Figure 17.7). Thus, for cities with estimated annual average concentrations of >50 μg/m^3, we assigned a maximum concentration, or C_m, equal to 50 μg/m^3, regardless of their actual estimated

Figure 17.7 Alternative concentration–response curves for mortality from cardiopulmonary disease, using different scenarios



concentration. This means that the excess risk is constrained to be no greater than that associated with an annual average concentration of 50 µg/m^3, regardless of the actual estimated annual average concentration. The counterfactual or theoretical minimum concentration was set at 7.5 µg/m^3, as discussed above.

We set the maximum city-specific concentration of $PM_{2.5}$, C_m, at 50 µg/m^3 to avoid producing unrealistically large estimates of mortality in the most extremely polluted subregions under a linear exposure model. With $C_m = 50$, the subregion-specific attributable fraction was restricted to no more than approximately 25% of the burden of a given health outcome, while not greatly exceeding the maximum observed annual average concentration in the ACS study. We also examined alternative values for the shape of the concentration–response function for $PM_{2.5}$ and mortality from cardiopulmonary disease and lung cancer in sensitivity analyses, as described below.

Mortality from acute respiratory infections in children aged 0–4 years

In view of the importance of mortality from ARI among children in developing countries and the suggestion from available evidence of an association with air pollution (Romieu et al. 2002), we decided to make

a summary estimate on the basis of these studies, in spite of their heterogeneity in outcomes and age groups.

To estimate the relationship between exposure to PM and mortality from ARI among children aged 0–4 years, we computed a summary estimate from the five published time-series studies discussed above (Table 17.6 and Figure 17.6). One study (Pereira et al. 1998) was excluded because the outcome, intrauterine mortality, was clearly unrelated to ARI. The five remaining studies were summarized as a weighted average of the estimates from individual studies (scaled to PM_{10}, as discussed above) with the weights determined by the inverse of the reported variance in the concentration–response function. We estimate that a 10 $\mu g/m^3$ increase in ambient concentrations of PM_{10} results in a 1.0% (95% CI—0.9%–3.1%) increase in daily mortality from ARI in children aged 0–4 years (Table 17.7).

When calculating the attributable fraction, this concentration–response function was applied to GBD baseline cause-specific rates of mortality from acute respiratory infection (GBD code 38) that includes ICD-9 codes 460–466, 480–487 and 381–382.

Numbers of deaths from all natural causes caused by short-term exposure to urban air pollution in adults

We also calculated an estimate of the numbers of deaths from all natural (non-injury) causes attributable to short-term exposure to urban air pollution using an estimate of concentration–response derived from international literature on air pollution and daily mortality. This estimate was not included in the total attributable deaths and disease burden because of the conceptual issue in quantifying the effects of short-term exposure discussed above.

Using the St George's Hospital Medical School database, described above, we identified 165 time-series studies of PM_{10} and daily mortality from all causes at all ages in all languages and countries, up to the end of July 2001. As we were concerned about the possibility of publication bias, we compared the summary estimates from 54 individual studies, a subset of the literature which would be expected to be susceptible to publication bias, with those of the two multi-city studies (Health Effects Institute 2000b; Katsouyanni et al. 2001), which selected cities from a pre-specified sampling frame, used uniform methods of analysis, and published all results. The pooled estimate for the cities of the combined APHEA and NMMAPS studies (n=111) was 1.005 (95% CI 1.004–1.006) with no evidence of publication bias in the funnel plot or on statistical testing. For the 54 studies of individual cities, graphical analysis showed some evidence of publication bias but when formally tested, this was weak (P=0.12). The pooled estimate was 1.007 (95% CI 1.006–1.008) but when adjusted for publication bias using Trim and Fill analysis, it was reduced to 1.006 (95% CI 1.004–1.007). We then examined the results for all 165 studies with results for $PM_{10.}$ There was

no evidence of publication bias on inspection of the funnel plot or on formal testing with Begg's or Eggar's tests (Begg and Mazumdar 1994). We calculated pooled estimates weighted according to the inverse of the variance of the individual study. Random effects models were used, as all showed significant heterogeneity. The pooled estimate was 1.006 (95% CI 1.005–1.007) (Table 17.7). This concentration–response function was applied to all GBD baseline cause-specific rates of mortality except GBD code 148 (injuries).[10]

4. Uncertainty estimates: statistical variability and sensitivity analyses

The total uncertainty in our estimates of the burden derives from both the statistical (sampling) variability of the parameter estimates in the models we chose to quantify disease burden, and our uncertainty with regard to those choices vs plausible alternatives, i.e. the form of our models (Morgan and Henrion 1998). We therefore quantified the statistical uncertainty of our estimates in terms of a combined, or propagated, uncertainty estimate, and used sensitivity analyses to quantify model uncertainty.

4.1 Statistical (sampling) variability

Our estimate of statistical uncertainty combined the sampling errors from two sources to derive an uncertainty distribution:

- sampling variability in the original concentration–response estimates from the ACS and time-series studies quantified in terms of their standard errors (Table 17.5); and
- sampling variability in the estimates of subregional concentration of PM from the exposure estimation model in terms of estimates of bootstrapped standard error described above (Table 17.4).

When presenting our results we show either the complete uncertainty distribution or the intervals between the 25th and 75th and/or 2.5th and 97.5th percentiles of that distribution, i.e. 50% and 95% uncertainty intervals.

4.2 Sensitivity analyses

We used sensitivity analyses, described below, to quantify the uncertainty in our base-case estimates, in which the burden of disease was estimated by applying the ACS concentration–response function over the range of 7.5 to 50 $\mu g/m^3$, as discussed above.

- *Cases 2–4: Shape of the concentration–response function for $PM_{2.5}$ and mortality attributable to cardiopulmonary disease and lung cancer.* We explored three alternatives to the base-case scenario for extrapolating the ACS concentration–response function beyond

30 μg/m^3, the highest annual average concentration observed in the ACS study (Figure 17.7).

- *Case 2: No incremental increase in excess mortality above 30 μg/m^3 $PM_{2.5}$.* Under this scenario, when calculating the population-weighted subregional average concentration of $PM_{2.5}$ we give the city-specific C_m a value of 30 μg/m^3, regardless of the estimated concentration, rather than the base-case concentration of 50 μg/m^3. This means that the excess risk is constrained to be no greater than that associated with an annual average concentration of 30 μg/m^3, regardless of the actual estimated annual average concentration. The counterfactual or theoretical minimum concentration was set at 7.5 μg/m^3. We considered this the estimator that would produce the smallest (i.e. most scientifically conservative) estimate of the impact of mortality consistent with the use of the ACS concentration–response function.
- *Case 3: Excess mortality increases linearly above 30 μg/m^3 $PM_{2.5}$.* Under this scenario, the city-specific concentration of $PM_{2.5}$ takes its actual estimated value when calculating the population-weighted subregional averages, i.e. in contrast to the base-case and case 2 scenarios. The counterfactual or theoretical minimum concentration was set at 7.5 μg/m^3. We considered this the estimator that would produce the largest estimate of mortality impact consistent with the use of the ACS concentration–response function.
- *Case 4: Excess mortality increases with the log of concentration of $PM_{2.5}$ across the entire range.* Under this scenario, the city-specific concentration of $PM_{2.5}$ takes the log of its actual estimated value when calculating the population-weighted subregional averages. Therefore, in contrast to case 3, the slope of the concentration–response function is constrained to decrease at higher concentrations. The counterfactual or theoretical minimum concentration was set at 7.5 μg/m^3.

  We included this estimator, proposed by an external reviewer after we had made our initial estimates (R. Burnett, personal communication), because it seemed a reasonable way to characterize an excess risk that we believed may: (i) increase directly with ambient levels over the entire range of annual average concentrations that we estimated for the world's cities, but (ii) be smaller at higher ambient concentrations, as has been observed for daily mortality in time-series studies (Daniels et al. 2000; Schwartz et al. 2002).
- *Cases 5 and 6: Choice of ACS concentration–response function.* In the base-case analyses, we used the ACS coefficients that were based on exposure of the cohort in 1979–1983. These arguably best represented the effects of long-term past exposure that some researchers assume are responsible for the increased mortality attributable to air pollution in that cohort through to 1998. There is, however, consid-

erable uncertainty regarding the timing of exposure with regard to risk of mortality (Krewski et al. 2000), so we calculated alternative estimates using the reported ACS coefficients based on the average of past (1979–1983) and recent (1999–2000) annual average concentrations using both a linear (case 5) and log-linear (case 6) extrapolation (Table 17.7).

- *Case 7: Change $PM_{2.5}$:PM_{10} ratio*. In the base-case analyses, we assumed a $PM_{2.5}$:PM_{10} ratio of 0.50, although higher and lower ratios have been observed in a number of locations, as discussed above. We examined the sensitivity of our base-case analyses by assigning cities in AMR-A, EUR-A, EUR-B, EUR-C and WPR-A a higher scaling factor of 0.65, while assigning a lower scaling factor of 0.35 to cities in all other subregions.
- *Cases 8 and 9: Choice of counterfactual or theoretical minimum concentration*. We evaluated the sensitivity of the base-case estimates to two different choices of counterfactual $PM_{2.5}$ concentration: 3 μg/m^3 (case 8) and 15 μg/m^3 (case 9). The former is close to the minimum background level of $PM_{2.5}$ observed in the United States, and the latter is the annual concentration of $PM_{2.5}$ proposed by the United States National Ambient Air Quality Standard (NAAQS) (U.S. Environmental Protection Agency 2002). Each value was substituted for the base-case concentration of 7.5 μg/m^3 in Equation 7 above, when calculating population-weighted subregional relative risks.

5. Results

5.1 Base-case estimates

We estimated that exposure to particulate air pollution caused approximately 800 000 excess deaths and 6.4 million DALYs (consisting only of years of life lost to premature mortality) in the year 2000 worldwide as a result of cardiopulmonary disease, lung cancer and ARI in children aged 0–4 years, combined (Table 17.8).

The worldwide estimates of attributable deaths and YLL from cardiopulmonary and lung cancer are subject to uncertainty contributed by the estimation of both the relative risks and the ambient concentrations of PM (Figure 17.8).

Cardiopulmonary disease in adults aged ≥30 years contributed 89% (712 000) and 78% (4.97 million) of attributable deaths and burden, respectively. Lung cancer contributed 8% and 9% and ARI in children, contributed 3% and 7% of deaths and YLL, respectively, to the total burden (Tables 17.9[a]–17.9[c]). The number of attributable deaths from all natural causes estimated from the daily time-series studies was roughly half the total attributable deaths, 378 000 vs 799 000 (Table

Table 17.8 Attributable deaths and DALYs in 2000, by subregion (50% and 95% confidence intervals)

Subregion	Deaths (000s)	50% CI	95% CI	DALYs (000s)	50% CI	95% CI
AFR-D	22	11.1–23.7	4.3–34.5	285	155.1–361.3	28.2–557.5
AFR-E	10	7.5–14.4	3.9–22.2	147	107.7–239.9	24.7–364.7
AMR-A	28	22.1–33.7	12.7–44.0	152	158.6–239.8	94.6–314.8
AMR-B	30	23.1–37.4	12.0–50.1	232	241.8–383.6	142.9–517.7
AMR-D	5	3.2–5.3	1.6–7.6	44	34.2–62.6	14.2–87.3
EMR-B	8	4.0–8.4	2.1–13.5	77	45.6–93.4	25.0–149.2
EMR-D	51	31.2–56.2	17.0–73.0	558	384.5–737.5	163.1–970.4
EUR-A	23	19.4–29.7	9.9–42.8	117	125.7–187.4	65.8–265.4
EUR-B	38	26.7–44.4	14.8–58.5	288	241.3–386.6	138.5–507.0
EUR-C	46	28.1–53.4	10.1–83.3	320	229.6–432.1	81.9–676.0
SEAR-B	32	19.2–37.5	5.5–51.5	282	191.0–388.9	67.0–532.6
SEAR-D	132	98.3–162.1	54.1–212.3	1 312	1 185.1–1 890.5	575.2–2 409.8
WPR-A	18	13.2–21.4	6.7–28.5	84	84.3–137.0	42.9–182.0
WPR-B	355	260.8–424.8	142.8–555.1	2 504	2 447.4–3 848.7	1 431.2–5 014.1
World	799	574.8–942.5	318.2–1 196.9	6 404	5 955.6–9 288.2	3 199.9–11 472.4

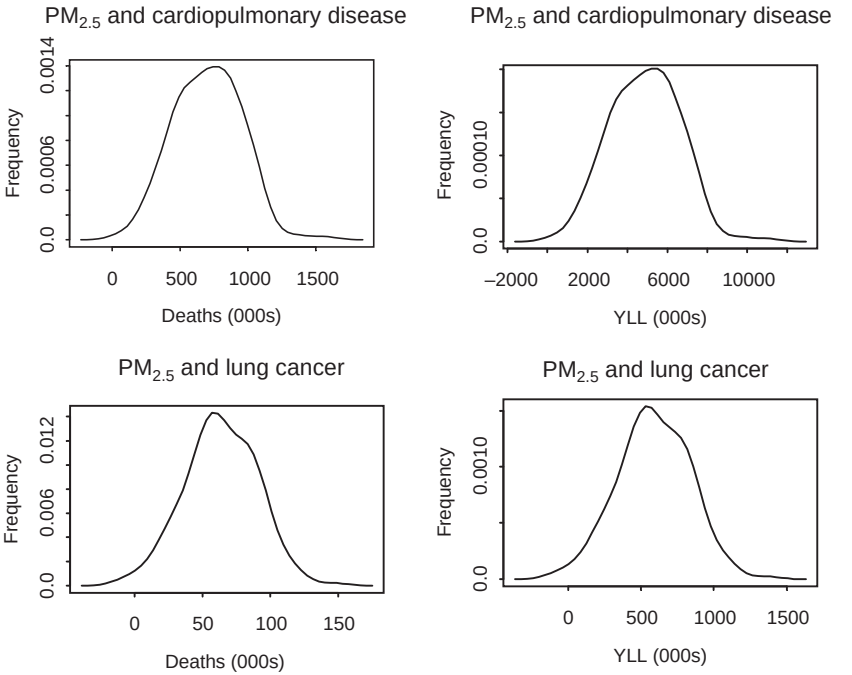
17.9[d]). The overall and cause-specific burden of disease varies across the 14 subregions, with the preponderance of the burden of air pollution contributed by cities in WPR-B, which includes China, and SEAR-D, which includes India. The variation in attributable deaths and YLL among the 14 subregions seen in Tables 17.9(a)–(d) reflects a subregional variation in the attributable fraction of approximately six-fold. For example, for all mortality end-points, EUR-A and WPR-B lie at the low and high ends, respectively, of the subregional distribution of attributable fractions. This largely reflects differences in the estimated population-weighted subregional ambient concentrations of PM_{10} and $PM_{2.5}$ (89 vs 26 $\mu g/m^3$ PM_{10} (see Table 17.4), rather than the proportion of the population that resides in cities. The proportion of the population of WPR-B that lives in cities is, in fact, lower than that in EUR-A (34% vs 39%).

5.2 Sensitivity analyses

Cases 2–4: Shape of the concentration–response function for $PM_{2.5}$

The estimates of attributable deaths from cardiopulmonary disease and lung cancer and YLL under the base-case and three alternative scenarios for the shape of the $PM_{2.5}$ concentration–response function are presented in Table 17.10. When the city-specific estimated concentrations

Figure 17.8 Uncertainty distributions for deaths and YLL from cardiopulmonary disease and lung cancer worldwide



of $PM_{2.5}$ are constrained to never exceed the concentrations observed in the most polluted city in the ACS study (annual average concentration of $PM_{2.5}$ of 30 μg/m^3), case 2, worldwide estimates of the number of deaths from cardiopulmonary disease and lung cancer are reduced by 29% and 27%, respectively. Extrapolation of the ACS coefficients to the highest estimated city-specific concentrations of $PM_{2.5}$ on the linear and logarithmic scales, cases 3 and 4, respectively, results in increases of 10% and 12% in the estimated number of attributable deaths from cardiopulmonary disease, and 8–24% increases in the estimated numbers of attributable deaths from lung cancer, relative to the base-case estimates.

These changes in worldwide estimates reflect underlying differences in the subregion-specific estimates (Table 17.11). Truncating the city-specific annual average concentrations at a given level leaves the burden unchanged in subregions with cities with estimated concentrations of PM that are lower than the truncation point, while reducing the burden in subregions with cities with estimated concentrations of PM that are above that point. Most cities in Europe and North America have

Table 17.9(a) Attributable deaths and YLL: base-case scenario for cardiopulmonary disease (50% and 95% confidence intervals)

Subregion	*Relative risk*	*Attributable fraction (%)*	*Deaths (000s)*	*50% CI*	*95% CI*	*YLL (000s)*	*50% CI*	*95% CI*
AFR-D	1.148	2	18	8.7–20.4	4–31	162	80.2–188.3	36–283
AFR-E	1.071	1	9	6.1–11.7	3–22	84	59.0–113.8	25–210
AMR-A	1.029	2	23	17.5–29.0	8–38	116	87.6–144.7	41–190
AMR-B	1.066	3	27	21.3–35.5	10–47	201	157.4–262.7	77–347
AMR-D	1.108	3	4	2.7–5.0	1–8	31	20.1–36.2	9–60
EMR-B	1.075	3	8	3.7–8.3	1–13	65	32.0–70.9	11–111
EMR-D	1.214	4	45	27.7–51.0	12–72	386	237.2–437.8	106–618
EUR-A	1.032	1	20	15.7–26.5	8–36	90	69.2–116.7	34–156
EUR-B	1.098	3	34	24.3–41.4	11–53	238	167.9–286.4	78–370
EUR-C	1.046	2	43	26.4–52.4	12–83	291	180.0–356.9	83–565
SEAR-B	1.250	4	30	17.1–33.8	5–50	240	136.9–271.2	38–400
SEAR-D	1.190	3	119	88.0–152.7	44–198	1 006	744.7–1 293.0	373–1 677
WPR-A	1.051	3	15	10.7–18.8	5–26	65	46.9–81.9	24–115
WPR-B	1.216	6	317	235.2–402.6	105–524	1 992	1 477.0–2 528.5	658–3 290
World		3	712	507.0–874.7	245–1 107	4 966	3 537.2–6 083.2	1 695–7 700

Table 17.9(b) Attributable deaths and YLL: base-case scenario for lung cancer (50% and 95% confidence intervals)

Subregion	*Relative risk*	*Attributable fraction (%)*	*Deaths (000s)*	*50% CI*	*95% CI*	*YLL (000s)*	*50% CI*	*95% CI*
AFR-D	1.210	5	0.41	0.2–0.5	0.06–0.78	4.2	1.9–5.0	0.6–8.0
AFR-E	1.100	2	0.27	0.2–0.4	0.05–0.69	2.9	2.1–4.2	0.5–7.4
AMR-A	1.040	3	4.85	3.8–6.7	0.75–8.48	36.9	28.9–50.8	5.7–64.7
AMR-B	1.093	4	2.10	1.7–2.9	0.39–4.25	19.9	16.0–27.5	3.7–40.3
AMR-D	1.153	6	0.15	0.1–0.2	0.02–0.27	1.5	1.0–1.8	0.2–2.7
EMR-B	1.105	4	0.45	0.2–0.5	0.05–0.86	4.7	2.1–5.3	0.5–8.8
EMR-D	1.309	8	1.55	1.0–1.9	0.18–2.70	16.8	10.4–20.6	2.0–29.3
EUR-A	1.044	2	3.52	2.8–5.0	0.58–6.61	27.4	22.1–38.9	4.5–51.5
EUR-B	1.138	5	3.00	2.1–3.9	0.41–5.43	30.2	21.5–38.9	4.1–54.7
EUR-C	1.065	3	2.82	1.6–3.6	0.52–5.63	27.4	15.7–34.8	5.1–54.8
SEAR-B	1.363	6	2.17	1.1–2.7	0.28–4.04	21.8	11.4–26.8	2.8–40.6
SEAR-D	1.272	4	5.63	4.3–7.8	0.90–11.60	55.9	42.4–77.2	8.9–115.2
WPR-A	1.071	4	2.71	2.0–3.7	0.45–4.94	17.6	13.3–23.9	2.9–32.1
WPR-B	1.311	10	32.37	24.5–44.0	5.15–59.01	308.5	233.2–419.1	49.1–562.4
World		5	62	47.0–83.1	9.95–114.32	576	436.9–767.5	92–1 063

Table 17.9(c) Attributable deaths and YLL: base-case scenario for ARI in children aged 0–4 years (50% and 95% confidence intervals)

Subregion	*Relative risk*	*Attributable fraction (%)*	*Deaths (000s)*	*50% CI*	*95% CI*	*YLL (000s)*	*50% CI*	*95% CI*
AFR-D	1.050	0.8	3.5	0.8–4.9	–2.9–9.0	119	26.8–165.6	–96.0–301.2
AFR-E	1.024	0.3	1.8	0.7–3.4	–2.0–5.8	61	22.6–113.5	–65.8–193.3
AMR-A	1.010	0.0	0.0	0.0–0.0	0.0–0.0	0	0.1–0.3	–0.2–0.5
AMR-B	1.023	0.3	0.3	0.1–0.6	–0.3–0.8	11	3.9–19.1	–11.3–27.4
AMR-D	1.037	0.9	0.3	0.1–0.5	–0.3–0.9	11	3.2–18.2	–10.9–30.3
EMR-B	1.026	0.6	0.2	0.1–0.3	–0.2–0.6	7	1.8–9.6	–6.3–20.7
EMR-D	1.070	1.4	4.6	1.5–7.0	–4.1–11.5	155	51.5–237.9	–137.6–389.2
EUR-A	1.011	0.0	0.0	0.0–0.0	0.0–0.0	0	0.0–0.1	–0.1–0.2
EUR-B	1.033	0.7	0.6	0.2–1.0	–0.6–1.5	20	7.0–33.1	–20.1–51.1
EUR-C	1.016	0.1	0.1	0.0–0.1	–0.1–0.2	2	0.6–3.5	–1.9–5.9
SEAR-B	1.082	0.5	0.6	0.2–1.0	–0.6–1.7	21	5.4–35.3	–19.7–58.7
SEAR-D	1.063	0.6	7.4	2.6–13.3	–7.2–20.1	250	86.6–448.4	–243.8–678.5
WPR-A	1.018	0.0	0.0	0.0–0.0	0.0–0.0	0	0.1–0.2	–0.1–0.4
WPR-B	1.071	1.2	6.1	2.0–10.7	–6.1–16.5	204	68.8–358.8	–203.3–555.0
World		0.7	25.6	8.2–43.9	–23.7–66.1	862	277.7–1 480.7	–798.6–2 228.0

Table 17.9(d) Attributable deaths: base-case scenario for mortality from all natural causes

Subregion	*Relative risk*	*Attributable fraction (%)*	*Deaths (000s)*
AFR-D	1.029	0.67	26
AFR-E	1.015	0.29	16
AMR-A	1.006	0.42	11
AMR-B	1.014	0.68	15
AMR-D	1.022	0.90	4
EMR-B	1.015	0.62	4
EMR-D	1.042	1.20	37
EUR-A	1.007	0.26	10
EUR-B	1.019	0.74	14
EUR-C	1.009	0.43	13
SEAR-B	1.048	0.87	17
SEAR-D	1.037	0.64	70
WPR-A	1.011	0.68	7
WPR-B	1.042	1.43	133
World		0.75	378

estimated annual average concentrations of $PM_{2.5}$ of $<30\,\mu g/m^3$, while estimated concentrations in the cities of developing countries are frequently much greater. More than 95% of the decrease in the worldwide burden in case 2 and the increase in case 3 occurs in four subregions: WPR-B, EMR-D, SEAR-B and SEAR-D.

Log-linear specification of the concentration–response function, as in case 4, allows for a more gradual increase in the relative risk at concentrations of PM of $>30\,\mu g/m^3$ than does the linear extrapolation model of case 3. This specification also means that the relative risk increases more steeply at concentrations of $<30\,\mu g/m^3$. Since the estimates for burden in both the log-linear case and the base case are measured with reference to a counterfactual annual average concentration of $PM_{2.5}$ of $7.5\,\mu g/m^3$, the burden under the log-linear specification is higher than that under the base case at low levels of exposure, but lower than the base case at high levels of exposure. Differences in the subregion-specific estimates for burden under the log-linear specification relative to the base case depend on the subregion-specific distributions of the city-specific concentrations of PM. The burden of disease in subregions where exposure is relatively low (AMR-A, EUR-A, EUR-C and WPR-A) increases by 63%, relative to the base case, while the burden in subregions where exposure is high remains unchanged or is slightly reduced.

Table 17.10 Sensitivity analyses of base-case estimates of attributable deaths and YLL, by cause

		Cardiopulmonary disease		*Lung cancer*		*Acute respiratory infections*[a]	
Case	*Conditions*	*Attributable deaths (000s) (% change)*	*YLL (000s) (% change)*	*Attributable deaths (000s) (% change)*	*YLL (000s) (% change)*	*Attributable deaths (000s) (% change)*	*YLL (000s) (% change)*
Base-case	Maximum concentration of $PM_{2.5} = 50\,\mu g/m^3$	712	4966	62	576	26	862
Case 2	Maximum concentration of $PM_{2.5} = 30\,\mu g/m^3$	506 (–29)	3498 (–30)	45 (–27)	414 (–28)	NA	NA
Case 3	Linear extrapolation	783 (10)	5507 (11)	67 (8)	623 (8)	NA	NA
Case 4	Log-linear extrapolation	794 (12)	5476 (10)	77 (24)	698 (21)	NA	NA
Case 5	Change ACS coefficient/linear extrapolation	1132 (59)	7908 (59)	101 (63)	939 (63)	NA	NA
Case 6	Change ACS coefficient/log-linear extrapolation	1069 (50)	7385 (49)	105 (69)	955 (66)	NA	NA
Case 7	Change $PM_{2.5}$: PM_{10} ratio	609 (–15)	4109 (–17)	58 (–7)	521 (–10)	NA	NA
Case 8	Theoretical minimum concentration of $PM = 3\,\mu g/m^3$	882 (24)	6081 (23)	80 (29)	731 (27)	30 (–15)	1012 (–17)
Case 9	Theoretical minimum concentration of $PM = 15\,\mu g/m^3$	474 (–33)	3365 (–32)	39 (–37)	369 (–36)	19 (27)	627 (27)

NA Not applicable.

[a] In children aged –4 years.

Table 17.11 Subregion-specific estimates for number of deaths and YLL from cardiopulmonary disease, lung cancer and ARI[a] for base-case and alternative scenarios

	Cardiopulmonary deaths (000s) and YLL (000s)																	
	Base-case		*Case 2*		*Case 3*		*Case 4*		*Case 5*		*Case 6*		*Case 7*		*Case 8*		*Case 9*	
Subregion	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*
AFR-D	18	162	13	123	19	179	19	180	28	259	26	243	12	106	21	194	12	112
AFR-E	9	84	8	79	9	84	12	113	14	132	16	152	5	44	12	115	4	41
AMR-A	23	116	23	116	23	116	41	205	36	181	55	273	40	202	44	220	3	13
AMR-B	27	201	25	187	27	203	35	261	43	315	47	349	15	109	37	277	14	106
AMR-D	4	31	4	28	4	31	5	38	7	49	7	51	2	18	5	39	3	18
EMR-B	8	65	7	62	8	65	10	89	12	102	14	120	4	34	10	88	4	30
EMR-D	45	386	26	219	64	545	46	397	72	617	63	537	37	317	51	440	36	306
EUR-A	20	90	20	90	20	90	34	151	32	141	46	202	35	154	37	164	4	18
EUR-B	34	238	33	229	34	238	44	304	54	375	59	408	49	339	44	304	20	136
EUR-C	43	291	43	291	43	291	68	460	67	457	91	616	67	458	66	449	11	74
SEAR-B	30	240	17	132	36	287	28	226	49	390	39	309	22	174	34	269	24	191
SEAR-D	119	1 006	79	667	136	1 150	123	1 037	192	1 625	167	1 413	82	697	137	1 164	88	749
WPR-A	15	65	15	65	15	65	23	99	23	102	30	132	23	101	23	98	5	21
WPR-B	317	1 992	192	1 209	344	2 163	305	1 915	504	3 163	411	2 580	216	1 357	360	2 261	247	1 550
World	712	4 966	506	3 498	783	5 507	794	5 476	1 132	7 908	1 069	7 385	609	4 109	882	6 081	474	3 365

continued

Table 17.11 Subregion-specific estimates for number of deaths and YLL from cardiopulmonary disease, lung cancer and ARI[a] for base-case and alternative scenarios (*continued*)

	Lung cancer deaths (000s) and YLL (000s)																	
	Base-case		*Case 2*		*Case 3*		*Case 4*		*Case 5*		*Case 6*		*Case 7*		*Case 8*		*Case 9*	
Subregion	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*	*Deaths*	*YLL*
AFR-D	0.4	4.2	0.3	3.2	0.5	4.7	0.5	5.0	1	7	1	7	0.3	2.8	0	5	0	3
AFR-E	0.3	2.9	0.3	2.7	0.3	2.9	0.4	4.1	0	5	1	6	0.1	1.5	0	4	0	1
AMR-A	4.8	36.9	4.8	36.9	4.8	36.9	9.1	69.5	8	59	12	94	8.4	64.4	9	70	1	4
AMR-B	2.1	19.9	2.0	18.6	2.1	20.1	2.9	27.5	3	32	4	37	1.1	10.8	3	27	1	11
AMR-D	0.1	1.5	0.1	1.4	0.1	1.5	0.2	2.0	0	2	0	3	0.1	0.9	0	2	0	1
EMR-B	0.5	4.7	0.4	4.5	0.5	4.7	0.7	6.8	1	7	1	9	0.2	2.4	1	6	0	2
EMR-D	1.6	16.8	0.9	9.5	2.2	23.8	1.7	18.4	3	28	2	25	1.3	13.8	2	19	1	13
EUR-A	3.5	27.4	3.5	27.4	3.5	27.4	6.3	49.2	6	44	9	67	6.0	46.9	6	50	1	6
EUR-B	3.0	30.2	2.9	29.0	3.0	30.2	4.1	41.0	5	49	6	56	4.3	43.0	4	39	2	17
EUR-C	2.8	27.4	2.8	27.4	2.8	27.4	4.8	46.2	5	44	6	63	4.4	43.1	4	42	1	7
SEAR-B	2.2	21.8	1.2	11.9	2.6	26.2	2.2	22.0	4	37	3	31	1.6	15.7	2	25	2	17
SEAR-D	5.6	55.9	3.7	36.8	6.4	64.0	6.2	61.6	9	94	9	86	3.9	38.5	7	65	4	41
WPR-A	2.7	17.6	2.7	17.6	2.7	17.6	4.4	28.3	4	28	6	38	4.2	27.1	4	26	1	6
WPR-B	32.4	308.5	19.6	186.7	35.2	335.2	33.2	316.1	53	503	45	433	22.0	209.6	37	350	25	240
World	62.0	576	45	414	67	623	77	698	101	939	105	955	58	521	80	731	39	369

	Deaths (000s) and YLL (000s) from acute respiratory infections					
	Base-case		Case 8		Case 9	
Subregion	Attributable deaths	YLL	Attributable deaths	YLL	Attributable deaths	YLL
AFR-D	3.5	118.8	4	141	2	82
AFR-E	1.8	61.2	3	84	1	30
AMR-A	0.0	0.2	0	0	0	0
AMR-B	0.3	10.6	0	15	0	6
AMR-D	0.3	11.2	0	14	0	7
EMR-B	0.2	7.3	0	10	0	3
EMR-D	4.6	154.8	5	176	4	123
EUR-A	0.0	0.1	0	0	0	0
EUR-B	0.6	20.2	1	26	0	12
EUR-C	0.1	2.0	0	3	0	1
SEAR-B	0.6	21.3	1	24	1	17
SEAR-D	7.4	249.9	9	288	6	188
WPR-A	0.0	0.1	0	0	0	0
WPR-B	6.1	204.3	7	231	5	159
World	25.6	862.1	30	1012	19	627

[a] In children aged 0–4 years.

Cases 5 and 6: choice of ACS coefficient

Linear extrapolation beyond concentrations of PM of 30 μg/m^3 of larger alternative coefficients from the ACS study on the basis of the average of ambient concentrations measured in 1979–1983 and 1999–2000 resulted in increases of 59% and 63% in deaths attributable to cardiopulmonary disease and lung cancer, respectively, relative to the base-case estimates. Log-linear extrapolation of the larger coefficients produced increases of 50% and 69% in the number of deaths attributable to cardiopulmonary disease and lung cancer, respectively (Table 17.10).

Attributable burdens increased in all subregions (Table 17.11). The differences between the linear and log-linear estimates followed the same subregional patterns as in cases 3 and 4, discussed above.

Case 7: Choice of $PM_{2.5}$:PM_{10} ratio

Allowing limited subregional variation in the ratio of $PM_{2.5}$ to PM_{10} produced reductions of 15% and 7% in the worldwide estimates of numbers of deaths attributable to cardiopulmonary disease and lung cancer, respectively, relative to the base-case scenario in which this ratio was fixed at 0.50 (Table 17.10).

As one might expect, the burden of disease increases by 57% in those subregions assigned a ratio of 0.65, that is, AMR-A, all of Europe, and WPR-B. This increase is more than offset by the rest of the world, assigned a ratio of 0.35, where the burden of disease falls by 31% (Table 17.11).

Cases 8 and 9: Choice of theoretical minimum level of exposure

Halving and doubling the base-case theoretical minimum concentration of $PM_{2.5}$ of 7.5 μg/m^3 resulted in a 24% increase and a 33% decrease in the number of deaths attributable to cardiopulmonary disease, and a 29% increase and a 37% decrease in deaths attributable to lung cancer, but only minor variations in mortality from ARI (Table 17.10).

All subregions experienced increases in attributable burden when the theoretical minimum concentration was halved, with the largest proportional increases in less polluted subregions (AMR-A, EUR-A and WPR-A). These subregions also experienced the largest reductions in burden when the theoretical minimum concentration was doubled. Highly polluted subregions (WPR-B and SEAR-D) also experienced marked reductions in the estimated burden when the theoretical minimum concentration was doubled (Table 17.11).

6. Discussion

Previously, most large-scale estimates of the health impacts of urban air pollution were conducted for countries or regions where data on expo-

sure and estimates of effect required for impact estimation were available (e.g. Brunekreef 1997; COMEAP 1998; Künzli et al. 2000; Ostro and Chestnut 1998). In the few previous global estimates, systematic methods were not applied to extrapolate exposures and exposure–response functions to other parts of the world (Hong 1995; WHO 1997; Working Group on Public Health and Fossil Fuel Combustion 1997). Although our estimates exceed those reported earlier, the differences are not large, given the variation in the approaches that were taken (Smith and Mehta 2003). The global scope of the present analysis required new approaches for estimating exposure, absent measurements of air pollution in many developing countries, extrapolating the results of epidemiological studies more widely than had previously been attempted, and describing and attempting to quantify, the many uncertainties this entailed. The results indicate that the impact of urban air pollution on the burden of disease in the cities of the world is large and, for a variety of reasons discussed below, have probably underestimated the burden. There is also considerable variation in our estimates among the 14 subregions, with the greatest burden occurring in the more polluted and rapidly growing cities of developing countries.

The availability of actual measurements of outdoor concentrations of PM varied widely across the globe. In order to have estimates for all 14 subregions, models developed by the World Bank were used to estimate concentrations of inhalable particles (PM_{10}) using economic, meteorological and demographic data and the available measurements of PM for 3211 cities with populations of >100 000, and also capital cities. To allow the most appropriate epidemiological studies to be used for the estimation of the burden of disease, the PM_{10} estimates were converted to estimates of fine particles ($PM_{2.5}$) using information on geographic variation in the ratio of $PM_{2.5}$ to PM_{10}. Population-weighted subregional annual average exposure estimates for $PM_{2.5}$ and for PM_{10} were obtained using the population of the city in the year 2000.

The estimates of the burden of disease were based on three health outcomes: mortality from cardiopulmonary causes in adults, mortality from lung cancer and mortality from ARI in children aged 0–4 years. Attributable numbers of deaths and YLL for adults and children (aged 0–4 years) were estimated using risk coefficients from a large cohort study of adults in the United States (Pope et al. 2002) and a meta-analytic summary of five time-series studies of mortality in children, respectively. Base-case estimates were calculated assuming that the risk of death increases linearly over a range of annual average concentrations of $PM_{2.5}$ between a counterfactual (or referent) concentration of 7.5 and a maximum of 50 $\mu g/m^3$. For comparison, an additional estimate of attributable deaths was calculated from time-series studies of daily mortality, on the basis of results of a meta-analysis of the world literature, but was not used in the final calculations. Worldwide and subregion-

specific estimates of attributable deaths and burden of disease in terms of YLL were calculated based on the standard methodology developed for this project (see chapters 1 and 25).

We estimated that urban air pollution, as measured by PM, is responsible for about 3% of mortality caused by cardiopulmonary disease in adults, about 5% of mortality caused by cancers of the trachea, bronchus and lung, and about 1% of mortality caused by ARI in children worldwide in the year 2000. The total burden was about 0.80 million (1.2% of total) premature deaths and 6.4 million (0.5% of total) DALYs. This burden occurred predominantly in developing countries, with 30% of attributable disease burden occurring in WPR-B and 19% in SEAR-D. The greatest contributions to the total burden of disease occurred in WPR-B, EUR-B and EUR-C, where urban air pollution caused 0.7–0.9% of the total burden of disease.

6.1 Identifying and quantifying uncertainty in the estimates

These estimates are subject to considerable uncertainty given the need to estimate exposures and to extrapolate concentration–response relationships. This is almost invariably the case in quantitative risk assessment of complex environmental exposures; and is certainly to be expected in this particular exercise, for reasons discussed above.

We quantified the statistical uncertainty of our base-case estimates by estimating the joint uncertainty in the estimates of annual average concentration and the estimates of the relative risks. Worldwide and most subregional estimates vary by less than two-fold (50% uncertainty intervals (Tables 17.8 and 17.9[a]–[c]). Uncertainty of the model owing to assumptions about the shape of the concentration–response function, the magnitude of the relative risk of disease attributable to urban air pollution, the choice of counterfactual level for PM, and the ratio of $PM_{2.5}$ to PM_{10} was assessed in sensitivity analyses. For the most part, the estimated worldwide burdens in the various sensitivity analyses are within the 50% uncertainty interval for the base-case estimate of worldwide burden. The sensitivity analyses indicate that base-case estimates were most sensitive to choice of coefficient from the ACS study and theoretical minimum concentration.

Although some sources of uncertainty could be quantified, others that were no less important or were perhaps more important, could not. These additional sources of uncertainty arise from the methods we used to estimate annual average exposure of the population and our choice of health end-points and concentration–response functions.

Estimates of exposure to urban air pollution

There are four key uncertainties related to exposure that have not been quantified, and that could affect the estimates of burden of disease.

First, we used PM as the sole indicator of exposure to urban air pollution, although urban air pollution is a complex mixture, as noted above. Other frequently measured pollutants, notably ozone, carbon monoxide, oxides of sulfur and nitrogen, and lead are associated with mortality and morbidity, albeit not as consistently as PM, although the effects of a number of these pollutants may be at least partially captured via the use of a PM metric (Sarnat et al. 2001). Estimating the health impacts of specific components poses challenges for both scientific research and risk assessment, including how to avoid attributing the same burdens to multiple pollutants (i.e. double counting), and how to quantify the effects of possible interactions (i.e. synergistic effects) among pollutants. Nonetheless, there is evidence for an effect of ozone on daily mortality that is independent of PM (e.g. Health Effects Institute 2000b). Future estimates of the burden of disease should include the health impacts of ozone. Unfortunately, lead in petrol remains an important toxic component of air pollution in some cities of the developing world and its contribution to the burden of disease has been estimated elsewhere in this book (see chapter 19). Some combination of the GBD estimates for urban air pollution and lead probably provide the best overall estimate of the burden of disease attributable to urban air pollution.

Second, use of estimated levels of exposure introduces some uncertainties and biases in the predicted levels of exposure that could not be addressed, owing to lack of data. The most important of these is the lack of city-specific data on the structure of economic activity and on fuel consumption. The exposure model uses national average data for these variables as a reasonable proxy, which can lead to bias in unknown directions with regard to city-specific estimates. The net bias in estimates of the aggregate burden at the subregional level is unclear. The use of long-run average climatic conditions instead of time-varying local data may result in biased estimates for specific years, but may not pose a serious problem as we are interested mainly in the long-term average health effects of air pollution. We have also explicitly examined our uncertainty regarding spatial variations in the size composition of PM through sensitivity analysis. The model for the estimation of exposure clearly suggests that coarser particles account for a larger fraction of the TSP in developing countries than in developed countries, all other things being equal. The limited data from monitoring available on $PM_{2.5}$ also indicate that spatial variations may also exist in the sizes of finer particles. Consequently, we have used conservative estimates for the fraction of PM_{10} accounted for by finer particles in our overall estimates and further tested the implications of using an even more conservative estimate. The burden estimates should be relatively insensitive to PM size fraction.

Third, misclassification of exposure may have led to underestimation of the burden of disease. Like the epidemiological studies used to quantify the estimates of health impact, we used the annual average ambient

concentration measured from a few stationary sources in each city to estimate average personal levels of exposure. Differences between personal levels of exposure and concentrations measured at fixed points depend on how well the pollutant mixes in the environment and the efficiency with which the pollutant penetrates indoors. The exposure estimates are based on a model developed from population-oriented monitors. Measurements of PM from these sites in well-designed monitoring networks would provide representative city-wide levels of exposure for a pollutant that mixes uniformly in the environment. They would underestimate the actual level of exposure of people living near pollution hotspots, such as busy roads or local sources of pollutant, which can contribute to spatial heterogeneity of exposure within cities (Hoek et al. 2002; Jerrett et al. 2001). This underestimate would probably be more pronounced for cities in developing countries, where nearly one third of the population resides in slums, which are often in heavily-polluted parts of cities, and even larger populations work near pollution hotspots.

Exposure misclassification from using outdoor concentrations to represent personal exposure to urban air pollution also results from differences in the efficiency with which PM penetrates indoors. Use of ambient concentrations as surrogates for exposure tends to underestimate the risk per unit exposure because the penetration of particles indoors, where most exposure occurs, is less than 100%. If average penetration is 66%, for example, actual exposure–response per $1\,\mu g/m^3$ would be 1.5-fold that indicated by outdoor concentrations of pollution. However, because of climate and housing, the rates of penetration of pollution in most, but not all, cities in developing countries can be expected to be somewhat greater than those in the average city in the United States where the epidemiology used here has been undertaken. Not being able to consider this factor because of lack of data on penetration of the pollutant would bias estimates of burden downward if actual changes in exposures in developing countries are better indicated by changes in outdoor concentrations than in developed countries.

An additional source of misclassification concerns the time referent of our exposure estimate. The current burden is related to past exposure, but our model estimates current (i.e. 1999) levels only. However, even if we had been able to retrospectively estimate a time series of annual average concentrations for each subregion, the ACS study provides little information as to how the concentration–response function varies over time (Krewski et al. 2000). It is not clear how this source of misclassification would affect our estimates.

Fourth, our estimates do not include the attributable burden of disease among the 800 million additional urban residents living either in suburban areas of some of the cities or in cities with populations of <100 000 and in the >3 billion residents of rural areas. Although lower levels of emissions per area combined with the differences in the built-up envi-

ronment in rural areas probably results in a small average exposure to ambient pollution and a modest increase in the global burden of disease from such pollution in rural areas, the same is not true for the urban residents that were not included in the target population. The magnitude of the missing burden depends on the actual exposures of those living in smaller cities. The target population was identified for the study on the basis of data available from the United Nations, which compiles the data reported by Member States from national censuses and makes projections from them on the basis of expected changes in demographics. In compiling the population statistics, Member States, hence the United Nations, do not use uniform definitions either for city area (the characteristic such as city size that defines an urban area) or the population included for each identified city (whether political boundaries are used or agglomerations of contiguous urban areas are used). For the target population, we have used the population of the city agglomeration when this choice was available. If all of the remaining 800 million residents lived in suburban areas next to a targeted city, exposures and hence, estimates of burden, for these residents could be expected to be similar to those in the identified city resulting in an aggregate underestimate of the attributable fraction of the population of about 28%. The exposure model suggests, however, that concentrations gradually decrease as the local population density decreases, suggesting that levels of exposure and hence estimates of burden are lower for these residents compared to those living in larger cities. The net result is that our focus on residents in cities with >100 000 inhabitants may underestimate the aggregate burden by between 0% and 28%.

Choice of health end-points and concentration–response functions

Our base-case estimates of burden in terms of disease burden considered only the impact of air pollution on mortality. This approach is likely to have underestimated the true attributable burden, since there is evidence from studies of both epidemiology and toxicology, to suggest that air pollution may play a role in the incidence of cardiopulmonary disease, and thus contribute to years lived with disability (YLD). Lacking estimates of the concentration–response function for air pollution and the incidence of cardiopulmonary disease, lung cancer, and ARI in children, we calculated disease burden under the assumption that air pollution multiplies both incidence and mortality to the same extent, i.e. the relative risk of unobserved morbidity equals the observed relative risk of mortality (Table 17.12), an approach taken to estimating the attributable burden caused by other factors other than urban air pollution. The total disease burden, including YLD for cardiopulmonary disease, exceeds the YLL by 23% worldwide in the base-case analyses. The effect on the estimated burden for lung cancer and ARI in children is negligible.

Table 17.12 Attributable YLL and DALYs for cardiopulmonary disease, lung cancer, ARI[a] and total mortality

	Cardiopulmonary disease		*Lung cancer*		*Acute respiratory infections*		*Total*		
	YLL	*DALYs*	*YLL*	*DALYs*	*YLL*	*DALYs*	*YLL*	*DALYs*	
Subregion	*(000s)*		*(000s)*		*(000s)*		*(000s)*		*% change*
AFR-D	162	193	4	4	119	121	285	319	12
AFR-E	84	100	3	3	61	62	147	166	13
AMR-A	116	161	37	38	0	0	152	200	32
AMR-B	201	273	20	20	11	14	232	307	32
AMR-D	31	39	1	2	11	12	44	53	21
EMR-B	65	77	5	5	7	9	77	91	18
EMR-D	386	457	17	17	155	162	558	636	14
EUR-A	90	122	27	28	0	0	117	151	29
EUR-B	238	286	30	31	20	21	288	338	17
EUR-C	291	340	27	28	2	2	320	360	13
SEAR-B	240	291	22	22	21	25	282	339	20
SEAR-D	1 006	1 195	56	57	250	261	1 312	1 513	15
WPR-A	65	95	18	18	0	0	84	114	36
WPR-B	1 992	2 732	304	317	204	224	2 504	3 272	31
World	4 966	6 360	572	591	862	913	6 404	7 865	23

[a] In children aged 0–4 years.

The estimates of the attributable burden caused by cardiopulmonary disease and lung cancer were derived from a single cohort study in the United States (the largest and most extensively reviewed study suitable for the estimation of the burden of disease). This raises questions concerning whether these results can be generalized to other populations, especially those in developing countries, owing to differences in susceptibility to the effects of air pollution and the nature of the mixture of air pollutants. The apparent qualitative and quantitative similarity of the relative risks of daily mortality in developed and developing countries, discussed above, provides some evidence that these results are generally applicable. In addition, trends in known risk factors for chronic cardiovascular and respiratory disease, such as diet and cigarette smoking, suggest that the populations of cities in developing countries may now be more comparable to populations of cities in Europe and North America with regard to susceptibility to air pollution conferred by pre-existing cardiovascular and respiratory morbidity (Reddy and Yusuf 1998). The increasing contribution of mobile sources to urban air pollution in developing countries also increases the similarity with cities in North America and Europe.

Other sources of uncertainty in our estimates cannot be readily quantified for the following reasons:

- *Lack of knowledge concerning differences between developed and developing countries in the physicochemical nature of PM produced by different sources*. The relative toxicity of PM may well vary according to the type of fuel burned and the type technology used to burn it. Increased burning of refuse outdoors and the prevalence of motor vehicles without emissions controls (e.g. vehicles powered by two-stroke engines) are two examples.

  Inhalable particles that are not the direct or indirect product of combustion sources may also be important. These particles are mainly of crustal origin and may be important, for example, in desert areas, or where there is disturbance of surface material owing to construction, use of unsurfaced roads, etc. They are largely found in the coarse fraction of inhalable particles, whereas combustion-derived particles tend to be found in the fine and ultra-fine fractions. The evidence concerning the toxicity of this fraction is mixed (Anderson 2000). Data on worldwide variation in the ratio of fine to coarse particles is limited, as discussed above, and our sensitivity analyses, which suggest relatively minor differences with our base-case estimates, may understate the uncertainty.

- *Lack of knowledge concerning differences in the susceptibility of the population*. Despite the trends discussed above, differences in demography and in the patterns of the incidence and prevalence of disease may be associated with differences in short-term and long-term vulnerability to air pollution. There exists the possibility of effect-modification factors related to health status, and behavioural factors, such as smoking and diet (Katsouyanni et al. 2001). The effects of previous or concurrent exposure to high levels of indoor air pollution may also play a role in determining susceptibility to urban air pollution. Poverty, which is a determinant of the factors just discussed, may also determine susceptibility in other ways. If the effects of air pollution are more severe among the poor, who comprise a large part of the world's population, then the magnitude of the burden would likely be greater than that which we estimated (Krewski et al. 2000; O'Neill et al. 2003).

- *The shape of the exposure–response relationship may differ between developing and developed countries in ways that were not captured in the sensitivity analyses*. For example, a recent time-series study of daily mortality in Mexico City did not observe a flattening of the PM_{10} concentration–response curve until 175 $\mu g/m^3$ (the daily mean) (Tellez-Rojo et al. 2000). These concentrations are measured in many mega-cities in developing countries.

We did not know which form of the concentration–response relationship should be used in extrapolating the results of the ACS study to the much higher concentrations observed in cities in India and China, for example. For this reason, we examined the sensitivity of the estimates to a range of scenarios, presenting a "base case", which we thought was a reasonable compromise between the conditions of the ACS and those of the rest of the world. Cohort evidence has recently been reported from Europe, although unfortunately it was unable to estimate concentrations of PM (Hoek et al. 2002). This study provides evidence that long-term exposure to urban pollution is associated with health effects elsewhere in developed countries, but we still lack cohort evidence from developing countries.

Mortality from acute respiratory infections in children aged 0–4 years

Despite limited evidence, discussed above, linking mortality from ARI to exposure to urban air pollution, we used the results from the small number of time-series studies in developing countries to estimate attributable deaths and YLL in children aged 0–4 years. In our view, most of these deaths are likely to be among children with temporary vulnerability owing to chest infections which would resolve eventually, and therefore represent, on average, the loss of many potential years of life, but this view is largely speculative.

Several studies that we used to derive the concentration–response function for ARI mortality actually reported results for all causes mortality in the 0–4 years age group (Ostro et al. 1999a), or total mortality in the first year of life (Loomis et al. 1999). We have assumed that the relationship between PM_{10} and mortality from ARI in children aged 0–4 years is similar to that for all-cause mortality. To some extent this is justified by the knowledge that mortality from ARI is an important component of all-cause mortality in developing countries.

6.2 Generalizability of our results

As a consequence of the uncertainties in this global assessment, its quantitative results cannot be confidently extrapolated to smaller geographic areas, such as specific countries or cities. The methods for estimation of exposure and extrapolation of concentration–response functions were developed specifically for estimating burdens for large geographic regions, often in the absence of essential data on exposure and response. Where better data exist, as they currently do in some parts of the world, they can, of course, be used.

Differences between our estimates and those of other groups may reflect other differences in methodology. For example, a tri-national European assessment recently estimated that some 40 000 deaths per year were attributable to exposure to ambient air pollution in a population of approximately 72 million, whereas the burden in EUR-A, in an urban

population of 80 million, was estimated to be 23 000 deaths per year, despite similar estimates of the concentration of ambient pollution (Künzli et al. 2000). The difference is largely owing to the different assumptions regarding the exposure reference level of 15 $\mu g/m^3$ PM_{10} in this work vs 7.5 $\mu g/m^3$ in the European study. In addition, the concentration–response functions were slightly higher in the tri-national project, which used estimates of total mortality from both the first ACS publication and the Harvard Six City estimates (Dockery et al. 1993; Pope et al. 1995).

6.3 Avoidable disease burden

We did not attempt to estimate the avoidable burden of disease, despite this being a specific objective of the project. Estimating the avoidable burden would have required making projections of concentrations of ambient air pollution and providing a model for the exposure time–response function for PM and mortality. Time constraints did not allow us to undertake the former task, although it is feasible. The latter information is not currently available from the existing cohort studies, although there is limited evidence that the induction time for mortality from lung cancer attributable to exposure to urban air pollution is in the order of decades (Nyberg et al. 2000), and that it is perhaps in the order of years for mortality from cardiovascular disease (Krewski et al. 2000). Evaluations of both "natural experiments" (Heinrich et al. 2000; Pope 1989), and regulatory interventions (Clancy et al. 2002; Hedley et al. 2002) provide further support for relatively rapid improvements in cardiovascular and respiratory outcomes. The latter studies also suggest that although rates of mortality may decrease after the successful implementation of air pollution reductions, the long-term benefits may extend well beyond that observed during the first years after the intervention is implemented.

6.4 How could a future risk assessment exercise provide better estimates?

There is a critical need for more information on the health effects of air pollution in developing counties. Research on exposure should aim to provide better estimates not only of ambient concentrations of pollutants, but also the characteristics of urban air pollution, including the contribution of various sources and the size distribution of PM. Epidemiological studies of mortality should be designed to provide age- and disease-specific estimates of the effects of air pollution, as well as identifying factors that confer susceptibility to air pollution. There is an obvious need for epidemiological studies of the effect of air pollution on the incidence of chronic cardiovascular and respiratory disease, and on the growth and development of children. Future estimates of the burden of disease attributable to urban air pollution should include morbidity outcomes, such as asthma exacerbation, which most certainly contribute to morbidity.

Estimates of uncertainty distributions should more fully incorporate model uncertainties, such as those related to the choice of concentration–response function. This could be accomplished via the elicitation and weighting of expert opinions in the context of a Bayesian approach to quantifying model uncertainty (Morgan and Henrion 1998; National Research Council 2002).

Acknowledgements

The authors wish to acknowledge the contribution of Richard Burnett, Uwe Deichmann, Majid Ezzati, Stephen Vander Hoorn and David Wheeler.

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Health Effects Institute (HEI) or its sponsors nor those of the World Bank.

Notes

1 Throughout the chapter we refer to urban air pollution using the terms "ambient air pollution" or "urban air pollution". For our current purposes, these terms are fully interchangeable.

2 See Preface for an explanation of this term.

3 Ambient particles fall into a trimodal size distribution, according to their aerodynamic diameter: coarse particles (>1 μm), fine particles (0.1–1 μm), and ultrafine particles (<0.1 μm). Ultrafine particles constitute a small percentage of the total mass of PM, but are present in very high numbers. Because of health concerns, the ambient concentrations (mass) of both coarse and fine PM are regulated by the United States Environmental Protection Agency (EPA) through the National Ambient Air Quality Standards for PM_{10} (PM <10 μm) and $PM_{2.5}$ (PM <2.5 μm) (USEPA 1997), and by the European Union through limit values for PM_{10}. $PM_{2.5}$, which includes only fine and ultrafine particles, is dominated by emissions from combustion processes; PM_{10}, which includes coarse as well as fine and ultrafine particles, has a much higher proportion of particles generated by mechanical processes from a variety of non-combustion sources. It is currently not clear how much particles of different sizes and composition differ in the effects on health that they cause.

4 Cities in the United States account for about 40% of the observations in the estimation model, even after this exclusion.

5 Data from monitoring were available for one additional city/country, Skopje in The former Yugoslav Republic of Macedonia, but were not used in the estimation model because of missing explanatory variables. In addition, 150 observations primarily from Germany (94), Lithuania (30) and other eastern European states (26) made during the early 1990s were excluded because of uncertainties in defining appropriate explanatory variables.

6 The model presented here is one of several versions of the GMAPS model developed at the World Bank. An alternative model jointly estimates concentrations of PM_{10} and TSP at residential and non-residential sites.

7 The climatic variables have been constructed from a global mean monthly climatology map with a resolution of 0.5° latitude × 0.5° longitude developed by researchers at the Climate Research Unit of the University of East Anglia. These data are available at http://ipcc-ddc.cru.uea.ac.uk/cru_data/examine/have_index.html. All of the climate variables are based on the conditions for the city centre.

8 Residential monitoring sites are located in residential areas but do not include pollution hotspots, such as locations that are immediately adjacent to industrial and commercial pollution sources or high traffic corridors. In contrast, mixed residential sites are characterized by both high population densities and the presence of some pollution sources that may result in elevated concentrations of PM in the immediate vicinity of the pollution source. Neither site includes areas of high pollution activity located in sparsely populated areas.

9 Had we instead included data from monitoring for cities with measured data for 1999, there would be an insignificant difference in the subregional average concentration, because a small fraction of the population in each subregion lives in monitored cities and most of the monitored cities are located in North America and western Europe, where the estimates of PM are more precise.

10 After these estimates had been made, investigators in the United States and Canada discovered several problems with the statistical software that had been used to estimate the relative risks associated with air pollution in the time-series studies (Health Effects Institute 2003). Correcting these problems reduced the magnitude of estimated relative risks and increased their standard errors.

References

Abbey DE, Nishino N, McDonnell WF et al. (1999) Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American Journal of Respiratory and Critical Care Medicine*, **159**:373–382.

Aerometric Information Retrieval System (2001) U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, AIRS. Available at http://www.epa.gov/oar/oaqps/airpdata.html#airs.

American Thoracic Society (ATS) (2000) What constitutes an adverse health effect of air pollution? *American Journal of Respiratory and Critical Care Medicine*, **161**:665–673.

Anonymous (1996a) Health effects of outdoor air pollution. Part 1. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. *American Journal of Respiratory and Critical Care Medicine*, **153**:3–50.

Anonymous (1996b) Health effects of outdoor air pollution. Part 2. Committee of the Environmental and Occupational Health Assembly of the American

Thoracic Society. *American Journal of Respiratory and Critical Care Medicine*, **153**:477–498.

Anonymous (1998) Worldwide variations in the prevalence of asthma symptoms: the International Study of Asthma and Allergies in Childhood (ISAAC). *European Respiratory Journal*. **12**:315–335.

Anderson HR (2000) Differential epidemiology of ambient aerosols. *Philosophical Transactions of the Royal Society of London, Series A*, **358**: 2771–2785.

Atkinson RW, Anderson HR, Sunyer J et al. (2001) Acute effects of particulate air pollution on respiratory admissions: results from APHEA 2 project. Air pollution and health: a European approach. *American Journal of Respiratory and Critical Care Medicine*, **164**:1860–1866.

Awasthi S, Glick HA, Fletcher RH, Ahmed N (1996) Ambient air pollution and respiratory symptoms complex in preschool children. *Indian Journal of Medical Research*, **104**:257–262.

Bacharova L, Fandakova K, Bratinka J, Budinska M, Bachar J, Gud-aba M (1996) The association between air pollution and the daily number of deaths: findings from the Slovak Republic contribution to the APHEA project. *Journal of Epidemiology and Community Health*, **50**:S19–21.

Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics*, **50**:1088–1101.

Bobak M, Leon DA (1999) The effect of air pollution on infant mortality appears specific for respiratory causes in the post neonatal period. *Epidemiology*, **10**:666–670.

Borja-Aburto VH, Castillejos M, Gold DR, Bierzwinski S, Loomis D (1998) Mortality and ambient fine particles in southwest Mexico City, 1993–1995. *Environmental Health Perspectives*, **106**:849–855.

Borja-Aburto VH, Loomis DP, Bangidiwala SI, Shy CM, Rascon-Pacheco RA (1997) Ozone, suspended particulates, and daily mortality in Mexico City. *American Journal of Epidemiology*, **145**:258–268.

Braga AL, Zanobetti A, Schwartz J (2001) The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities. *Journal of Occupational and Environmental Medicine*, **43**:927–933.

Brunekreef B (1997) Air pollution and life expectancy: is there a relation? *Occupational and Environmental Medicine*, **54**:781–784.

Burnett RT, Smith-Doiron M, Stieb D, Cakmak S, Brook JR (1999) Effects of particulate and gaseous air pollution on cardiorespiratory hospitalizations. *Archives of Environmental Health*, **54**:130–139.

Burney PG, Luczynska C, Chinn S, Jarvis D (1994) The European Community Respiratory Health Survey. *European Respiratory Journal*, **7**:954–960.

California Air Resources Board (2002) *Staff report: Public hearing to consider amendments to the ambient air quality standards for particulate matter and sulfates*. California Air Resources Board and Office of Environmental Health Hazard Assessment, Sacramento, CA.

Calle EE, Rodriguez C, Jacobs EJ et al. (2002) The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer*, **94**:2490–2501.

Castillejos M, BorjaAburto VH, Dockery DW, Gold DR, Loomis D (2000) Airborne coarse particles and mortality. *Inhalation Toxicology*, **12**:61–72.

CIESIN (2000) *Gridded population of the world (GPW), version 2*. Center for International Earth Science Information Network (CIESIN), Columbia University; International Food Policy Research Institute (IFPRI), World Resources Institute (WRI) Columbia University, Palisades, NY. Available at http://sedac.ciesin.columbia.edu/plue/gpw.

Chay KY, Greenstone M (1999) *The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession*. Unpublished document. (Working Paper No. 17.) Available from: Center for Labor Economics, University of California, Berkeley, CA.

Chay KY, Greenstone M (2001) *Air quality, infant mortality, and the Clean Air Act of 1970*. Unpublished document. (Working Paper No. 42.) Available from: Center for Labor Economics, University of California, Berkeley, CA.

Chhabra SK, Chhabra P, Rajpal S, Gupta RK (2001) Ambient air pollution and chronic respiratory morbidity in Delhi. *Archives of Environmental Health*, **56**:58–64.

Cifuentes L, Borja-Aburto VH, Gouveia N, Thurston G, Davis DL (2001) Assessing the health benefits of urban air pollution reductions associated with climate change mitigation (2000–2020): Santiago, São Paulo, Mexico City, and New York City. *Environmental Health Perspectives*, **109**: S419–425.

Cifuentes L, Vega J, Kopfer K, Lava LB (2000) Effect of the fine fraction of particulate matter versus the coarse mass and other pollutants on daily mortality in Santiago, Chile. *Journal of the Air Waste Management Association*, **50**:1287–1298.

Clancy L, Goodman P, Sinclair H, Dockery DW (2002) Effect of air pollution control on death rates in Dublin, Ireland: an intervention study. *The Lancet*, **360**:1210–1214.

Cohen AJ, Nikula K (1999) Health effects of diesel exhaust: laboratory and epidemiologic studies. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

COMEAP (1998) *Quantification of the effects of air pollution on health in the United Kingdom*. Department of Health Committee on the Medical Effects of Air Pollutants. Stationery Office, London.

COMEAP (2001) *Statement on long term effects of particles on mortality*. Available at http://www.doh.gov.uk/comeap/state.htm, or as a printable document at http://www.doh.gov.uk/comeap/statementsreports/longtermeffects.pdf.

Conceiao GMS, Miraglia SGEK, Kishi HS, Saldiva PNH, Singer JM (2001) Air pollution and child mortality: a time series study in São Paulo, Brazil. *Environmental Health Perspectives*, **109**:347–350.

Daniels MJ, Dominici F, Samet JM, Zeger SL (2000) Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. *American Journal of Epidemiology*, **152**:397–406.

Derwent RG (1999) Atmospheric chemistry. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Diesel Working Group (1995) *Diesel exhaust: a critical analysis of emissions, exposure, and health effects. Special report*. Health Effects Institute, Cambridge, MA.

Dockery DW, Pope AC, Xu X et al. (1993) An association between air pollution and mortality in six US cities. *New England Journal of Medicine*, **329**: 1753–1759.

Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJL, Comparative Risk Assessment Collaborating Group (2002). Selected major risk factors and global and regional burden of disease. *The Lancet*, **360**: 1347–1360.

Gao J (1993) [Relationship between air pollution and mortality in Dongcheng and Xicheng Districts, Beijing]. *Chung-Hua Yu Fang i Hsueh Tsa Chih [Chinese Journal of Preventive Medicine]*, **27**:340–343.

Gouveia N, Fletcher T (2000) Time series analysis of air pollution and mortality: effects by cause, age and socioeconomic status. *Journal of Epidemiology and Community Health*, **54**:750–755.

Grant LD, Shoaf CR, Davis JM (1999) Sources of air pollution. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Grønskei KE, Gram F, Hagen LO et al. (1997a) *URBAIR Urban air quality management strategy in Asia: Jakarta report*. Shah J, Nagpal T, eds. World Bank, Washington, DC.

Grønskei KE, Gram F, Hagen LO et al. (1997b) *URBAIR Urban air quality management strategy in Asia: Kathmandu Valley report*. Shah J, Nagpal T, eds. World Bank, Washington, DC.

Hajat S, Anderson HR, Atkinson RW, Haines A (2002) Effects of air pollution on general practitioner consultations for upper respiratory diseases in London. *Occupational and Environmental Medicine*, **59**:294–299.

Hajat S, Haines A, Atkinson RW, Bremner SA, Anderson HR, Emberlin J (2001) Association between air pollution and daily consultations with general practitioners for allergic rhinitis in London, United Kingdom. *American Journal of Epidemiology*, **153**:704–714.

Hales S, Salmond C, Town GI, Kjellstrom T, Woodward A (2000) Daily mortality in relation to weather and air pollution in Christchurch, New Zealand. *Australia and New Zealand Journal of Public Health*, **24**: 89–91.

Health Effects Institute (2000a) *National morbidity, mortality and air pollution study.* (HEI Report No. 94, Part 1. Methods and methodologic issues) Health Effects Institute. Cambridge, MA.

Health Effects Institute (2000b) *National morbidity, mortality and air pollution study*. (HEI Report No. 94. Part 2.) Health Effects Institute. Cambridge, MA.

Health Effects Institute (2001) *Airborne particles and health: HEI epidemiologic evidence*. HEI perspectives. Health Effects Institute, Cambridge, MA.

Health Effects Institute (2002) *Understanding the health effects of components of the particulate matter mix: progress and next steps*. HEI Perspectives. Health Effects Institute, Boston, MA.

Health Effects Institute (2003) *Revised analyses of time-series studies of air pollution and health*. Health Effects Institute. Boston, MA.

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR (2002) Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. *The Lancet*, **360**: 1646–1652.

Heinrich J, Hoelscher B, Wichmann HE (2000) Decline of ambient air pollution and respiratory symptoms in children. *American Journal of Respiratory and Critical Care Medicine*, **161**:1930–1936.

Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA (2002) Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *The Lancet*, **360**:1203–1209.

Holman C (1999) Sources of air pollution. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Hong C (1995) *Global burden of disease from air pollution*. World Health Organization, Geneva. Available at http://www.cru.uea.ac.uk/~markn/cru05/cru05_intro.html.

Hong YC, Leem JH, Ha EH (1999) Air pollution and daily mortality in Inchon, Korea. *Journal of Korean Medical Science*, **14**:239–244.

Hong YC, Leem JH, Ha EH, Christiani DC (1999) PM10 exposure, gaseous pollutants, and daily mortality in Inchon, South Korea. *Environmental Health Perspectives*, **107**:873–888.

Hurley F, Holland MR, Markandya A et al. (2000) *Towards assessing and costing the health impacts of ambient particulate air pollution in the UK*. (Report TM/00/07.) Institute of Occupational Medicine, Edinburgh.

Instituto Nacional de Ecología (2000) *Almanaque de datos y tendencias de la calidad del aire en ciudades Mexicanas [Temperature data and trends in Mexican cities]*, Dirección General de Gestión e Información Ambiental del Instituto Nacional de Ecología, Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAP) [The Secretariat of Environment and Natural Resources], Mexico. Unpublished document, available at http://www.ine.gob.mx.

International Energy Agency (2001a) *Energy balance of non-OECD countries 1995–1999*. International Energy Agency, Vienna.

International Energy Agency (2001b) *Energy Balance of OECD Countries 1995–1999*. International Energy Agency, Vienna.

Jerrett M, Burnett RT, Kanaroglou P et al. (2001) A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A*, **33**:955–973.

Katsouyanni K, Schwartz J, Spix C et al. (1996) Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. *Journal of Epidemiology and Community Health*, **50**:S12–18.

Katsouyanni K, Touloumi G, Samoli E et al. (2001) Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. *Epidemiology*, **12**:521–531.

Krewski D, Burnett RT, Goldberg MS et al. (2000) *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of particulate air pollution and mortality. A special report of the institute's particle epidemiology reanalysis project*. Health Effects Institute, Cambridge, MA.

Krzyzanowski M, Schwela D (1999) Patterns of air pollution in developing countries. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Krzyzanowski M, Wojtyniak B (1991) Air pollution and daily mortality in Cracow. *Public Health Reviews*, **19**:73–81.

Künzli N, Kaiser R, Medina S et al. (2000) Public-health impact of outdoor and traffic-related air pollution: a European assessment. *The Lancet*, **356**: 795–801.

Künzli N, Medina S, Kaiser R et al. (2001) Assessment of deaths attributable to air pollution: should we use risk estimates based on time series or cohort studies? *American Journal of Epidemiology*, **153**:1050–1055.

Lee J-T, Kim H, Hong YC, Kwon HJ, Schwartz J, Christiani DC (2000) Air pollution and daily mortality in seven major cities of Korea, 1991–1997. *Environmental Research*, **84**:247–254.

Lee J-T, Shin D, Chung Y (1999) Air pollution and daily mortality in Seoul and Ulsan, Korea. *Environmental Health Perspectives*, **107**:149–154.

Leksell I, Rabl A (2001) Air pollution and mortality: quantification and valuation of years of life lost. *Risk Analysis*, **21**:843–857.

Levy JI, Hammitt JK, Spengler JD (2000) Estimating the mortality impacts of particulate matter: what can be learned from between-study variability? *Environmental Health Perspectives*, **108**:109–117.

Lipfert FW, Perry HM Jr, Miller JP, Baty JD, Wyzga RE, Carmody SE (2003) Air pollution, blood pressure, and their long-term associations with mortality. *Inhalation Toxicology*, **15**:493–512.

Loomis D, Castillejos M, Gold DR et al. (1999) Air pollution and infant mortality in Mexico City. *Epidemiology*, **10**:118–123.

McDonnell WF, Nishino-Ishikawa N, Petersen FF, Chen LH, Abbey DE (2000) Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. *Journal of Exposure Analysis and Environmental Epidemiology*, **10**:427–436.

McMichael AJ, Anderson HR, Brunekreef B, Cohen AJ (1998) Inappropriate use of daily mortality analyses to estimate longer-term mortality effects of air pollution. *International Journal of Epidemiology*, **27**:450–453.

Morgan MG, Henrion M (1998) *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, Cambridge.

National EPA of China (2000) *Environmental statistical yearbook*. China Environmental Publishing House, Beijing.

National Research Council (2002) *Estimating the public health benefits of proposed air pollution regulations*. National Academies Press, Washington, DC.

Navidi W, Lurmann F (1995) Measurement error in air pollution exposure assessment. *Journal of Exposure Analysis and Environmental Epidemiology*, **5**:111–124.

New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. Part I. Development of a 1961–90 mean monthly terrestrial climatology, *Journal of Climate*, **12**:829–856.

Nyberg F, Gustavsson P, Jarup L et al. (2000) Urban air pollution and lung cancer in Stockholm. *Epidemiology*, **11**:487–495.

O'Neill M, Jerrett M, Kawachi I et al. (2003) Health, wealth and air pollution. *Environmental Health Perspectives,* **111**:1861–1870.

Ostro B, Chestnut L (1998) Assessing the health benefits of reducing particulate matter air pollution in the United States. *Environmental Research*, **76**:94–106.

Ostro B, Chestnut L, Vichit-Vadakan N, Laixuthai A (1999a). The impact of particulate matter on daily mortality in Bangkok, Thailand. *Journal of the Air and Waste Management Association* **49**:100–107.

Ostro B, Hurley S, Lipsett MJ (1999b) Air pollution and daily mortality in the Coachella Valley, California: a study of PM10 in an area dominated by coarse particles. *Environmental Research*, **81**:231–238.

Ostro BD, Sanchez JM, Aranda C, Eskeland GS (1996) Air pollution and mortality: results from a study of Santiago, Chile. *Journal of Exposure Analysis and Environmental Epidemiology*, **6**:97–114.

Ozkaynak H (1999) Exposure assessment. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Pandey KD, Wheeler D, Ostro B, Deichmann U, Hamilton K, Bolt K (forthcoming) *Ambient particulate matter concentrations in residential areas of world cities: new estimates based on global model of ambient particulates (GMAPS)*. The Development Research Group and the Environment Department, World Bank, Washington, DC.

Pereira LA, Loomis D, Conceicao GM et al. (1998) Association between air pollution and intrauterine mortality in São Paulo, Brazil. *Environmental Health Perspectives*, **106**:325–329.

Peters A, Dockery DW, Muller JE, Mittleman MA (2001) Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*, **103**: 2810–2815.

Pope CA III (1989) Respiratory disease associated with community air pollution and a steel mill, Utah Valley. *American Journal of Public Health*, **79**:623–628.

Pope CA III, Burnett RT, Thun MJ et al. (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association*, **287**:1132–1141.

Pope CA III, Dockery DW (1999) Epidemiology of particle effects. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Pope CA III, Thun MJ, Namboodiri MM et al. (1995) Particulate air pollution as a predictor of mortality in a prospective study of US adults. *American Journal of Respiratory and Critical Care Medicine*, **151**:669–674.

Qian Z, Chapman RS, Tian Q, Chen Y, Lioy PJ, Zhang J (2000) Effects of air pollution on children's respiratory health in three Chinese cities. *Archives of Environmental Health*, **55**:126–133.

Qian Z, Zhang J, Wei F, Wilson WE, Chapman RS (2001) Long term ambient air pollution levels in four Chinese cities: inter-city and intra-city concentration gradients for epidemiological studies. *Journal of Exposure Analysis and Environmental Epidemiology*, **11**:341–351.

Reddy KS, Yusuf S (1998) Emerging epidemic of cardiovascular disease in developing countries. *Circulation*, **97**:596–601.

Roemer W, Clench-Aas J, Englert N et al. (1999) Inhomogeneity in response to air pollution in European children (PEACE project). *Occupational and Environmental Medicine*, **56**:86–92.

Romieu I, Samet JM, Smith KR, Bruce N (2002) Outdoor air pollution and acute respiratory infections among children in developing countries. *Journal of Occupational and Environmental Medicine*, **44**:640–649.

Saldiva PHN, Lichtenfels AJFC, Paiva PSO et al. (1994) Association between air pollution and mortality due to respiratory diseases in children in São Paulo, Brazil: a preliminary report. *Environmental Research*, **65**:218–225.

Saldiva PH, Pope CA, III, Schwartz J et al. (1995) Air pollution and mortality in elderly people: a time-series study in São Paulo, Brazil. *Archives of Environmental Health*, **50**:159–163.

Samet JM (1999) Air pollution and risk assessment. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Samet JM, Cohen AJ (1999) Air pollution and lung cancer. In: *Air pollution and health*. Holgate ST, Samet JM, Koren HS, Maynard R, eds. Academic Press, London.

Sanhueza P, Vargas C, Jimenez J (1999) Daily mortality in Santiago and its relationship with air pollution. *Revista Medica de Chile*, **127**:235–242.

Sarnat JA, Schwartz J, Catalano PJ, Suh HH (2001) Gaseous pollutants in particulate matter epidemiology: confounders or surrogates? *Environmental Health Perspectives*, **109**:1053–1061.

Schwartz J (2000) Harvesting and long-term exposure effects in the relation between air pollution and mortality. *American Journal of Epidemiology*, **151**:440–448.

Schwartz J, Laden F, Zanobetti A (2002) The concentration-response relation between PM(2.5) and daily deaths. *Environmental Health Perspectives*, **110**:1025–1029.

Simpson R, Denison L, Petroeschevsky A, Thalib L, Williams G (2000) Effects of ambient particle pollution on daily mortality in Melbourne, 1991–1996. *Journal of Exposure Analysis and Environmental Epidemiology*, **10**:488–496.

Smith KR, Mehta S (2003) The burden of disease from indoor air pollution in developing countries: comparison of estimates. *International Journal of Hygiene and Environmental Health*, **206**:279–89.

Sommer H, Künzli N, Seethaler R et al. (2000) *An impact assessment project of Austria, France and Switzerland. Ancillary benefits and costs of greenhouse gas mitigation* (Proceedings from an OECD Expert Workshop). OECD, New York.

Stanfield JP (1993) Some aspects of the long-term effects of malnutrition on the behaviour of children in the Third World. *Proceedings of the Nutritional Society*, **52**:201–210.

Stieb DM, Judek S, Burnett RT (2003) Meta-analysis of time-series studies of air pollution and mortality: update in relation to the use of generalized additive models. *Journal of the Air and Waste Management Association*, **53**:258–261.

Szafraniec K, Tecza W (1999) The effect of short-term changes in levels of air pollution on mortality from cardiovascular diseases among inhabitants of Krakow [in Polish]. *Przeglad Lekarski*, **56**:698–703.

Tager IB, Künzli N, Lurmann F, Ngo L, Segal M, Balmes J (1998) *Methods development for epidemiologic investigations of the health effects of prolonged ozone exposure. Part II. An approach to retrospective estimation of lifetime ozone exposure using a questionnaire and ambient monitoring data (California sites)*. (Research Report No. 81.) Health Effects Institute, Cambridge, MA.

Tellez-Rojo MM, Romieu I, Ruiz-Velasco S, Lezana MA, Hernandez-Avila MM (2000) Daily respiratory mortality and PM10 pollution in Mexico City: importance of considering place of death. *European Respiratory Journal*, **16**:391–396.

UN (2001) *Demographic yearbook 1998*. (United Nations Publication, Sales No. E/F.97.XIII.1.) Data also available at http://unstats.un.org/unsd/citydata/.

U.S. Environmental Protection Agency (1996) *Review of the national ambient air quality standards for particulate matter: policy assessment of scientific and technical information: OAQPS staff paper* (EPA-452\R-96-013.) Office of Air Quality Planning and Standards, Research Triangle Park, NC.

U.S. Environmental Protection Agency (1997) *National ambient air quality standards for particulate matter: final rule*. (CFR Part 50, Federal register 62:1–54.) U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency (2002) *Third external review draft of air quality criteria for particulate matter, April 2002*. (EPA/600/p-99/002aC.) Office of Research and Development, Washington, DC.

USGS (1996) *GTOPO30—Global digital elevation model, US geological survey*. Eros Data Center, Sioux Falls, SD. Available at http://edcdaac.usgs.gov/gtopo30/gtopo30.asp.

WHO (1997) *Health and environment in sustainable development: five years after the earth summit*. World Health Organization, Geneva.

WHO (2000a) *Air quality guidelines for Europe. Second edition*. (WHO Regional Publication, European Series, No. 91.) WHO Regional Office for Europe, Copenhagen.

WHO (2000b) *Evaluation and use of epidemiological evidence for environmental health risk assessment*. (EUR/00/5020369.) WHO Regional Office for Europe, Copenhagen.

WHO (2001a) *Health impact assessment of air pollution in the WHO European Region*. (Technical Report from WHO/ECEH Project.) WHO European Centre for Environment and Health, Bonn.

WHO (2001b) *Quantification of health effects of exposure to air pollution*. WHO Regional Office for Europe. Available at http://www.euro.who.int/document/e74256.pdf.

WHO (2001c) *Air quality and health*. Air Management Information System AMIS3.0. World Health Organization, Geneva.

WHO (2002) *World health report 2002*. Reducing risks, promoting healthy life. World Health Organization, Geneva.

WHO (2003) *Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide*. WHO Regional Office for Europe, Copenhagen. Available at http://www.euro.who.int/document/e79097.pdf.

Wilhelm M, Ritz B (2003) Residential proximity to traffic and adverse birth outcomes in Los Angeles county, California, 1994–1996. *Environmental Health Perspectives*, **111**:207–216.

Wojtyniak B, Piekarska T (1996) Short term effect of air pollution on mortality in Polish urban populations—what is different? *Journal of Epidemiology and Community Health*, **50**:S36–41.

Wong CM, Atkinson RW, Anderson HR et al. (2002) A tale of two cities: effects of air pollution on hospital admissions in Hong Kong and London compared. *Environmental Health Perspectives*, **110**:67–77.

Wong CM, Ma S, Hedley AJ, Lam TH (2001) Effect of air pollution on daily mortality in Hong Kong. *Environmental Health Perspectives*, **109**:335–340.

Woodruff TJ, Grillo J, Schoendorf KC (1997) The relationship between selected causes of post neonatal mortality and particulate air pollution in the United States. *Environmental Health Perspectives*, **105**:608–612.

Working Group on Public Health and Fossil Fuel Combustion (1997) Short-term improvements in public health from global-climate policies on fossil fuel combustion: an interim report. *The Lancet*, **350**:1341–1349.

World Bank (2002) *World development indicators 2002*. World Bank, Washington.

Xu X, Gao J, Dockery DW, Chen Y (1994) Air pollution and daily mortality in residential areas of Beijing, China. *Archives of Environmental Health*, **49**:216–222.

Xu ZY, Yu DG, Jing LB, Xu XP (2000) Air pollution and daily mortality in Shenyang, China. *Archives of Environmental Health*, **55**:115–120.

Zeger SL, Dominici F, Samet J (1999) Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology*, **10**:171–175.

Zeger SL, Thomas D, Dominici F et al. (2000) Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental Health Perspectives*, **108**:419–426.

Zhang J, Qian Z, Kong L, Zhou L, Yan L, Chapman RS (1999) Effects of air pollution on respiratory health of adults in three Chinese cities. *Archives of Environmental Health*, **54**:373–381.

Zhang JJ, Hu W, Wei F, Wu G, Korn LR, Chapman RS (2002) Children's respiratory morbidity prevalence in relation to air pollution in four Chinese cities. *Environmental Health Perspectives*, **110**:961–967.

Chapter 18

Indoor air pollution from household use of solid fuels

Kirk R. Smith, Sumi Mehta and
Mirjam Maeusezahl-Feuz

Summary

This chapter summarizes the methodology used to assess the burden of disease caused by indoor air pollution from household use of solid fuels. Most research into and control of indoor air pollution worldwide has focused on sources of particular concern in developed countries, such as environmental tobacco smoke (ETS), volatile organic compounds from furnishings and radon from soil. Although these pollutants have impacts on health, little is known about their global distribution. Thus, we focus solely on indoor smoke from household use of solid fuels, the most widespread traditional source of indoor air pollution on a global scale.

In order to be consistent with the epidemiological literature, binary classifications of household use of solid fuels (biomass and coal) were used as a practical surrogate for actual exposure to indoor air pollution. Specifically, household solid fuel use was estimated at the national level using binary classifications of exposure to household fuel use, i.e. solid fuel and non-solid fuel (gas, kerosene, electricity). We estimated exposure to smoke from solid fuel by combining a number of national surveys of household fuel use into a regression model that predicts use according to independent, development-related variables, such as income and urbanization. Although this method was necessary owing to the current paucity of quantitative data on exposure, we acknowledge that it overlooks the large variability of exposure within households using solid fuels. As pollution emissions from the use of solid fuel may not always indicate high exposures, we have adjusted exposure estimates by a second term, the ventilation factor, which is based on qualitative measures of ventilation.

Estimates of relative risk obtained from epidemiological studies were combined in meta-analyses for three disease end-points for which there is strong evidence of an association with use of solid fuels: acute lower respiratory infections (ALRI) in children aged <5 years, chronic

obstructive pulmonary disease (COPD) and lung cancer (estimates for lung cancer are only for use of coal).

More than 1.6 million deaths and over 38.5 million disability-adjusted life years (DALYs) were attributable to indoor smoke from solid fuels in 2000. Cooking with solid fuels is thus responsible for a significant proportion, about 3%, of the global burden of disease. Although trends are highly uncertain, attributable risks are likely to be greater than avoidable risks.

Several potentially important health outcomes, including tuberculosis, cardiovascular disease, and adverse pregnancy outcomes, were not included, owing to insufficient epidemiological evidence. In addition, there was insufficient evidence to assess the associated health effects for children aged 5–14 years. The burden of disease caused by use of solid fuel is thus likely to be underestimated.

1. Introduction

The use of solid fuels for cooking and heating is likely to be the largest source of indoor air pollution on a global scale. Nearly half the world continues to cook with solid fuels such as dung, wood, agricultural residues and coal. When used in simple cooking stoves, these fuels emit substantial amounts of toxic pollutants. These pollutants, called solid-fuel "smoke" in this chapter, include respirable particles, carbon monoxide, oxides of nitrogen and sulfur, benzene, formaldehyde, 1,3-butadiene, and polyaromatic compounds, such as benzo(α)pyrene (Smith 1987). In households with limited ventilation (as is common in many developing countries), exposures experienced by household members, particularly women and young children who spend a large proportion of their time indoors, have been measured to be many times higher than World Health Organization (WHO) guidelines and national standards (Bruce et al. 2000; Smith 1987).

Most research into and control of indoor air pollution worldwide has focused on sources of particular concern in developed countries, such as ETS, volatile organic compounds from furnishings and radon from soil (Table 18.1) (Spengler et al. 2001). Although these pollutants have impacts upon health, little is known about their global distribution.

In an initial attempt to estimate the burden of disease and death caused by indoor sources of air pollution, this chapter focuses solely on the burning of solid fuels. Studies of the health effects of exposure to emissions from the two major sources of energy used for cooking in developed countries, gas and electricity, have been inconsistent, although small but statistically significant increased risks of childhood respiratory disease and other effects associated with use of gas have emerged from meta-analyses (Basu and Samet 1999). This is in contrast to the growing quantity of literature reporting reasonably consistent and strong relationships for a number of health end-points in households burning solid

Table 18.1 Major toxic pollutants of indoor air

Pollutant	*Major indoor sources*
Fine particles	Fuel/tobacco combustion, cleaning, fumes from food being cooked, e.g. from cooking oil
Carbon monoxide	Fuel/tobacco combustion
Polycyclic aromatic hydrocarbons	Fuel/tobacco combustion, fumes from food being cooked, e.g. from cooking oil
Nitrogen oxides	Fuel combustion
Sulfur oxides	Coal combustion
Arsenic and fluorine	Coal combustion
Volatile and semi-volatile organic compounds	Fuel/tobacco combustion, consumer products, furnishings, construction materials, fumes from food being cooked, e.g. from cooking oil
Aldehydes	Furnishing, construction materials, cooking
Pesticides	Consumer products, dust from outside
Asbestos	Remodelling/demolition of construction materials
Lead[a]	Remodelling/demolition of painted surfaces
Biological pollutants	Moist areas, ventilation systems, furnishings
Free radicals and other short-lived, highly reactive compounds	Indoor chemistry
Radon	Soil under building, construction materials

[a] Lead-containing dust from deteriorating paint is an important indoor pollutant for occupants of many households, but the most critical exposure pathways are not usually through air. See chapter 19.

Source: Zhang and Smith (2003).

fuels (biomass or coal), particularly those with poorly-vented stoves and homes, which are common throughout developing countries. In many circumstances, it is difficult to distinguish use of solid fuels for cooking from use for heating the home. There may also be effects associated with the use of kerosene, a common cooking fuel in many parts of the world, for which emissions and exposures are intermediate between those for solid and for gaseous fuels (Smith 1987), but on which few studies of health effects seem to have been conducted.

2. Estimating risk factor levels

2.1 Exposure variables

One way to determine the health effects of indoor smoke from solid fuels would be to apply the well-established exposure–response relationships from epidemiological studies of outdoor, or ambient, concentrations of

the same pollutants (see chapter 17) to the household exposures, called here the "pollutant-based approach" (Smith and Mehta 2003).

There are a number of potential problems with such an approach, however, including:

- *Differences in pollutant mixtures:* Although particles are often used as an indicator pollutant, the composition of particles (size, chemical composition, etc.) as well as that of other pollutants varies from source to source, and also changes with dispersion (Rossi et al. 1999).
- *Differences in exposure patterns:* The daily pattern of indoor air pollution sources varies from that of ambient sources, with large peaks corresponding to cooking and heating schedules (Naeher et al. 2000b).
- *Differences in exposure levels*: Concentrations of particulates from the indoor combustion of biomass have been measured at levels that are 10–50 times greater than in urban areas of developed countries, where the main epidemiology of pollutants has been performed. Extrapolating exposure–response relationships by such a large factor is problematic, particularly as there are indications that the relationship becomes more shallow at higher exposures (Bruce et al. 2000).
- *Relevance of health outcomes addressed:* Most studies of outdoor air pollution have attempted to associate short-term changes in exposure with acute health outcomes. This does not address the long-term impact on chronic health outcomes, nor does it necessarily focus on the health outcomes that are responsible for the bulk of the burden of disease. In particular, ALRI, mostly in the form of pneumonia, are likely to be responsible for the largest burden of disease caused by exposure to indoor air pollution.
- Data on concentrations of particulate matter (PM) in indoor air[1] are sparse. In addition, most measurements have been made for concentrations of total particulates, which are less reliable indicators of risk than smaller particles (PM_{10} or $PM_{2.5}$).

An alternative approach, consistent with that used in most epidemiological studies in developing countries, is to divide the population into categories of people that are exposed or not exposed to smoke from solid fuel, on the basis of fuel use and ventilation. Although necessary here, owing to the current lack of exposure data, this method overlooks the large variability of exposure within each of these groups (Naeher et al. 2000a). Furthermore, the method based on use of fuel is affected by the first of the shortcomings listed above, as the same broad category of fuels may produce different mixtures of pollutants in different settings. We also recognize that exposures from cooking and heating

can differ considerably because of different conversion technologies. It was not possible to distinguish between the two end-uses in most cases, however.

To account for differences in other factors (e.g. housing) that would affect levels of pollution (Mehta and Smith 2002), we included a second component in the exposure variable, which we refer to as the "ventilation factor". The final exposure variable in the population was defined as:

$$\textit{Household-equivalent solid-fuel exposed population} = (\textit{Population using solid fuel}) \times (\textit{Ventilation factor})$$

We compiled a database of household use of solid fuel, from which the prevalence of household use of solid fuel was estimated for each subregion.[2] Using known values from this database, a statistical model was developed to predict national use of solid fuel for countries without data. Ventilation factors were assigned on the basis of qualitative evidence, to account for differences in types of cooking and heating appliances and housing.

2.2 Theoretical-minimum-risk exposure distribution

The theoretical minimum for this risk factor is clearly no use of solid fuels for the production of household energy; this has already been achieved in many populations. In reality, of course, there would still be exposure to pollution from liquid and gaseous fuels, which might be further reduced through a switch to use of electricity or of very well-ventilated cooking conditions.

2.3 A database of household use of solid fuel

A database of households using solid fuel, expressed as a percentage of all households, was compiled for 52 countries in 10 subregions, in order to estimate global household use of solid fuel (see Table 18.2). Although the data were acquired from studies conducted at different times in the past decade, fuel-use patterns are unlikely to have changed drastically within this time frame (International Energy Agency 2002; World Resources Institute 2000). Out of necessity, the data were gathered from various sources using different and, at times, non-validated methodology. We thus had to make many assumptions in order to facilitate subregional comparison and data manipulation associated with solid fuel use. No households were reported to be using solid fuels for cooking in AMR-A, EUR-A, EUR-C and WPR-A, presumably because countries in these subregions have already shifted to cleaner fuels.

In many countries where large proportions of the population cook with solid fuels, data on household energy are widely, although not

Table 18.2 Estimates of data for the database of households using solid fuels

Subregion	*Country*	*Households using solid fuel (%)*	*Type of data source*	*Year*	*Reference*
AFR-D	Algeria	4	National energy statistics	1999	World Resources Institute (2003)
	Angola	100	National energy statistics	1999	International Energy Agency (1999)
	Burkina Faso	97	Household survey	1994/1995	World Bank (2000)
	Chad	95	Household survey	1991	World Bank (1998)
	Gambia	98	Household survey	1992	World Bank (2000)
	Ghana	95	Household survey	1997	World Bank (2000)
	Guinea	99	Household survey	1994/1995	World Bank (2000)
	Guinea-Bissau	95	Household survey	1992	World Bank (2000)
	Madagascar	99	Household survey	1993/1994	World Bank (2000)
	Mali	100	Household survey	1994	World Bank (2000)
	Mauritania	69	Household survey	1995	World Bank (2000)
	Niger	98	Household survey	1995	World Bank (2000)
	Nigeria	67	Household survey and census data	1992	World Bank (2000)
	Senegal	79	Household survey	1994/1995	World Bank (2000)
	Sierra Leone	92	Household survey	1989/1990	World Bank (2000)
AFR-E	Botswana	65	National census	1991	Government of Botswana (1991)
	Central African Republic	99	Household survey	1993	World Bank (2000)
	Congo	67	Household survey	1988	World Bank (1988)
	Côte d'Ivoire	93	Household survey	1995	World Bank (2000)
	Democratic Republic of the Congo	100	National energy statistics	1999	World Resources Institute (2003)
	Ethiopia and Eritrea	97	Household survey and census data	1994	Government of Ethiopia (1998)
	Kenya	85	Household survey	1994	World Bank (2000)
	South Africa	28	Household survey	1993	World Bank (2000)
	Swaziland	88	Household survey	1994	World Bank (2000)
	United Republic of Tanzania	96	Household survey	1993	World Bank (2000)
	Uganda	97	Household survey	1992/1993	World Bank (2000)
	Zambia	87	Household survey	1996	World Bank (2000)
	Zimbabwe	67	National census	1992	Government of Zimbabwe (1992)

AMR-A	—	—	—	—	—
AMR-B	Brazil	27	National census	1991	Government of Brazil (1991)
	Mexico	22	National census	1990	Government of Mexico (1990a)
AMR-D	Ecuador	28	National census	1990	Government of Ecuador (1990a)
EMR-B	Iran (Islamic Republic of)	2	National energy statistics	1999	World Resources Institute (2003)
	Lebanon	0	National energy statistics	1996/1997	World Resources Institute (2003)
	Libyan Arab Jamahiriya	3	National energy statistics	1996/1997	International Energy Agency (1999)
	Tunisia	29	National energy statistics	1999	World Resources Institute (2003)
EMR-D	Afghanistan	98	National energy statistics	1999	World Resources Institute (2003)
	Djibouti	6	Household survey	1996	World Bank (2000)
	Egypt	23	Household survey	1993	World Energy Council (1999)
	Iraq	2	National energy statistics	1999	World Resources Institute (2003)
	Morocco	11	National energy statistics	1999	World Resources Institute (2003)
	Pakistan	76	National energy statistics	1997	Government of Pakistan (1997)
	Sudan	100	National energy statistics	1999	International Energy Agency (1999)
EUR-A	—	—	—	—	—
EUR-B	Turkey	11	National energy statistics	1999	World Resources Institute (2003)
EUR-C	—	—	—	—	—
SEAR-B	Indonesia	63	Personal communication	1995/1996	Government of Indonesia (1996b)
	Thailand	72	National energy statistics	1997	FAO (1997a)
SEAR-D	Bangladesh	96	National energy statistics	1997	FAO (1997a)
	India	81	National census	1991	Government of India (1991a)
	Myanmar	100	National energy statistics	1997	FAO (1997a)
	Nepal	97	National energy statistics	1997	FAO (1997a)
WPR-A	—	—	—	—	—
WPR-B	China	80	National energy statistics	1996	Government of China (1996)
	Philippines	85	National energy statistics	1997	FAO (1997a)
	Viet Nam	98	National energy statistics	1997	FAO (1997a)

— No data.

universally, available. In some cases, the data come directly from national census information or energy use statistics, which state explicitly the number or fraction of households that rely predominantly on solid fuels for their energy needs (Government of Botswana 1991; Government of Brazil 1991; Government of Ecuador 1990b; Government of Ethiopia 1998; Government of India 1991b; Government of Mexico 1990b; Government of Nigeria 1990; Government of Zimbabwe 1992). For example, information on the main fuel used for cooking is collected during the house listing of the census of India each decade (Government of India 1991b). These data, disaggregated into urban and rural sectors, are available at the district level (in India, a district contains about 2 million people).

In some countries, where censuses are infrequent and/or data on residential energy use are not collected, household surveys are an important source of information. Some of these household surveys, such as the widely conducted Demographic Health Surveys are repeated, while others may be conducted only once. For example, primary household energy estimates for 22 countries in Africa, based on household surveys with sample sizes ranging from 1000 to >14 000 households, are included in a database of development indicators for Africa, compiled by the World Bank (2000). In China, data are available in the form of aggregate annual residential fuel consumption at the provincial level, disaggregated by urban and rural areas (Government of China 1996). Cooking and heating energies were distinguished using a simple model that accounted for the average number of "heating days" in each province, based on a 30-year average from 1951–1980 (Lin 1995). A small amount of energy (2 kg-coal equivalent per household per heating day) was considered to be heating fuel and subtracted from the mix of solid fuels in each province. The remaining heating-adjusted cooking fuel was then normalized to "useful energy" using typical conversion efficiencies for each fuel–stove combination reported (Zhang et al. 2000). The proportion of useful cooking energy attributed to each fuel type per household in each province was taken to represent the number of households using that fuel. This analysis was repeated for each of the provinces in China[3] and aggregated to give a national total. It was estimated that in 1996 nearly 80% of the households in China used solid fuels.

Many countries produce national estimates of solid-fuel use, but only a minority collect specific information on fuel use at the household level. Evidence from 10 countries (Bangladesh, Ecuador, Indonesia, Mexico, Myanmar, Nepal, Pakistan, the Philippines, Thailand and Viet Nam) indicates that national and household levels of solid-fuel use are highly correlated ($R^2=0.75$). It should be noted, however, that this relationship holds true when solid fuels are not heavily used in industry. This correlation was used to estimate use of solid fuel by households in nine coun-

tries (Afghanistan, Algeria, Egypt, the Islamic Republic of Iran, Lebanon, the Libyan Arab Jamahiriya, Morocco, Tunisia and Turkey) where only information on national use of solid fuel was available. For three countries (Angola, the Democratic Republic of the Congo and the Sudan), in which a large fraction of the total national energy consumed (>70%) comprised biomass fuels (World Resources Institute 2003), household use of solid fuel was assumed to be 100%. In other countries, including Bangladesh, Indonesia, Myanmar, Nepal, the Philippines, Thailand, Viet Nam and Pakistan (FAO 1997a, 1997b; Government of Indonesia 1995, 1996a; Government of Pakistan 1997), aggregate data on annual residential fuel consumption are available. In these cases, the percentage of households using solid fuels was estimated according to the quantity of fuel consumed.

The fraction of the population of each subregion covered by the countries for which some data were available, and the prevalence of solid-fuel use according to these data are given in Table 18.3. Data on specific types of solid fuel (i.e. use of coal vs biomass) are limited to India and China, but this factor is also likely to be important in other countries in which no estimates were made, including South Africa and Pakistan.

Table 18.3 Estimates of the prevalence of households using solid fuel, by subregion, using the household fuels database

Subregion	*Population covered by available data (000s)*	*Population covered by available data (% of total population of subregion)*	*Households using solid fuel in population covered (%)*
AFR-D	260515	88.8	72.5
AFR-E	284784	84.4	84.5
AMR-A	—	—	—
AMR-B	268997	62.5	24.9
AMR-D	12646	17.7	28.1
EMR-B	86174	61.8	5.6
EMR-D	260797	73.0	66.8
EUR-A	—	—	—
EUR-B	66591	30.7	10.8
EUR-C	—	—	—
SEAR-B	273507	93.6	64.9
SEAR-D	1212359	97.9	83.8
WPR-A	—	—	—
WPR-B	1433356	93.8	81.1

— No data.

2.4 A model to predict national use of solid fuel

Using known values from the database of households using solid fuel, a statistical model was built to predict national use of solid fuel according to a number of development parameters. The model was then applied to countries where no data on household fuel use existed. This method also allowed for the estimation of statistical uncertainty (i.e. excluding uncertainty in available data and the validity of model) surrounding each prediction.[4]

As a country develops, households gradually switch from using solid fuels to using cleaner liquid and/or gaseous fuels. Although the picture is often more complex at local and household levels, it is assumed here that this generally holds true over the long term on a subregional scale, a trend well-established by current, albeit cross-sectional, international comparisons. After a certain level of economic growth has been achieved, it is assumed that countries will shift away from cooking entirely with solid fuels. The use of solid fuel for heating may continue, however, especially in areas that are rich in coal and wood.

For countries for which data were not available, a model based on the parameters described in Table 18.4 was used with stepwise linear regression. With a gross national product (GNP) of US$4420 per capita

Table 18.4 Parameters in the fuel use prediction model[a]

Indicator	*Source*
Solid-fuel use (dependent variable)	Table 18.3
Adult female illiteracy, 1998	World Bank (2001)
Average annual growth rate, 1998–1999	World Bank (2001)
Dummy variables for all subregions	NA
Electricity consumption, per capita, 1997 (kilowatt-hours)	World Bank (2001)
Fuel-wood production	UN (1993)
Population in 2000	UN (1998)
Fuel-wood production per capita (kg)	Author calculation
Gini coefficient	World Bank (2001)
GNP per capita, 1999	World Bank (2001)
ln (GNP per capita, 1999)	Author calculation
Petroleum use per capita	UN (1993)
ln (petroleum use per capita)	Author calculation
Rural population, 1999	World Bank (2001)
Traditional fuel use (national), 1993	UN (1993)

NA Not applicable.

[a] Variables already entered were tested for removal at each step, so that variables in the model that became insignificant with inclusion of additional variables were removed. Missing values were replaced with mean values for each variable.

Table 18.5 Models to predict fuel use: GNP per capita vs use of traditional fuel as a predictor variable

Model	*Predictors*[a]	*R*	R^2	*Adjusted* R^2
1	Use of traditional fuel, EMR,[b] petroleum use per capita, rural population, constant	0.869	0.756	0.735
2	GNP per capita, EMR, petroleum use per capita, rural population, constant	0.864	0.746	0.724

[a] Dependent variable in both models is the percentage of households using solid fuels.

[b] Each subregional dummy variable was entered separately into the model. EMR was the only subregional dummy variable that was significant in the final model, perhaps because of a combination of low biomass resources and high access to petroleum fuels in some countries in these subregions.

in 1999, Brazil was the richest country in the database to have significant levels of cooking with solid fuels (27% of households). To avoid extrapolating the model to areas where it may be inappropriate, estimates were made only for countries with a GNP of <US$ 5000 per capita in 1999. All countries with a GNP of >US$ 5000 per capita in 1999 were assumed to have made a complete transition to clean household-cooking systems, either with cleaner liquid or gaseous fuels, or electricity or, where solid fuel was still used for cooking or heating, to fully ventilated appliances.

As use of traditional fuel (as a percentage of national energy use) is highly correlated with GNP per capita, stepwise linear regression eliminates GNP per capita when both variables are entered together. If use of traditional fuel is not entered, it is essentially replaced by GNP per capita in the model, with little impact on model fit or standard error (Table 18.5). Two models to predict fuel use were assessed, one employing GNP per capita and the other use of traditional fuel (as a percentage of national energy use) as predictor variables. Use of traditional fuel, which includes use of fuel-wood, bagasse (biomass remaining after processing sugar-cane), charcoal, animal wastes, agricultural residues, and other vegetable biomass wastes, is expressed as a percentage of total fuel use at the national (as opposed to the household) level, on an energy-equivalent basis. Like household use of solid fuel, use of traditional fuel at the national level is highly correlated with GNP per capita (Figures 18.1 and 18.2).

Information on GNP per capita is more reliable, is updated more routinely, and is available at the national level for nearly all countries. Therefore, we used the model including GNP per capita as a predictor, rather than the model using use of traditional fuel. The final model is shown in Table 18.6 and includes percentage of the rural population, GNP per capita (log-transformed), petroleum use per capita, and location within the EMR subregions (entered as a dummy variable). Other

Figure 18.1 The relationship between use of traditional fuel at the national level (as a fraction of national energy use) and GNP per capita

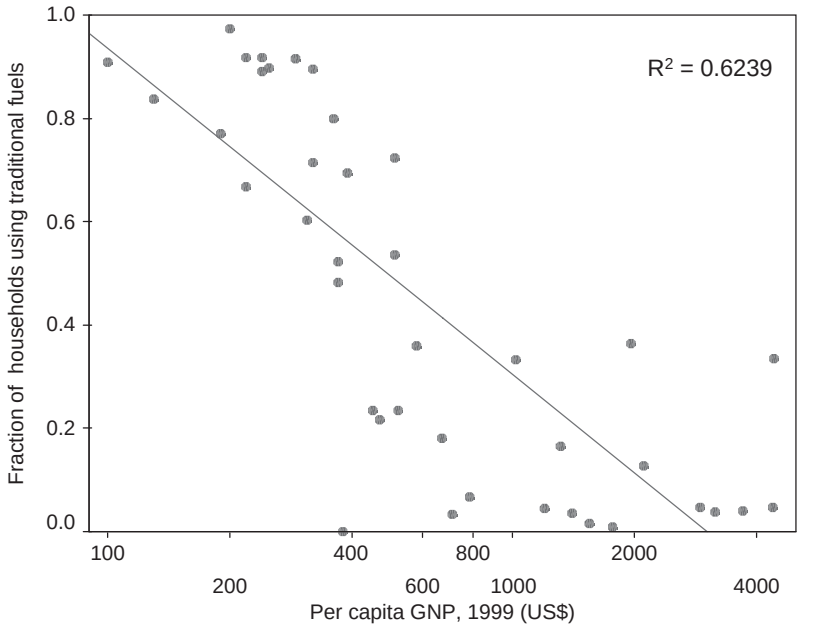
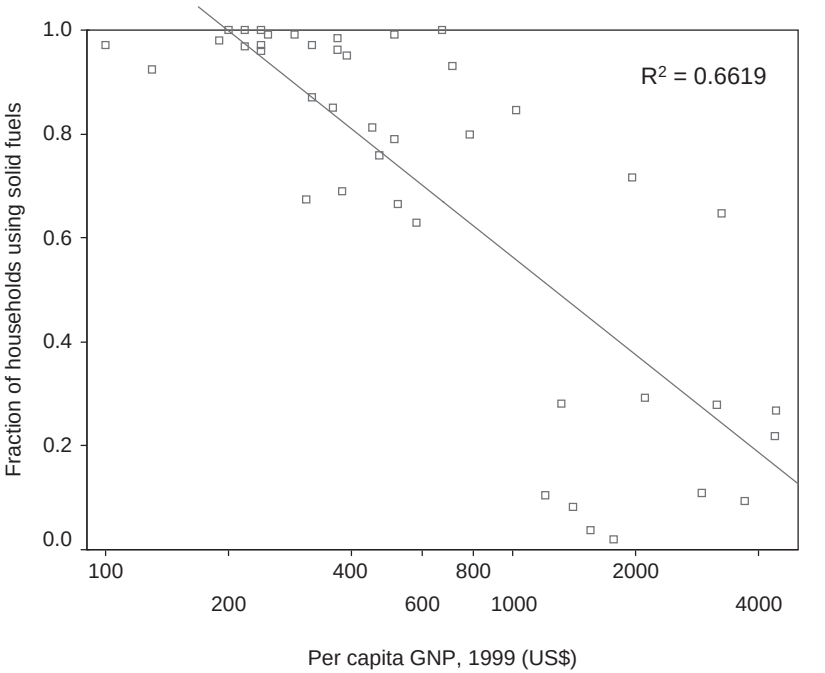


Table 18.6 Final model used to predict household use of solid fuel at the national level[a]

	Unstandardized coefficients	*Standardized coefficients*			
	Beta	*Standard error*	*Beta*	*t*	p
(Constant)	1.12	0.350	NA (0.414–1.82)	3.19	0.0025
Rural	0.661	0.214	0.353 (0.231–1.09)	3.09	0.0033
EMR	–0.248	0.0709	–0.284 (–0.390––0.105)	–3.50	0.0010
GNP(log transformed)	–0.104	0.0405	–0.265 (–0.185––0.0224)	–2.56	0.0136
Per capita petroleum use	–0.0003	0.0001	–0.224 (–0.0006––0.0001)	–2.55	0.0143

[a] Dependent variable is the percentage of households using solid fuels.

Figure 18.2 Relationship between use of traditional fuel at the household level and GNP per capita



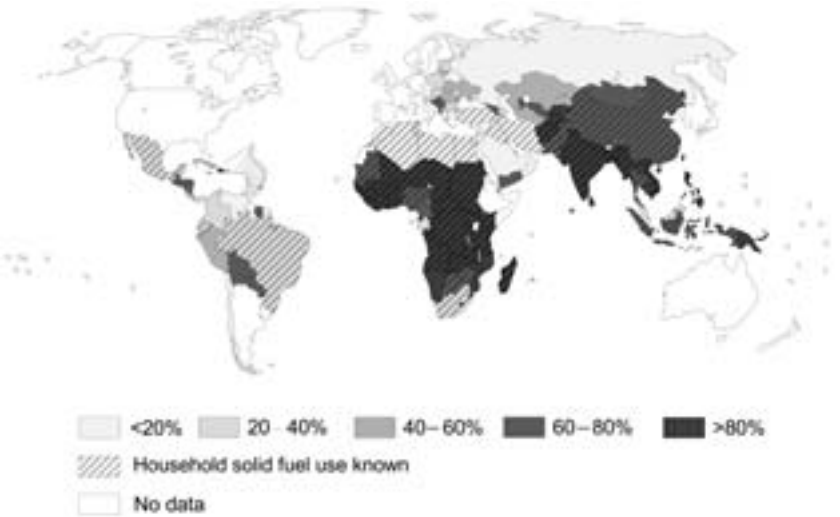
potential variables were dropped from the model in stepwise linear regression.

This model was used to predict percentages of households using solid fuel in all countries where these values were unknown (see Figure 18.3). In order to force the percentage of households using solid fuel to lie between 0% and 100%, estimates for the 23 countries with predicted values of <0 or >100 were converted to 0 and 100, respectively.

Known (for all countries in the household fuel-use database) and predicted estimates of use of solid fuel at the country level were aggregated into subregional estimates of household solid-fuel use (Table 18.7). The subregions with the least coverage are those that have the highest levels of economic development, i.e. those subregions that are least likely to have high proportions of household solid-fuel use because people have, for the most part, already shifted to cleaner fuels and cooking technologies.

We assumed that the fraction of the population exposed is the same as the fraction of households using solid fuel. This assumption is likely

Figure 18.3 Household use of solid fuel, at the national level, 2000



Note: Household solid fuel use estimates are predictions in areas without striations.

Table 18.7 Estimated household use of solid fuel, by subregion

Subregion	*Subregional population (000s)*	*Total population covered by fuel use prediction and by available data*		*Household use of solid-fuel (% of population)*
		n (000s)	*%*	*Point estimate*
AFR-D	293 440	292 317	99.6	73.4 (68.1–77.7)
AFR-E	337 547	333 697	98.9	85.8 (80.5–89.2)
AMR-A	320 704	11 201	3.5	1.5 (0.9–2.0)
AMR-B	430 674	388 897	90.3	24.6 (18.8–30.8)
AMR-D	71 318	71 318	100.0	52.9 (42.6–63.2)
EMR-B	139 532	145 137	100.0	6.1 (2.0–12.1)
EMR-D	357 476	278 909	78.0	55.2 (49.8–60.1)
EUR-A	410 714	10 689	2.6	0.2 (0.0–0.5)
EUR-B	216 930	216 930	100.0	41.5 (32.0–50.7)
EUR-C	245 688	245 688	100.0	22.8 (13.9–41.0)
SEAR-B	292 334	292 334	100.0	66.5 (61.1–71.8)
SEAR-D	1 238 808	1 236 398	99.8	83.5 (78.3–88.3)
WPR-A	153 357	328	0.2	0.2 (0.1–0.2)
WPR-B	1 528 144	1 479 669	96.8	78.1 (73.0–82.8)
World	6 036 664	5 003 510	82.9	56.5 (51.7–61.5)

to underestimate exposure since solid-fuel-using households are more likely to be rural and of low socioeconomic status, and are thus likely to have more members than the subregional average.

2.5 Assigning ventilation factors

Since people in different parts of the world use different cooking and heating appliances and have different types of housing, ventilation must also be taken into account when estimating exposure. Here, the term "ventilation" encompasses both ventilation-related characteristics of the stove (such as the presence of a chimney that vents to the outside of the house) and characteristics of the kitchen (building material, architectural features that influence indoor air quality, location of the kitchen with relation to living area, etc.).

Although we had no data on ventilation conditions according to subregion, we hypothesized that ventilation was a function of climate and development (UNCHS 1996). As described above, countries with a GNP per capita of >US$ 5000 were essentially assigned an estimated exposure of 0, that is, any use of solid fuel in the household was assumed to be undertaken in fully-vented appliances, with no re-entry of the pollution into the household. In the absence of further information (as described below), all other countries were assigned a ventilation factor of 1.0.

In countries of eastern Europe and the former Soviet Union, a long history of household use of solid fuel under cold climatic conditions and relatively high standards of living, before the recent economic decline, led to the development of energy technologies with far fewer indoor emissions and, consequently, less exposure per unit of solid fuel burned. Therefore, we set the ventilation factor at 0.2 for EUR-B and EUR-C.

In China, the widespread national improved-stove programme has disseminated cooking stoves with chimneys to three-quarters of rural households using solid fuel since 1981 (Goldemberg et al. 2000; Smith et al. 1993), resulting in decreased effective exposure. The ventilation factor for China was set at 0.25 for child health outcomes, because even well-operating, improved biomass stoves with chimneys are still responsible for some exposure (Sinton et al. 1995). We set China's ventilation factor at 0.5 for adult health outcomes, as current disease patterns for adults partly reflect exposure before the introduction of improved stoves. India, the only other country with a long-term national stove-improvement programme, has had only mixed success, with relatively low stove lifetimes and national coverage (NCAER 2002). The ventilation factor was therefore maintained at 1.0 for India.

Tables 18.8 and 18.9 detail estimated exposures as defined above for children aged <5 years and for adults. Separate estimates of exposure resulting from use of coal are presented in Table 18.10 for adults only, as adults are affected by chronic health outcomes (see section 3).

Table 18.8 Exposure of children (aged <5 years) to indoor smoke from solid fuels

Subregion	*Household solid-fuel use (%)*	*Ventilation factor*	*Exposure (% population) Point estimate (95% CI)*
AFR-D	73.4	1.00	73.4 (68.1–77.7)
AFR-E	85.8	1.00	85.8 (80.5–89.2)
AMR-A	1.5	1.00	1.5 (0.9–2.0)
AMR-B	24.6	1.00	24.6 (18.8–30.8)
AMR-D	52.9	1.00	52.9 (42.6–63.2)
EMR-B	6.1	1.00	6.1 (2.0–12.1)
EMR-D	55.2	1.00	55.2 (49.8–60.1)
EUR-A	0.2	0.97	0.0 (0.0–0.5)
EUR-B	41.5	0.65	26.0 (20.6–31.1)
EUR-C	22.8	0.25	7.2 (5.0–11.3)
SEAR-B	66.5	1.00	66.5 (61.1–71.8)
SEAR-D	83.5	1.00	83.5 (78.3–88.3)
WPR-A	0.2	1.00	0.2 (0.1–0.2)
WPR-B	78.1	0.37	28.0 (26.1–29.6)

Table 18.9 Exposure of adults (aged ≥15 years) to indoor smoke from solid fuels

Subregion	*Household solid-fuel use (%)*	*Ventilation factor*	*Exposure (%) Point estimate (95% CI)*
AFR-D	73.4	1.00	73.4 (68.1–77.7)
AFR-E	85.8	1.00	85.8 (80.5–89.2)
AMR-A	1.5	1.00	1.5 (0.9–2.0)
AMR-B	24.6	1.00	24.6 (18.8–30.8)
AMR-D	52.9	1.00	52.9 (42.6–63.2)
EMR-B	6.1	1.00	6.1 (2.0–12.1)
EMR-D	55.2	1.00	41.4 (37.4–45.1)
EUR-A	0.2	0.97	0.0 (0.0–0.5)
EUR-B	41.5	0.65	26.0 (20.6–31.1)
EUR-C	22.8	0.25	7.2 (5.0–11.3)
SEAR-B	66.5	1.00	66.5 (61.1–71.8)
SEAR-D	83.5	1.00	83.5 (78.3–88.3)
WPR-A	0.2	1.00	0.2 (0.1–0.2)
WPR-B	78.1	0.58	44.7 (41.7–47.4)

Table 18.10 Exposure of adults (aged ≥15 years) to coal smoke[a]

Subregion	*Exposure (%) Point estimate (95% CI)*
SEAR-D	2.1 (0.0–7.1)
WPR-B	12.9 (7.9–17.9)

[a] Assumed to be zero in all other subregions owing to lack of disaggregated data.

2.6 Quantitative and qualitative sources of uncertainty

Estimates of use of solid fuel for countries in the household fuel-use database were arbitrarily assigned an uncertainty range of 5%. The exposure classification system used here is binary (exposed to solid fuels or not exposed), which is consistent with the available epidemiological literature. In reality, exposure to indoor air pollution from use of solid fuel results in a wide range of exposures, which vary according to different types and quality of fuel and stove housing characteristics (e.g. ventilation and size), cooking and heating methods, differences in time-activity patterns (time spent within the household and in close proximity to the pollution source) and season (Saksena et al. 1992). Since the distribution of exposures is continuous, exposures would best be characterized as a continuous outcome, or at least better characterized by multiple categories. As a result, the above binary categorization and uncertainty values significantly underestimate the true uncertainty in levels of exposure. In addition, the need to use the fuel-prediction model for countries without data obviously introduces uncertainty, only part of which may be reflected in the variance of the results obtained from the model.

3. Estimating risk factor–disease relationships

3.1 Health outcomes: evidence for causality and inclusion criteria

Health outcomes caused by indoor exposure to smoke from use of solid fuel were chosen after a review of the epidemiological evidence available for each end-point, using electronic databases, including Medline and TCMLARS (Traditional Chinese Medical Literature Analysis and Retrieval System, an electronic database of Chinese journals). In addition, given that a large body of evidence comes from developing countries, literature was also obtained from other researchers and reputable developing-country journals not currently indexed in international databases. Only articles written or abstracted in English were used, except for articles on lung cancer, for which both the Chinese and the English

Table 18.11 Diseases associated with use of solid fuels and populations affected that were included in the analysis

Disease	*Population affected*
Acute lower respiratory infections (ALRI)	Children aged <5 years
Chronic obstructive pulmonary disease (COPD)	Females and males aged ≥30 years
Lung cancer (coal use only)	Females and males aged ≥30 years

literature were accessed, since, to our knowledge, only in China has there been significant use of coal in unvented household devices in recent decades.

General assessment of causality

The strength of the evidence for each end-point was determined on the basis of a structured assessment of causality, using Bradford Hill's criteria for causality, including temporal relationship, strength of association, specificity, the presence of a dose–response relationship, biological plausibility, coherence, the existence of experimental evidence and consistency of association.

As specificity, dose–response relationships, and experimental evidence are often difficult to assess for environmental exposures and health outcomes with multiple causes or long latency periods, we used the epidemiological evidence in conjunction with available information on emissions, exposures and mechanisms for indoor air pollution (Smith et al. 2000; Zelikoff et al. 2003). Three health outcomes were determined to have strong enough evidence to be included: ALRI, COPD and lung cancer (Table 18.11). Information on assessing causality for these outcomes is given in section 3.3 and excluded outcomes are discussed in section 3.2.

Children aged >5 years (of school-age) were excluded as they spend less time in the house than women and children aged <5 years; this is a conservative assumption as there is some exposure of this group, although levels are unknown on a global scale (Ezzati and Kammen 2001; Saksena et al. 1992). Because of the limitations of the available epidemiological studies, only risks in young children (aged <5 years) and adults were included. Available data indicate that men are also affected by those outcomes considered for women, but presumably at lower risks than women because of lower exposures. Adults aged 15–30 years were excluded because the chronic diseases of concern (COPD and lung cancer) have not yet become manifest in this group. Obviously, however, development of these diseases in later years is partly caused by exposures at these and younger ages.

3.2 Excluded health outcomes

Outcomes with insufficient evidence

A number of important diseases that are potentially associated with use of solid fuels have not been included in this analysis owing to insufficient or lack of direct evidence on causality. Lack of inclusion does not necessarily imply inconclusive findings. Rather, it refers to a relatively small set of findings, suggesting that additional, carefully conducted studies are needed to strengthen the evidence base.

Asthma

On the basis of the usual measures (concentrations of small particles, $PM_{2.5}$), typical exposures to indoor smoke from use of solid fuels are much higher than those for urban outdoor pollution (García-Marcos et al. 1999) and ETS (Strachan and Cook 1998), with which asthma has been frequently associated. In addition, a study of children aged <5 years in Malaysia found increased risk associated with the burning of mosquito coils, another important indoor source of $PM_{2.5}$ (Azizi et al. 1995). Studies in China (Xu et al. 1996a) and Kenya (Mohamed et al. 1995) have quantitatively associated asthma in children of school age and in adults with various measures of indoor pollution from solid-fuel use. As the reported background rate is low in most developing countries, however, asthma contributes relatively little to the total burden of deaths or DALYs from indoor air pollution.

Cataracts and other visual impairments

Two case–control studies in India have found an increased risk of cataracts among people using biomass fuel; Mohan et al. (1989) determined an odds ratio of 1.6; Zodpey and Ughade (1999) found an adjusted odds ratio of 2.4. Evaluation of the National Family Health Survey of India (NFHS 1995) found a somewhat lower rate for partial blindness (odds ratio of 1.3; Mishra et al. 1999a), but no significant difference for total blindness. There is also evidence that exposure to ETS is associated with cataracts (West 1992) and animal studies show that cataracts can be caused by wood smoke (Rao et al. 1995; Shalini et al. 1994).

Indoor air pollution may also be linked to blindness through trachoma (Prüss and Mariotti 2000). Two unadjusted studies in the United Republic of Tanzania found such a link (Taylor and West 1989; West and Lynch 1989) although another in Ethiopia found cooking in a central room to be protective, perhaps through reduction of flies (Sahlu and Larson 1992).

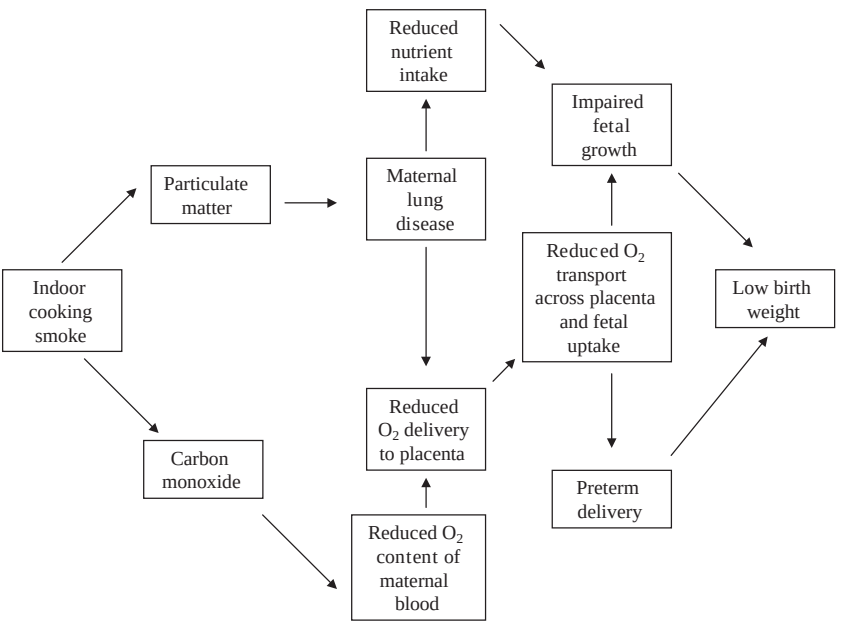
Perinatal effects

One study in India found an adjusted excess risk of stillbirth of 50% among women using biomass fuels during pregnancy (Mavalankar et al. 1991), and two Chinese studies of urban ambient pollution, from the

same group of researchers, also found a strong relationship between concentrations of particulates and pre-term delivery (Xu et al. 1995) and low birth weight (Wang et al. 1997). Low birth weight was also found to be associated with household exposure to biomass smoke in Guatemala (Boy et al. 2002). Intrauterine mortality, low birth weight, prematurity, and early infant death have been significantly associated with urban outdoor pollution at much lower concentrations than those typically found in households that use biomass (Bobak 2000; Loomis et al. 1999; Pereira et al. 1998; Ritz and Yu 1999; Scram 1999; Woodruff et al. 1997). Exposure of non-smoking pregnant women to ETS has been associated with low birth weight in a meta-analysis of 17 studies (Windham et al. 1999a), with low cognitive development (Johnson et al. 1999), but not with spontaneous abortion (Windham et al. 1999b).

Low birth weight is a risk factor for a number of childhood (Walsh 1993) and, probably, adult (Barker 1997) diseases, not just those of the respiratory system. The potential pathways by which indoor cooking smoke may cause low birth weight are given in Figure 18.4. Although this mechanism seems likely to be important in some parts of the world,

Figure 18.4 Possible mechanisms for indoor air pollution and low birth weight



Source: Adapted from Jere D. Haas' schematic diagram of the causal pathway for indoor cooking smoke and birth weight (Smith et al. 2000).

at present it is difficult to provide a quantitative estimate of the potential burden, and it is not attempted here.

Tuberculosis

Recent studies in India and Mexico have suggested that indoor air pollution from use of solid fuel may be a risk factor for active tuberculosis. A statistically significant relationship has been found between reported use of biomass fuel and incidence of tuberculosis in 260 000 adults aged >20 years. Indeed, women in households using biomass fuels were found to be 2.7 (95% CI 1.9–4.0) times more likely to have tuberculosis than women in households using cleaner fuels, even after correction for a range of socioeconomic factors (Mishra et al. 1999b). In addition, an unadjusted but significant odds ratio of 2.5 has been reported for clinically-confirmed tuberculosis in adult male and female householders aged 16–60 years using wood or dung cakes as fuel (Gupta and Mathur 1997). Although these studies were not able to address smoking as a possible confounder, two studies in Mexico City have found an association between exposure to wood smoke and incidence of tuberculosis, after taking smoking into account (Perez-Padilla et al. 1996, 2001). A study in China also found exposure to outdoor air pollution to be associated with tuberculosis (Xu et al. 1995). Animal studies have shown that wood smoke causes immune suppression in the respiratory system (Thomas and Zelikoff 1999; Zelikoff 1994).

Other health effects not included

- Interstitial lung disease has been associated with long-term exposures in several studies (Dhar and Pathania 1991; Gold 2000; Ramage et al. 1988; Sandoval et al. 1993).
- Early studies in Africa seemed to implicate wood smoke as a cause of nasopharyngeal cancer, but this association was not borne out by later studies in Asia (Smith 1987; Smith and Liu 1994).
- Two studies in Brazil have shown increased risk of upper aerodigestive tract cancers, with adjusted odds ratios of 2.7 (Pintos et al. 1998) and 2.5 (Franco et al. 1989).
- An association has been shown with cervical neoplasia among HPV-infected women in Honduras, with an adjusted odds ratio of 5.7 after 35 years or more of cooking over an open fire (Velema et al. 2002).
- Ischaemic heart disease has been associated with exposure to outdoor particulate air pollution (Ponka and Virtanen 1996; Pope et al. 1992; Schwartz 1993; Schwartz and Dockery 1992; Schwartz and Morris 1995) and ETS (Steenland et al. 1998) in some studies, both at much lower levels of exposure than for indoor air pollution (see chapter 17).

Excluded outcomes associated with use of solid fuel, but not caused by exposure to air pollution

The use of solid fuels for household cooking and heating involves a range of activities with potential health implications that are separate from those involving the pollution created. The most important involve the harvesting of the two major types of fuel.

- The harvesting of biomass in rural settings in developing countries may involve regular carriage of heavy loads for long distances, with consequent physical strain and food energy demands, along with exposure to such hazards as snake-bite, leeches and assault (crime). Women and children typically bear the greatest burden of such harvesting, although there is much variation across the world.
- Coal mining is one of the most hazardous occupations in the world, particularly in developing countries in small mines from which much household fuel is obtained.

In addition, the extra time taken to harvest, store, and prepare solid fuels is time that is potentially deducted from other pursuits that are associated with health benefits, such as child care or the generation of the household income.

Considering that the counterfactual distribution is cooking with non-solid fuels (rather than no cooking at all), there are also categories of health risk that are avoided by the use of solid fuels:

- fires and explosions related to household use of liquid and gaseous fuels;
- poisoning caused by ingestion of household kerosene;
- risk inherent in the operation of the national and international petroleum fuel cycles required to provide liquid and gaseous fuels;
- risks involved in providing electricity for household cooking, such as coal mining, air pollution from power plants, accidents involving nuclear and hydroelectric dams, etc.; and
- additional risk of mosquito-borne diseases owing to absence of repellence from household smoke produced by solid fuel.

In its current form, the system limits of this comparative risk assessment (CRA) do not encompass any of these health effects, positive or negative, that are not directly caused by exposure of humans to pollution in the household.

Excluded outcomes associated with specialized airborne products of indoor combustion

There are several related sources of indoor pollution not covered by this analysis that may be locally important in some countries. However, too

few data are available regarding exposures to extrapolate these risks to global burdens, although we suggest that these sources represent potential research topics, as well as priorities for determining exposure distributions, in order to improve the estimated burden of indoor air pollution.

- Smoke from cooking oil: studies in China (including the Province of Taiwan) show relative risks for lung cancer of 3–5 for Chinese-style cooking in a wok with certain cooking oils (Ko 2000; Zhong et al. 1999b).
- Exposures to trace quantities of toxic elements resulting from indoor use of coal in China and elsewhere: significant and widespread impacts from exposures to fluorine and arsenic have been reported in China (Finkelman et al. 1999) and can be expected to occur wherever coal fuels are contaminated with such toxic elements.
- Smoke from incense and mosquito coils, which have been associated with ill-health in some Asian studies (Azizi et al. 1995).

3.3 Evidence and exposure–risk relationships

The estimates of relative risk[5] and confidence intervals used for ALRI, COPD and lung cancer were derived through formal meta-analyses of the available literature.

Searches of the scientific literature were conducted using the Medline computerized bibliographic database, review of bibliographies from previously-retrieved articles and personal communications. In some cases, the authors of articles that were lacking data that were of interest for this analysis were contacted and asked for clarification, and specific requests for information were sent to researchers in this field.

Medline searches were conducted using the following key words:

- For ALRI: indoor air pollution, household fuel, smoke, acute respiratory infections (ARI), pneumonia and ALRI
- For COPD: indoor air, fuel, COPD, chronic obstructive lung disease (COLD) and chronic bronchitis
- For lung cancer: indoor, air, fuel and lung cancer

To be eligible for inclusion in the meta-analysis, studies had to fulfil the following criteria:

- to be a primary study, not a re-analysis or review;
- to examine some proxy for exposure to indoor smoke from the use of solid fuels for cooking and/or heating purposes;
- to report an odds ratio and its variance, or sufficient data with which to estimate them; and

- to be written or abstracted in English. Additionally, for lung cancer only, a Chinese colleague assisted in a comprehensive search of the Chinese literature, extraction of the relevant data and translation.

We considered both fixed- and random-effects models for the meta-analysis. As the results from both were similar, we used those from the fixed-effects model only. Owing to heterogeneity within studies, we performed sensitivity analyses by stratifying the studies by potential sources of heterogeneity, including assessment of exposure and adjustment for confounders. We did not use a random-effects model, even when statistical significance for heterogeneity was present, for the following reasons.

- Random-effects models assume that studies are selected from a population with a single underlying variance. This would be violated given the heterogeneity among the studies in measuring exposure.
- Random-effects models assign the same weight to small and large studies. This would be problematic for the studies of this analysis because the number of cases ranged from 45 to 500.

Smoking is an important risk factor for the diseases associated with indoor smoke from use of solid fuel, especially lung cancer and COPD. At present, information on the combined effects of smoking and use of solid fuel is rare. To avoid possible overestimation of the burden of disease, therefore, attributable fractions for lung cancer and COPD caused by use of solid fuel were applied to disease burdens remaining after removal of the burden attributable to smoking. This is conservative in that some of the effect attributable to smoking could also be attributed to use of solid fuel. To ensure internal consistency within the CRA project, burdens attributable to smoking were obtained from chapter 11. Globally, about 51% and 62%, for men and women respectively, of the total burden of COPD is not attributable to tobacco.

Acute lower respiratory infections

A number of studies in developing countries (Argentina, Brazil, the Gambia, India, Kenya, Nepal, Nigeria, South Africa, the United Republic of Tanzania and Zimbabwe) have quantified the relative risk of ALRI for children in households that burn biomass (Armstrong and Campbell 1991; Campbell 1997; Cerqueiro et al. 1990; Collings et al. 1990; de Francisco et al. 1993; Ezzati and Kammen 2001; Johnson and Aderele 1992; Kossove 1982; Mtango et al. 1992; O'Dempsey et al. 1996; Pandey et al. 1989b; Shah et al. 1994; Victora et al. 1994). Some work has also been done to identify possible mechanisms in the developing countries (Verma and Thakur 1995).

Studies among native Americans (Navajos in the south-western United States of America) show a strong and significant association between ALRI and use of wood stoves, at much lower levels of indoor pollution than found in developing countries (Morris et al. 1990; Robin et al.

1996). There is a larger group of studies that show various childhood respiratory symptoms (e.g. cough, wheezing) to be associated with exposure to smoke from solid fuel, but do not provide sufficient evidence to calculate odds ratios of ALRI itself.

As all studies included here used either ARI or ALRI, or death caused by ARI or ALRI, in children aged <5 years as a health outcome, we only estimated the burden of disease for children in this age group. A recent study in Kenya (Ezzati and Kammen 2001) found associations between use of solid fuels and ARI in adults (both men and women), suggesting that, once time–activity patterns and spatial dispersion of smoke have been taken into account, men and women may have similar patterns of exposure–response.

A single statistical analysis of all 15 studies identified (Table 18.12) was not appropriate because of the heterogeneous exposure variables and the diverse analytical strategies used by the investigators, especially with respect to potential confounding factors. To address this diversity, different subgroups of these studies were used to conduct several meta-analyses, the results of which were remarkably consistent; pooled relative risk estimates increased with improved precision of exposure measure.[6]

Characteristics of excluded studies

Of the 15 studies identified (Table 18.12), we excluded the study by Kossove (1982), which had an inappropriately-small comparison group. Two studies in South America focused on use of solid fuels in urban populations (Cerqueiro et al. 1990; Victora et al. 1994). The study in Buenos Aires, Argentina, was excluded owing to a very low prevalence of households using solid fuels and, in one of the case groups, missing data on exposure to heating fuelled by charcoal (Cerqueiro et al. 1990). In the study in Brazil (Victora et al. 1994), only a small proportion of the study population was exposed (6%) and exposure was defined loosely, encompassing a wide range of sources of pollution, from open fires to enclosed metal heating stoves and vented fireplaces. The study by Shah et al. (1994) was excluded because its definition of non-exposure (use of stove with chimney provided by the government improved-stove programme) has been shown to produce concentrations of indoor pollutants that were not statistically different from those produced by open fires at that time in India (Ramakrishna et al. 1989) and no observations of direct pollution were made. The study by Mtango et al. (1992) was excluded because, as the study focused on mortality from all causes, no information was given on exposure status for the proportion of deaths caused by ALRI (in this case, pneumonia). Two studies reported on the same study population (Armstrong and Campbell 1991; Campbell et al. 1989). We chose to include the older report by Campbell, which included the odds ratio for girls and boys combined. A recently-published longitudinal study examining rates of episodes of

Table 18.12 Studies on the risk of acute lower respiratory infection associated with use of solid fuels, in children aged <5 years

Study location	*Reference*	*Study design (n)* *Study population*	*Exposure assessment*	*Outcome assessment*	*Covariates adjusted for*	*Odds ratio (95% CI)*
Argentina	Cerqueiro et al. (1990) *Excluded*	Case–control (616–669) Children aged <5 years	Questionnaire: type of cooking fuel used (wood, kerosene, gas)	ALRI within the last 12 days, at a well-baby clinic	None	9.9 (1.8–31.4)
Brazil (urban)	Victora et al. (1994) *Excluded*	Case–control (510–510) Children aged <2 years	Questionnaire: presence of indoor smoke	ALRI hospital cases, clinical signs and X-ray	Smoking, housing, no. of siblings, income, education, history of respiratory illness	1.1 (0.6–2.0)
Gambia	Armstrong and Campbell (1991) *Excluded*	Cohort (500) Children aged <5 years	Questionnaire: mother carries child on her back while cooking	ALRI, by weekly home visits	Birth interval, ETS, crowding, socioeconomic status, nutrition, vaccination, education	Males: 0.5 (0.2–1.2) Females: 1.9 (1.0–3.9)
Gambia	Campbell et al. (1989)	Cohort (271) Children aged <1 year	Questionnaire: mother carries child on her back while cooking	ALRI, by weekly home visits	Birth interval, ETS, crowding, socioeconomic status, nutrition, vaccination, education	2.8 (1.3–6.1)
Gambia Upper River Division	de Francisco et al. (1993)	Case–control (129–270) Children aged <2 years	Questionnaire: mother carries child on her back while cooking	Death from ALRI by verbal autopsy confirmed by three independent physicians	Socioeconomic status, ETS, maternal education, crowding, nutrition	5.2 (1.7–15.9)
Gambia Upper River Division	O'Dempsey et al. (1996)	Case–control (80–159) Children aged <5 years	Questionnaire: mother carries child on her back while cooking	ALRI hospital cases, clinical signs, X-ray and laboratory	ETS, mother's income, weight slope, recent illness, nutrition	2.5 (1.0–6.6)
India	Shah et al. (1994) *Excluded*	Case–control (400) Children aged ≤5 years	Household has a smoke-producing stove	Severe ARI hospital cases, clinical symptoms	Smoking, housing, no. of siblings, income, education, birth weight	1.2 (0.7–2.3)

Kenya	Ezzati and Kammen (2001) *Excluded*	Cohort (93) Children aged <5 years	Mean daily personal PM_{10} exposure from pollution and time-location data	Rate of ALRI during study period by Integrated Management of Childhood Illness (IMCI) diagnosis criteria	Age, sex, crowding, smoking, village type	2.93 (1.34–6.39) Highest vs lowest exposure category plus exposure–response trend
Nepal	Pandey et al. (1989b)	Cohort (280) Children aged <2 years	Questionnaire: Average time spent near the fireplace	ARI, by bi-weekly home visits	None	2.3 (1.8–2.9)
Nigeria	Johnson and Aderele (1992)	Case–control (103–103) Children aged <5 years	Questionnaire: type of cooking fuel used (wood, kerosene, gas)	ALRI hospital cases, clinical signs, X-ray and laboratory	None	0.8 (0.4–1.7)
South Africa	Kossove (1982) *Excluded*	Case–control (132–18) Children aged ≤1 year	Questionnaire: does the child stay in the smoke?	ALRI hospital cases, clinical signs and X-ray	None	4.8 (1.7–13.6)
United Republic of Tanzania	Mtango et al. (1992) *Excluded*	Case–control (456–1160) Children aged <5 years	Questionnaire: child sleeps in room where cooking is done	Death from all causes, by verbal autopsy and physician	Village, age, questionnaire respondent, maternal education, parity, water source, child eating habits	2.8 (1.8–4.3)
USA Arizona	Morris et al. (1990)	Case–control (58–58) Children aged <2 years	Questionnaire: primary source for heating and cooking	ALRI hospital cases, clinical signs and X-ray	Family history of asthma, recent respiratory illness, dirt floor, running water	4.9 (1.7–12.9)
USA Arizona	Robin et al. (1996)	Case–control (45–45) Children aged <2 years	Questionnaire: household uses wood for cooking	ALRI hospital cases	No. of siblings, electricity, running water, difficulty in transport to clinic, ETS, housing	5.0 (0.6–42.8)
Zimbabwe	Collings et al. (1990)	Case–control (244–500) Children aged <3 years	Questionnaire: household uses open wood-fire for cooking	ALRI hospital cases, clinical signs and X-ray	ETS, crowding, housing, number of siblings	2.2 (1.4–3.3)

ALRI in a range of age groups across several categories of exposure to smoke from combustion of biomass in Kenya (Ezzati and Kammen 2001) was excluded from the formal meta-analysis because the outcome, expressed as "fraction of weeks with illness", could not be translated into an odds ratio in a manner consistent with the other epidemiological studies. This study did provide strong collaborative evidence, nevertheless, for it showed effects in older children and women as well as in young children and demonstrated a statistically significant trend in the exposure–response relationship. In a subsequent analysis, the authors reported an odds ratio of 2.14 for children exposed to PM_{10} concentrations of >1000 $\mu g/m^3$ (Ezzati 2002).

Estimating risk factor–disease relationships

After the exclusions noted above, there remained eight studies that reported relative risks of acute respiratory illness for young children exposed to indoor smoke from use of solid fuel (Campbell et al. 1989; Collings et al. 1990; de Francisco et al. 1993; Johnson and Aderele 1992; Morris et al. 1990; O'Dempsey et al. 1996; Pandey et al. 1989b; Robin et al. 1996). Of these, the majority were case–control studies. One study used the outcome "pneumococcal infection", which includes meningitis and septicaemia (O'Dempsey et al. 1996). However, 80% of patients in this study were diagnosed with pneumonia. Although most of the studies were conducted in developing countries, two were carried out in populations of Navajo and Hopi Indians in the United States (Morris et al. 1990; Robin et al. 1996). The populations in the United States are likely to differ in socioeconomic characteristics from the rest of the studies, thus potentially influencing the rates of incidence of ALRI. As the overall odds ratio did not change substantially with the exclusion or inclusion of these studies, all subsequent analyses included these two studies.

Exposure assessment used in the studies

The studies provide relatively little information on the indoor concentrations of or exposures to specific pollutants produced by use of solid fuels, or on the baseline concentrations within similarly-constructed households that do not use solid fuels. All but one study used binary classifications of exposure (Table 18.12). On the basis of evidence for an exposure–response relationship between ARI and exposure to smoke from solid fuels (Ezzati and Kammen 2001; Pandey et al. 1989a), we attempted to analyse the studies according to the precision of the exposure measure used and the likely intensity of exposure. Exposure measures used were grouped in three major categories, in what was assumed to be an increasing order of precision: type of fuel used, duration of exposure to smoke from solid fuels, and using solid fuel and carrying the child on the mother's back (Table 18.13). Although it is generally true that concentrations of pollutants are likely to be lower in households using cleaner fuels, such as kerosene or gas, there is a wide variation in con-

Table 18.13 The risk of ALRI associated with use of solid fuels, in children aged <5 years: subgroup analyses

Subgroup analyses	*Studies included*	*Odds ratio (95% CI)*
All studies	Campbell et al. (1989); Collings et al. (1990); de Francisco et al. (1993); Johnson and Aderele (1992); Morris et al. (1990); O'Dempsey et al. (1996); Pandey et al. (1989b); Robin et al. (1996)	2.3 (1.9–2.7)
Use of solid fuel	Johnson and Aderele (1992); Collings et al. (1990); Morris et al. (1990); Robin et al. (1996)	2.0 (1.4–2.8)
Duration of time child spent near the cooking fire	Pandey et al. (1989b)	2.3 (1.8–2.9)
Child is carried on the mother's back	Campbell et al. (1989); de Francisco et al. (1993); O'Dempsey et al. (1996)	3.1 (1.8–5.3)
Studies adjusting for nutritional status	Campbell et al. (1989); de Francisco et al. (1993); O'Dempsey et al. (1996)	3.1 (1.8–5.3)
Studies not adjusting for nutritional status	Collings et al. (1990); Johnson and Aderele (1992); Morris et al. (1990); Pandey et al. (1989b); Robin et al. (1996)	2.2 (1.8–2.6)
Children aged <2 years old	Campbell et al. (1989); de Francisco et al. (1993); Morris et al. (1990); Pandey et al. (1989b); Robin et al. (1996)	2.5 (2.0–3.0)
Children aged <5 years old	Collings et al. (1990); Johnson and Aderele (1992); O'Dempsey et al. (1996)	1.8 (1.3–2.5)

centrations reported in households using solid fuels (Mehta and Smith 2002). Some studies report whether or not children remained indoors when the mother was cooking, but, for reasons noted above, all of these studies were excluded (Awasthi et al. 1996; Kossove 1982; Mtango et al. 1992). Only one study reported the average time that the child spent near the cooking fire (Pandey et al. 1989b). We assumed that carrying the child on the mother's back during cooking represented the most precise measure of exposure, as this suggests that the child was in close proximity to the fire, where exposures are generally higher (although the type of fuel used in control households in these studies was not specified).

We performed separate analyses for each category of exposure, as summarized in Table 18.13. Cooking with wood or other biomass was associated with an odds ratio of 2.0, 95% CI 1.4–2.8. The Pandey study reported an intermediate estimate of relative risk of 2.3, 95% CI 1.8–2.9, for children spending more than two hours near the cooking fire each day. The highest odds ratio was found to be associated with the child being carried on the mother's back during cooking (odds ratio of 3.1, 95% CI 1.8–5.3).

In only three of the studies were the results adjusted for nutritional status in multivariate analyses, an important confounding variable for

ARI in young children (Victora et al. 1999). The odds ratio found by those studies that did adjust was 3.1, 95% CI 1.8–5.3, whereas the effect was slightly smaller in the studies that did not adjust, with an odds ratio of 2.2, 95% CI 1.8–2.6. This may be explained, however, by the fact that the studies that controlled for nutrition also used a different exposure proxy (child was carried on mother's back during cooking).

Age is another potential confounding variable because younger children are more likely to remain close to their mothers and are therefore also more likely to be exposed to indoor smoke from cooking or heating, and because age is independently associated with ALRI, with younger children being more susceptible than older children. Most case–control studies adjusted for age by matching controls to cases. When the analysis was restricted to include only studies in children aged <2 years, the risk of ALRI was found to be slightly higher (odds ratio of 2.3, 95% CI 1.9–2.7) than that obtained from studies in children aged ≤5 years (odds ratio of 1.6, 95% CI 1.2–2.2). Armstrong and Campbell (1991) noted that, in their study population, girls were more likely to be carried on their mothers' backs than boys and were thus exposed to higher concentrations of pollutants for a longer duration of time. This study found that girls who were carried on the mother's back during cooking had an increased risk of ALRI; no association was observed for boys. The risk in girls was much higher (odds ratio of 6.0 vs odds ratio of 1.9) when only the first episode of ALRI (rather than all episodes) was included in the analysis, although the confidence interval was also much wider, owing to the smaller sample. Data were not disaggregated by sex in any of the other studies (although several did control for sex in the multivariate analyses).

As we could not separate the effects of measures of exposure from adjustment for nutritional status, we used the combined odds ratios for all eight studies remaining after exclusions. The results of this approach are similar to those that would be produced if the difference between the most and least precise exposure measures were to be used as the range, i.e. 2.0–3.1 (geometric mean, GM=2.4). This is also consistent with the differences in the odds ratios for the two age groups, that is, 1.8 for children aged <5 years and 2.5 for children aged <2 years. The overall estimate, from all eight studies, of the risk of ALRI in young children exposed to indoor air pollution caused by use of solid fuels was 2.3, CI 95% 1.9–2.7.

Chronic obstructive pulmonary disease

Globally, the most important risk factor for COPD is thought to be smoking of tobacco (NHLBI/WHO 2001) (see also chapter 11). A number of studies have examined various symptoms of chronic respiratory ill-health in women who cook with open stoves burning biomass (Smith 2000). Eight studies in six countries—Bolivia (Albalak et al. 1999), Colombia (Dennis et al. 1996), India (Gupta and Mathur 1997;

Malik 1985), Mexico (Perez-Padilla et al. 1996), Nepal (Pandey 1984b; Pandey et al. 1988) and Saudi Arabia (Døssing et al. 1994)—have quantified the association between indoor air pollution and COPD. Although there are no comparable studies reporting odds ratios in China, the high rates of COPD in non-smoking Chinese women argue that this risk can be related to exposure to coal smoke (Liu et al. 1998).

Cor pulmonale, a heart condition that is secondary to COPD and that is also found among non-smoking rural women in south Asia (Smith 1987), has long been attributed to long-term exposure to smoke from biomass (Padmavati and Pathak 1959). Other studies have attributed silicosis (Norboo et al. 1991; Saiyed et al. 1991), reductions in lung function, cough and various other respiratory conditions to exposure to smoke from biomass, in women,[7] but were not however included here, owing to limited evidence and the relatively small burden of disease associated with these conditions.

Studies that were included in the meta-analysis used a specific definition of COPD or chronic bronchitis, such as cough and sputum on every day for at least three consecutive months for two successive years, and/or a forced expiratory volume in first second/forced vital capacity (FEV1/FVC) ratio of <70% or a FEV1 of <70% of the predicted value. We identified 11 studies reporting the relative risks of chronic airway disease in adults exposed to indoor smoke (Albalak et al. 1999; Behera et al. 1991; Dennis et al. 1996; Døssing et al. 1994; Dutt et al. 1996; Gupta and Mathur 1997; Malik 1985; Menezes et al. 1994; Pandey 1984a; Perez-Padilla et al. 1996; Qureshi 1994). Of these, one was a cohort study (Dutt et al. 1996) and three were case–control studies (Dennis et al. 1996; Døssing et al. 1994; Perez-Padilla et al. 1996). The remaining six studies were cross-sectional (Table 18.14).

Where studies reported exposure as a continuous variable, categories were constructed post hoc to be consistent with studies that presented the same exposure or a similar exposure as a categorical variable (e.g. average time spent daily near the stove, <2 hours and >2 hours). More than half of the study populations in Table 18.14 originated from rural areas where cooking on an open fire in ill-ventilated huts was common. Five study sites, however, were in urban or peri-urban settings where a mixture of fuels might be used (see Table 18.14) and where exposure to indoor smoke is likely to be lower than for women living in rural areas.

Estimating the relationship between risk factor and disease

Smoking is an important potential confounding variable for COPD and particularly so if men are included in the analysis, given the higher prevalence of smoking in men than in women in developing countries. Only two studies adjusted for smoking (Dennis et al. 1996; Menezes et al. 1994). Of the studies that did not adjust for smoking, two included non-smokers only (Behera et al. 1991; Dutt et al. 1996), another reported an overall prevalence of smoking of <1% in the entire study population

Table 18.14 Studies on the risk of chronic obstructive pulmonary disease associated with use of solid fuels

Study location	*Author (year of publication)*	*Study design (n) Study population*	*Exposure assessment*	*Outcome assessment*	*Covariates adjusted for*	*Odds ratio (95% CI)*
Bolivia	Albalak et al. (1999)	Cross-sectional (241) Females+males aged >20 years	Cooking inside or outside	CB	Age, sex	2.5 (1.25–5)
Brazil (urban)	Menezes et al. (1994)	Cross-sectional (1053) Females+males aged >40 years	Presence of at least two of the following: open fire, charcoal stove, paraffin lamp or coal heater	CB	Age, sex, race, income, schooling, smoking, childhood respiratory illnesses, occupational exposures	1.3 (0.75–2.27)
Colombia (urban)	Dennis et al. (1996)	Case–control (104–104) Females aged >35 years	Use of solid bio-fuel for cooking (wood)	COPD,[a] COPD+CB	Age, smoking, hospital	3.92 (1.16–9.1)
India (rural)	Gupta and Mathur (1997)	Cross-sectional (707) Females+males aged >15 years	Use of solid bio-fuel for cooking (wood+dung)	CB+bronchial asthma	Age	7.9 (2.84–21.8)
Northern India	Behera et al. (1991)	Cross-sectional (3718) Females involved in cooking	Use of solid bio-fuel for cooking (wood+dung)	CB	None	1.97 (1.16–3.22)
Northern India	Malik (1985)	Cross-sectional (2180) Females aged >20 years	Use of solid bio-fuel for cooking (wood)	CB, COPD+CB	None	3.0 (1.77–4.93)
Southern India (urban)	Dutt et al. (1996)	Cohort (315) Females aged 15–60 years	Use of solid bio-fuel (wood) for cooking	CB	None, age-stratified sampling	2.8 (0.7–11.4)

India Ladakh, Himalaya	Norboo et al. (1991) *Excluded*	Cross-sectional (364) Females+males aged >20 years	Carbon monoxide level	CB	NA	NA
Kashmir	Qureshi (1994)	Cross-sectional (560) Females+males aged >15 years	Average time spent near the fireplace (>4 hours vs <4 hours)	CB	None	3.5 (1.4–8.77)
Mexico (urban)	Perez-Padilla et al. (1996)	Case–control (127–375) Females aged >40 years	Use of solid bio-fuel for cooking and heating (wood)	CB	Age, place of residence, education, income, smoking	4.1 (2.3–9.4)
Mozambique (urban)	Ellegard (1996) *Excluded*	Cross-sectional (1188) Females+males aged >14 years	Estimate of lifetime exposure to cooking fuel	Cough index	Economic and environmental variables	NA
Nepal	Pandey (1984a)	Cross-sectional (1375) Females+males aged >20 years	Use of solid bio-fuel for cooking (wood+straw)	CB	Age	5.4 (2.96–9.78)
Nepal	Pandey et al. (1985) *Excluded*	Cross-sectional (150) Females aged 30–44 years	Daily duration of exposure to domestic smoke	FVC	None	NA
Saudi Arabia	Døssing et al. (1994)	Case–control (50–71) Females+males admitted to the hospital	Ever exposed to open cooking fire	COPD[b]	None, matched for age and sex	14.4 (5.5–37.5)

Key: CB, chronic bronchitis, defined as cough and sputum on most days for at least three consecutive months of two successive years; NA, not applicable; FEV, forced expiratory volume; FEV1, forced expiratory volume in first second; FVC, forced vital capacity.

[a] COPD = FEV1/FVC <70% without asthma or FEV1<70% of predicted value.

[b] COPD = FEV1/FVC <70%, FEV1<70% of predicted value and <15% or <250 cm^3 absolute increase after 200 μg of aerosolized salbutamol.

(Albalak et al. 1999). Pandey (1984a) reported the data stratified by smoking status and finally, the study by Perez-Padilla et al. (1996) reported that 70–80% of the subjects indicated that they had never smoked.

Two studies (Døssing et al. 1994; Gupta and Mathur 1997), which included men and women and reported a relatively high prevalence of smoking in their study populations (not equally distributed between COPD cases and controls), did not adjust for smoking. The combined estimate of risk from the group of studies that accounted for smoking, and excluding the Døssing et al. and Gupta and Mathur studies, was 2.5, 95% CI 1.9–3.3. The combined estimate of relative risk for the studies by Døssing et al. and Gupta and Mathur that did not adjust for smoking, and which is thus likely to be an overestimation, was substantially higher at 10.8, 95% CI 5.4–21.8. Another major confounding variable in the association between risk of COPD and exposure to indoor smoke is age, with absolute risk increasing with age. Most studies adjusted for age by matching, stratified sampling (Dutt et al. 1996), or by adjustment in the analysis; two studies (Malik 1985; Qureshi 1994) reported the mean age to be similar in the exposed vs unexposed subjects. A potential problem of confounding by age remains with the studies by Pandey (1984a) and Behera et al. (1991), which showed no data on the age distribution. The combined estimate of the relative risk excluding these two studies was 2.9, 95% CI 2.2–3.6.

This analysis primarily included women as they comprise the population that is most frequently exposed to smoke from wood during cooking and which is thus at greatest risk of developing chronic airway disease. Therefore, we included estimates for women or the combined estimate adjusted for sex, if available. With two exceptions (Døssing et al. 1994; Gupta and Mathur 1997), all studies reported the data separately for men and women, or combined the data while adjusting for sex. The overall estimate of relative risk for all studies included was 2.9, 95% CI 2.2–3.8. For men, it was 2.8, 95% CI 1.4–5.7, but this was based on only two studies, one of which did not correct for age (Døssing et al. 1994; Qureshi 1994). See Table 18.15 for details.

All three case–control studies were hospital-based; control groups consisted of visitors to patients other than the study subjects (Døssing et al. 1994), patients with illnesses other than those of the respiratory tract (Dennis et al. 1996) and a mixture of visitors, patients diagnosed with tuberculosis or interstitial lung disease and patients with otolaryngological problems (Perez-Padilla et al. 1996). Bias could have been introduced by the choice of visitor controls if exposure to indoor smoke was related to the likelihood to come to the hospital to visit a patient, or by the selection of inpatient controls, if exposure to indoor smoke made the patients with the control diseases less or more likely to be referred to the hospital (e.g. tuberculosis).

Table 18.15 The risk of chronic obstructive pulmonary disease associated with use of solid fuels: subgroup analyses

	Subgroup analyses	*Studies included*	*Odds ratio (95% CI)*
Males and females	Rural population	Too few studies available to allow odds ratio to be calculated	NA
	Urban population	Too few studies available to allow odds ratio to be calculated	NA
	Adjusted for smoking	Albalak et al. (1999); Menezes et al. (1994); Pandey (1984a)	2.51 (1.76–3.56)
	Not adjusted for smoking	Qureshi (1994); Døssing et al. (1994); Gupta and Mathur (1997)	5.8 (3.74–8.99)
	Adjusted for age	Albalak et al. (1999); Døssing et al. (1994); Gupta and Mathur (1997); Menezes et al. (1994)	3.3 (2.32–4.69)
Females only	Adjusted for smoking but not for age	Behera et al. (1991); Pandey (1984a); Qureshi (1994)	2.56 (1.75–3.75)
	Adjusted for smoking and age	Dutt et al. (1996); Perez-Padilla et al. (1996); Dennis et al. (1996); Malik (1985)	2.83 (2.0–3.97)
Males only	Not adjusted	NA, too few studies—Døssing et al. (1994) adjusted for age; Qureshi (1994) adjusted for none	NA, see also text
	Adjusted for smoking and age	None of the studies in males adjusted for both age and smoking	NA

NA Not applicable.

The final model for women excluded the three studies that did not adjust for age and/or smoking status. The overall risk of COPD in women exposed to indoor air pollution from use of solid fuels was estimated as 3.2, 95% CI 2.3–4.8. There is much less evidence available about the impact on men, but the risk seems to be lower, 1.8, 95% CI 1.0–3.2,[8] presumably because of lower exposure.

LUNG CANCER

Lung cancer in women has been associated with cooking with open coal stoves in China on the basis of a number of studies. In China, there is also evidence that lung cancer is caused by use of certain cooking oils

(Zhong et al. 1999a, 1999b) as well as by exposures to known carcinogens contained in coal smoke, such as arsenic (Finkelman et al. 1999). There is limited evidence available for an association between lung cancer and use of biomass fuels in women, but not in men (Gao et al. 1987; Liu et al. 1993; Sobue 1990), although several pollutants in biomass smoke are known or suspected human carcinogens (Smith 1987).

The majority of the internationally published studies on lung cancer and indoor air pollution that we were able to locate were conducted in China. One took place in Japan (Sobue 1990) and one in the United States (Wu et al. 1985). Two eligible studies were published in Chinese only (Huang 1999; Wu et al. 1999). All 19 studies identified were case–control studies, including either newly-diagnosed cases of lung cancer at a hospital or using death registries, and of these, 14 studies were hospital-based. Inherent in the choice of this design is Berkson's bias, referring to the possibility that controls (men and women hospitalized with other diseases) are not selected independently of exposure in the source population. With two exceptions (Ko et al. 1997; Sobue 1990), all studies used population controls, which minimizes such bias (Table 18.16).

Characteristics of excluded studies

Of the 19 studies identified, we excluded three (Du et al. 1988; Xu et al. 1996b; Yang et al. 1990). The ecologic study by Yang et al. (1990) neither adjusted for smoking or other risk factors nor provided sufficient information to calculate odds ratios. Of two articles which reported on the same study population (Du et al. 1988, 1996), we included the more recent, which provided 95% CIs for the relative risk. More than one article reported on a collaborative study that included men and women of two major cities in the Province of Liaoning (Wu-Williams et al. 1990; Xu et al. 1996b); we included only the study by Wu-Williams et al. (1990), which combined all female lung cancer cases from the death registries of the two cities. The study by Xu et al. (1996b) considered cases in males and females from one city only.

Estimating risk factor–disease relationship

Although the 16 studies included in this analysis were all case–control designs, measurement of exposure to indoor air pollution was carried out by a multitude of methods. Seven studies assessed exposure to indoor air pollution in terms of years of exposure (Dai et al. 1996; Ko et al. 1997; Lei et al. 1996; Liu et al. 1991; Sobue 1990; Wu et al. 1999; Wu-Williams et al. 1990). The remaining eight studies merely determined whether coal and/or bio-fuel were generally used for cooking or heating (Du et al. 1996; Gao et al. 1987; Huang 1999; J. Liu and H. Hu, unpublished data, 1996; Liu et al. 1993; Shen et al. 1996; Wang et al. 1996; Wu et al. 1999). In order to explore the characteristics responsible for

Table 18.16 Studies on the risk of lung cancer associated with use of solid fuels

Study location	*Reference*	*Study design (n)* *Study population*	*Exposure assessment*	*Outcome assessment*[a]	*Covariates adjusted for*	*Odds ratio (95% CI)*
China Fujian Province Fuzhou	Luo et al. (1996)	Case–control (102–306) Females+males	Indoor combustion of coal	Newly-diagnosed lung cancer	Smoking, passive smoking, chronic bronchitis and matched for age and sex	ADC: 6.0 (1.36–23.49) SCC: 14.1 (1.67–119.4)
China Gaunxi Province Nanning	Huang (1999)	Case–control (122–244) Females+males	Use of coal	Newly-diagnosed lung cancer	Smoking, chronic lung disease, meat consumption, depression, SES, BMI, exercise	1.76 (1.3–2.38)
China Guangzhou	Du and Ou (1990) *Excluded*	Case–control (662–662) Females+males	Exposed to coal fumes yes/no	Deaths from lung cancer over 5 years	Matched for age, sex, residence	14.52 (—)
China Guangzhou	Du et al. (1996)	Case–control (120–240) Non-smoking females+males	Exposed to coal fumes yes/no	Death from lung cancer	Smoking and chronic respiratory disease	Females: 1.56 (0.57–4.25) Males: 1.5 (0.69–3.27)
China Guangzhou	Lei et al. (1996)	Case–control (792–792) Females+males	Cooking for >40 years	Death from lung cancer	Matched for age and sex	0.93 (0.67–1.21)
China Guangzhou	Liu et al. (1993)	Case–control (316–316) Females+males	Use of coal and wood for cooking	Newly-diagnosed lung cancer	Smoking, passive smoking, education, SES, history of cancer	Coal: 1.46 (0.83–2.56) Bio-fuel: 1.19 (0.46–3.11)
China Guangzhou	Wu et al. (1999)	Case–control (258–258) Females	Use of coal as residential fuel	Newly-diagnosed lung cancer	Smoking, history of tuberculosis, fruit consumption, ventilation of kitchen	1.57 (0.89–2.82)

continued

Table 18.16 Studies on the risk of lung cancer associated with use of solid fuels (*continued*)

Study location	*Reference*	*Study design (n)* *Study population*	*Exposure assessment*	*Outcome assessment*[a]	*Covariates adjusted for*	*Odds ratio (95% CI)*
China Liaoning Province Harbin	Dai et al. (1996)	Case–control (120–120) Non-smokers, females	Use of coal heater for 25–34 years	Newly-diagnosed lung cancer	History of family cancer, income, carrot consumption, deep fried cooking	4.7 (1.28–17.18)
China Liaoning Province Harbin and Shenyang	Wu-Williams et al. (1990)	Case–control (956–952) Females	Use of coal stove for >40 years	Newly-diagnosed lung cancer	Age, education, smoking	1.3 (1–1.7)
China Liaoning Province Shenyang	Wang et al. (1996)	Case–control (135–135) Females	Use of coal for cooking	Newly-diagnosed lung cancer	Family history of cancer, ETS	0.75 (0.43–1.31)
China Liaoning Province Shenyang	Xu et al. (1996b) *Excluded*	Case–control (1249–1345) Females+males	Use of coal stove for cooking	Newly-diagnosed lung cancer from cancer registry	None	Females: 1.5 (—) Males: 2.3 (—)
China Shanghai	Gao et al. (1987)	Case–control (672–735) Females	Cooking with coal or bio-fuel	Newly-diagnosed lung cancer	Smoking, education, age	Coal: 0.9 (0.7–1.3) Bio-fuel: 1.0 (0.6–1.8)
China Beijing and Shunyi	J. Liu and H. Hu, unpublished data, 1996	Case–control (220–440) Females+males, farmers	Combustion of coal cakes	Death from lung cancer	Smoking, chronic respiratory disease and matched for age	1.9 (1.16–3.43)

China Hubei Province Wuhan	Yang et al. (1990) *Excluded*	Cross-sectional; ecologic design (50–200) Females+males, two parts of city	Use of coal for cooking	Death from lung cancer	None	—
China Jiangsu Province Nanjing	Shen et al. (1996)	Case–control (263–263) Females+males	Use of solid fuels	Newly-diagnosed lung cancer	Matched for age and sex and multivariate (final model not shown)	4.97 (0.8–30.88)
China Yunnan Province Xuanwei	Liu et al. (1991)	Case–control (110–426) Females+males, farmers	Starting to cook before 10 years of age	Newly-diagnosed lung cancer	Smoking and matched for age, sex, and village	Females: 1.25 (0.45–3.49) Males: 3.36 (1.27–8.88)
China Province of Taiwan	Ko et al. (1997)	Case–control (117–117) Females	Start cooking either coal or bio-fuel between 20–40 years of age	Newly-diagnosed lung cancer	Education, place of residence, SES	Coal: 1.3 (0.3–5.8) Bio-fuel: 2.7 (0.9–8.9)
Japan Osaka	Sobue (1990)	Case–control (144–731) Non-smoking females	Use of bio-fuel for cooking at 15 or 30 years of age	Newly-diagnosed lung cancer	Age, education	1.77 (1.08–2.91)
USA Los Angeles	Wu et al. (1985)	Case–control (220–220) Females+males	Use of coal for cooking and heating during childhood	Newly-diagnosed lung cancer	Smoking and matched for age and place of residence	ADC: 2.3 (1.0–5.5) SCC: 1.9 (0.5–6.5)

Key: ADC, adenocarcinoma; SCC, squamous cell carcinoma; Bio-fuel, wood, straw; BMI, body mass index; SES, socioeconomic status.

— No data.

[a] Studies that were included in the meta-analysis examined as primary outcome either cases of lung cancer (as defined by either clinical signs and symptoms, X-ray and/or histological and pathological findings) or death due to lung cancer.

heterogeneity found in the results of a meta-analysis of all studies, several subgroup analyses were conducted, in which stratification by type of fuel used (mostly coal and some wood) and sex was used. The variability of exposure categories was too great and the number of studies too small to be grouped for duration of exposure. If a study reported an estimate of relative risk for several exposure categories, the odds ratio for the category representing the longest period of exposure was used (Lei et al. 1996; Wu-Williams et al. 1990). Two studies (Luo et al. 1996; Wu et al. 1985) reported separate estimates for adenocarcinoma and squamous cell carcinoma; these were entered as separate studies as we were unable to achieve a combined estimate. Whenever possible, separate estimates for men and women were extracted and entered as individual studies (Du et al. 1996; Liu et al. 1991).

In a recent review of the literature on indoor air pollution and several health outcomes (Bruce et al. 2000), the most prominent concern voiced was regarding the lack of control for confounders. Therefore, we conducted stratified analyses based on studies that accounted for the most common potential confounders, such as smoking and the presence of a chronic respiratory disease. All studies included in the meta-analysis either adjusted for smoking or included only non-smokers. It has been suggested that chronic respiratory diseases such as chronic bronchitis, tuberculosis, asthma and emphysema that originate from infections or other predispositions may increase the probability of developing lung cancer later in life (Luo et al. 1996). We examined the effect of indoor air pollution from coal smoke on men and women separately. Nine studies either only included women or presented risk estimates for men and women separately (Dai et al. 1996; Du et al. 1996; Gao et al. 1987; Ko et al. 1997; Liu et al. 1991; Sobue 1990; Wang et al. 1996; Wu et al. 1985, 1999). The overall estimate for females was 1.17, 95% CI 1.02–1.35. The analysis restricted to studies that adjusted for smoking and chronic respiratory disease indicated a substantial increase in risk for women of almost two-fold (odds ratio of 1.94, 95% CI 1.09–3.47).

Five studies presented a combined risk estimate for men and women (Huang 1999; Lei et al. 1996; Liu et al. 1993; Luo et al. 1996; Shen et al. 1996), producing a summary odds ratio of 1.86 (95% CI 1.48–2.35). Restricting the analysis to the three studies that controlled for smoking and chronic respiratory disease showed a substantial increase in risk (odds ratio of 2.55, 95% CI 1.58–4.10).

Only three studies either included males only (Wu et al. 1999) or presented sufficient data to extract a separate estimate for males (Du et al. 1996; Liu et al. 1991). The risk associated with coal use for the male population was 1.79, 95% CI 1.18–2.72, and slightly lower when taking into account confounding by smoking and chronic airway disease (odds ratio of 1.51, 95% CI 0.97–2.46). Although the results of the two studies

Table 18.17 Summary of results of subgroup meta-analyses

	Odds ratio (95% CI)	
Subgroup analyses	*Not adjusted*	*Adjusted for smoking and chronic airway disease*
Males and females—coal use	1.86 (1.48–2.35)	2.55 (1.58–4.10)
Males only—coal use	1.79 (1.18–2.72)	1.51 (0.97–2.46
Females only—coal use	1.17 (1.02–1.35)	1.94 (1.09–3.47)

comprising this model were not quite statistically significant (lower confidence limit was 0.97), the pattern of significance of the five studies assessing risks for men and women combined, give confidence that there is likely to be a real effect on men. Odds ratios are shown in Table 18.17.

3.4 Sources of uncertainty

Uncertainty estimates were generated through the use of meta-analyses for all the disease end-points included. A critical problem with extrapolating the results of epidemiological studies from one subregion to another, particularly between developed and developing regions, is the difference in other potentially interactive risk factors, such as malnutrition, which are not addressed by the methodology. That all the studies used for the calculations of solid-fuel use were done in developing countries, however, does provide some confidence that differences in competing risks were not excessive. Meta-analytical confidence intervals probably underestimate true uncertainty because of variations in the way different studies dealt with measures of exposure, adjustment for confounding, and outcome definitions, as well as the need to extrapolate results across populations.

3.5 Risk reversibility

There are few studies on the reversibility of the health effects of smoke from solid fuel. For acute outcomes (ALRI), evidence from risk factors for other childhood infectious diseases may provide some guidance (Jones et al. 2003). For the chronic conditions, COPD and lung cancer, the timing is less clear, however, since the increased risk presumably results from many years of exposure. A retrospective cohort study in China, however, did find a statistically significant drop in lung cancer rates associated with introduction of improved stoves with flues in around 1980 (Lan et al. 2002). The delay between intervention and a discernible reduction in lung cancer incidence was about 10 years, consistent with that observed after smoking cessation (see chapter 11).

4. Results

4.1 Attributable burden of disease

As shown in Table 18.18, the burden of disease attributed to use of household solid fuels is dominated by that caused by ALRI in young children, which accounts for 59% of all attributed premature deaths and 78% of DALYs. COPD accounts for nearly all the remainder, with the burden from lung cancer a relatively minor contributor, owing to the concentration of estimated use of coal in two subregions only. Because ALRI in children does not cause many years lost due to disability, however, COPD is responsible for a much larger portion of the total disability.

As shown in Table 18.18, five subregions account for nearly all deaths (94%) and DALYs (93%) attributable to indoor air pollution from solid fuel. The subregions with the largest numbers of DALYs, in descending order, are SEAR-D, WPR-B, AFR-E, AFR-D and EMR-D. When the subregions are ranked according to numbers of deaths, the relative positions of SEAR-D and WPR-B shift, because there are more deaths in SEAR-D in a younger age group (ALRI-related deaths in children) compared to WPR-B (mortality is dominated by COPD in adults).

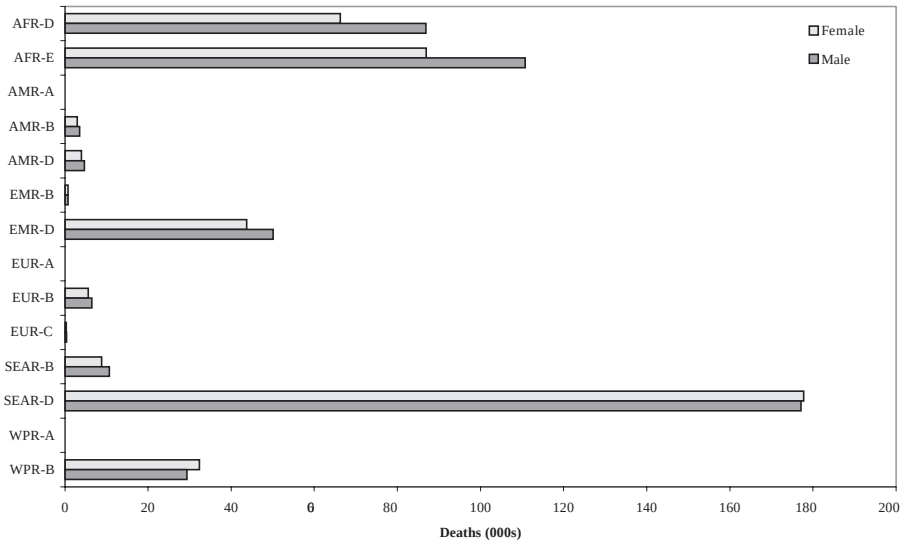
As shown in Figure 18.5, because of differences in baseline rates of disease, not exposure or risk from use of solid fuel, effects on mortality

Table 18.18 Burden of disease from use of solid fuel, 2000

	Deaths (000s)				*DALYs (000s)*			
Subregion	*ALRI*	*COPD*	*Lung cancer*	*All causes*	*ALRI*	*COPD*	*Lung cancer*	*All causes*
AFR-D	153	20	NA	173	5221	173	NA	5394
AFR-E	198	21	NA	219	6746	178	NA	6924
AMR-A	0	0	NA	1	1	6	NA	6
AMR-B	6	9	NA	16	291	153	NA	444
AMR-D	9	2	NA	10	314	16	NA	330
EMR-B	2	0	NA	2	59	5	NA	64
EMR-D	94	22	NA	116	3306	203	NA	3508
EUR-A	0	0	NA	0	0	0	NA	0
EUR-B	12	5	NA	17	417	60	NA	477
EUR-C	1	4	NA	4	22	44	NA	67
SEAR-B	19	17	NA	37	761	229	NA	990
SEAR-D	355	167	1	522	12506	1724	8	14237
WPR-A	0	0	NA	0	0	0	NA	0
WPR-B	62	426	15	503	2275	3662	160	6097
World	910	693	16	1619	31919	6453	168	38539

NA Not applicable.

Figure 18.5 Deaths from acute lower respiratory infection attributable to indoor smoke from use of solid fuels, 2000



attributable to ALRI are larger for males than females in AFR-D and AFR-E, similar in EMR-D and WPR-B, and greater for females in SEAR-D.

As shown in figure 18.6, the vast majority of attributable deaths from COPD and lung cancer appear to be experienced by the women of SEAR-D and WPR-B. This is partially because lung cancer deaths associated with solid fuel use were only estimated in these two subregions, due to lack of information on coal use in the other subregions. In addition, women appear to bear a higher proportion of the burden not only because they are likely to be more exposed, but because smoking attributable deaths (which are a higher proportion of male deaths) have been removed.

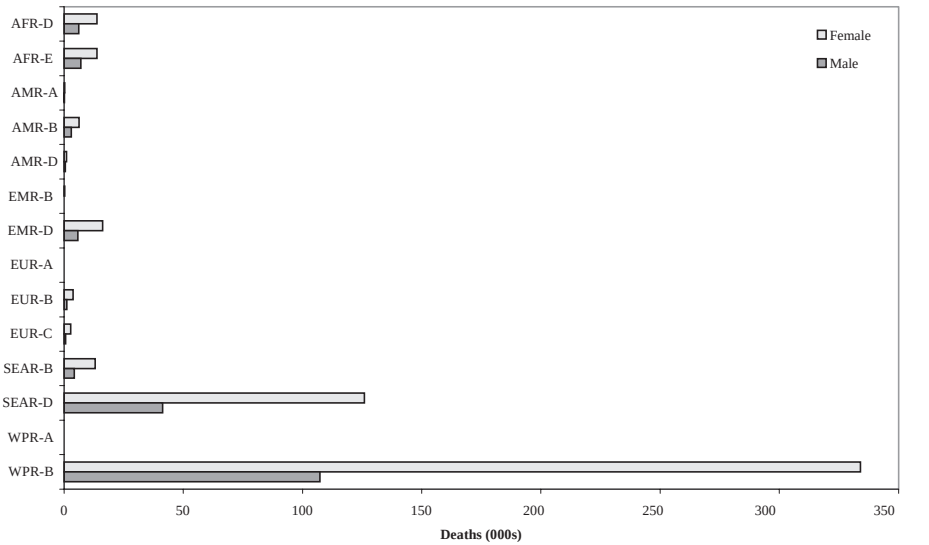
5. DISCUSSION

5.1 SOURCES OF UNCERTAINTY

Of a large number of sources of uncertainty, four major factors dominate these estimates.

- The choice of exposure variable, which, although necessary to match with current epidemiological studies, only roughly captures the population distribution of exposure and its variability in different populations.

Figure 18.6 Deaths from chronic obstructive pulmonary disease and lung cancer attributable to indoor smoke from use of solid fuels, 2000



- Distribution of the ventilation factor worldwide, i.e. what fractions of solid-fuel-using households do so in ways that vent some or all of the smoke outside and away from the householders.
- The different patterns of competing and confounding risks for ALRI in different circumstances, particularly those related to the severe forms affecting mortality.
- The relationship between the risks of indoor pollution and tobacco smoking, particularly for COPD and lung cancer in China where tobacco smoking is an important contributor (Liu et al. 1998).

5.2 Possible interventions

Although not included in the primary calculations here, as previously noted, there is growing evidence that other important health end-points can be attributed to exposure to indoor air pollution. Three of these, in particular, are of increasing concern worldwide: tuberculosis (because it is so closely related to the HIV epidemic); ischaemic heart disease (because of the shift in age and diet occurring in developing countries); and asthma (because of rising trends in diagnosed asthma in many parts of the world) (ISAAC 1998). There is some urgency that the associations of all potentially policy-sensitive risk factors (including use of solid household fuels) with these diseases be investigated.

There are four general categories of interventions that have been identified to reduce the health impacts of household use of solid fuel (Barnes et al. 1993; Ezzati and Kammen 2002; NCAER 2002; Smith and Desai 2002; Smith 1987, 1989).

- Behavioural changes to reduce exposure, for example, encouraging women to keep their young babies away from the fire.
- Changes in household ventilation, such as increasing the number of window openings in the kitchen, providing gaps between roof and wall, and moving the stove out of the living area.
- Improvements in stoves, either through venting by use of flues or hoods and/or improvements in stove combustion efficiency that reduce the emissions of toxic pollutants, nearly all of which are products of incomplete combustion.
- Shifts to higher-quality, low-emission liquid or gaseous fuels, such as kerosene and liquefied petroleum gas (which are based on petroleum) or biomass-based alcohol and biomass-based gaseous fuels derived either from biological processes (bio-gas) or thermochemical processing (producer gas).

Most research has focused on improvements in stoves and shifts to higher-quality, low-emission liquid or gaseous fuels; it seems that the efficacy of the interventions listed above generally increases as one moves down the list. The extent to which they can be successfully applied varies across different populations depending on income, housing, biomass availability, cultural factors and climate. It seems possible, however, that programmes can be designed to encourage many urban and peri-urban solid-fuel-using populations to move to using liquefied petroleum gas or kerosene, at lower incomes (i.e. sooner) than would occur without intervention. On the other hand, the poorest rural populations with nearly no cash income, but with access to wood and/or agricultural waste, are unlikely to move to clean fuels or use significantly improved stoves without large subsidies, which are usually not sustainable. There do seem to be large populations between these extremes, however, that can be targeted by efforts to introduce improved stoves. Although the fraction of improved-stove programmes that have succeeded is small, the total number of stoves successfully introduced is impressive because of the remarkable achievement of the Chinese programme, which has apparently been responsible for the introduction of nearly 200 million stoves since the early 1980s (Goldemberg et al. 2000; Smith 1993). More research and development work is needed, however, to learn how to successfully translate the lessons learned in China and elsewhere to other parts of the world in a sustainable cost-effective manner.

6. Exposure projections

The use of solid fuel will probably slowly decrease in absolute, as well as relative, terms, as economic development proceeds. This shift is occurring most rapidly in China and Latin America, at interim rates in south Asia, and slowest or not at all in sub-Saharan Africa (World Resources Institute 2000). Cooking outdoors, on the other hand, is likely to decrease with development, but as the number of separate kitchens may increase, it is not clear how exposures will change overall. Current trends in vented stoves are less certain outside China. The Indian national stove programme, for example, had mixed success (NCAER 2002) and was dismantled in 2002 (Mahapatra 2003). In China, however, nearly 90% of the rural population seems to have adopted higher-efficiency vented stoves in recent years.

Table 18.19 Use of solid fuel and exposure to its smoke: estimates for 2000 and predictions for 2010

	Estimated fuel use[a]		*Estimated exposure of adults*[b]	
Subregion	*2000*	*2010*	*2000*	*2010*
AFR-D	73.4	69.0	55.1	52.0
AFR-E	85.8	83.0	64.3	62.0
AMR-A	1.5	1.0	1.1	1.0
AMR-B	24.6	20.0	18.4	15.0
AMR-D	52.9	52.0	39.7	39.0
EMR-B	6.1	5.0	4.6	4.0
EMR-D	55.2	50.0	41.4	37.0
EUR-A	0.2	0.2	0.0	0.0
EUR-B	41.5	35.0	20.5	19.0
EUR-C	22.8	21.0	6.4	6.0
SEAR-B	66.5	62.0	49.9	46.0
SEAR-D	83.5	77.0	62.6	58.0
WPR-A	0.2	0.0	0.1	0.0
WPR-B[c]	78.1	70.0	41.8	23.0

[a] These projections only address changes in biomass use, i.e. for India and China, rates of coal use are not predicted to decline in the same manner. Indeed, recent trends in China indicate that coal is being substituted by gas in urban households, but is substituting for biomass in many rural households (Fridley et al. 2001).

[b] Children's exposures differ from adult exposures at present in that they are modified by a different ventilation factor, since adults experience the health effects of exposures that took place before improvements in ventilation occurred. In the future, child and adult exposures will converge.

[c] We assumed that the Chinese improved-stove programme would reach 90% penetration for biomass but that rates of coal use would not decrease (Goldemberg et al. 2000). When estimating exposure, the ventilation factor for China was therefore fixed at 0.25 for both adults and children, making the exposures of these two groups the same.

Some insight can be gleaned about the potential for reduction in exposure by application of the model of solid-fuel use employed in this chapter. Estimates of income growth and shift of the population from rural to urban areas have different impacts on use of solid fuels in different subregions. Economic growth and urbanization over the next 10 years, for example, might substantially reduce the fraction of households that use solid fuel in the subregions that currently have the largest burdens. We examined changes that might occur over a 10-year period in two major model parameters: GNP per capita and rural–urban population shift (World Bank 2001). Estimates based on changes in income and urbanization beyond 2010 would be highly unstable, since current trends are unlikely to be sustained over several decades. Countries for which data are lacking are assigned the global average values for GNP per capita (equivalent to a 1.3% annual growth rate) and global rate of urbanization (rural population decreases from around 58% to 51% of the total population. Among many other assumptions, of course, such an extrapolation supposes that the structure of the model remains valid over this period. Table 18.19 shows how predicted changes in GNP per capita and urbanization affect predictions of future household use of solid fuel and of future exposure in each subregion. The net impact of shifts in these factors seems to indicate that, globally, exposure to indoor smoke from use of solid fuel is likely to decrease. There are subregional variations in the pattern, however, with continuing large exposures in sub-Saharan Africa and south-east Asia (Indian subcontinent).

Acknowledgements

We appreciate the excellent comments of many anonymous reviewers; assistance with extracting information from Chinese-language literature from Linwei Tian; advice on meta-analysis from Daniel Mäusezahl, Jack Colford and the Berkeley Meta-analysis Group; and the patience and thoughtfulness of the book's editors, particularly Majid Ezzati. The authors would also like to recognize the millions of poor women and children around the world who are exposed daily to toxic indoor air pollution produced by smoke from solid fuel, as well as to a range of other risks. We hope that these results will help to generate recognition of the potential magnitude of this problem and spark commensurate efforts to ameliorate the situation.

Notes

1 Particulate matter, often abbreviated as PM, is categorized by size, specifically by aerodynamic diameter in microns (millionths of a meter or μm). For example, $PM_{2.5}$ refers to particulate matter with a diameter of less than 2.5 μm. In general, small particles are thought to be more damaging to health.

2 See preface for an explanation of this term.

3 Seven urban and three rural areas were omitted because of missing data or likely errors in the government statistical publications, which suggested improbable levels of energy consumption per household (i.e. in provincial households, average levels of consumption that were more than one standard deviation from the mean).

4 All analysis was done using SPSS Version 8.0 (SPSS Inc., USA) and STATA 7.0 (Stata Corporation, USA).

5 Cross-sectional studies report odds ratios rather than relative risks. These terms are used interchangeably in this chapter.

6 Two hospital-based case–control studies in India came to our attention too late for inclusion in the meta-analysis. In New Delhi, Broor et al. (2001) found an adjusted odds ratio of 2.5 (95% CI 1.5–4.2) for ALRI in children aged <5 years in homes not using liquefied petroleum gas. In Calcutta, Mahalanabis et al. (2002) found an adjusted odds ratio of 4.0 (95% CI 2.0–7.9) for pneumonia in children aged 2–35 months living in homes using solid fuels.

7 For further discussions, see reviews by Bruce et al. (2000), Chen et al. (1990) and Smith (1987).

8 For males, it did not seem appropriate to use the unadjusted estimate of risk, particularly when the adjusted estimates for both sexes were lower than either the unadjusted estimate for males only or the adjusted estimates for females only. Simple averaging of the risk chosen for males, 1.8, with the adjusted risk for females, 3.2, results in the combined mean risk of 2.5 observed when analyses included both sexes. The lower bound of the confidence interval was set at 1.0 (no effect) and the higher bound only at the unadjusted risk for males, 3.2.

References

Albalak R, Frisancho AR, Keeler GJ (1999) Domestic biomass fuel combustion and chronic bronchitis in two rural Bolivian villages. *Thorax*, **54**:1004–1008.

Armstrong JR, Campbell H (1991) Indoor air pollution exposure and lower respiratory infections in young Gambian children. *International Journal of Epidemiology*, **20**:424–429.

Awasthi S, Glick HA, Fletcher RH (1996) Effect of cooking fuels on respiratory diseases in preschool children in Lucknow, India. *American Journal of Tropical Medicine and Hygiene*, **55**:48–51.

Azizi BH, Zulkifli HI, Kasim S (1995) Indoor air pollution and asthma in hospitalized children in a tropical environment. *Journal of Asthma*, **32**:413–418.

Barker D (1997) Maternal nutrition, fetal nutrition, and disease in later life. *Nutrition*, **13**:807–813.

Barnes DF, Openshaw K, Smith KR, van der Plas R (1993) The design and diffusion of improved cookstoves. **8**:119–141.

Basu R, Samet J (1999) A review of the epidemiological evidence on health effects of nitrogen dioxide exposure from gas stoves. *Journal of Environmental Medicine*, **1**:173–187.

Behera D, Dash S, Yadav SP (1991) Carboxyhaemoglobin in women exposed to different cooking fuels. *Thorax*, **46**:344–346.

Bobak M (2000) Outdoor air pollution, low birth weight, and prematurity. *Environmental Health Perspectives*, **108**:173–176.

Boy E, Bruce N, Delgado H (2002) Birth weight and exposure to kitchen wood smoke during pregnancy in rural Guatemala. *Environmental Health Perspectives*, **110**:109–114.

Broor S, Pandey RM, Ghosh M et al. (2001) Risk factors for severe acute lower respiratory infection in under-five children. *Indian Pediatrics*, **38**:1361–1369.

Bruce N, Perez-Padilla R, Albalak R (2000) Indoor air pollution in developing countries: a major environmental and public health challenge. *Bulletin of the World Health Organization*, **78**:1078–1092.

Campbell H (1997) Indoor air pollution and acute lower respiratory infections in young Gambian children. *Health Bulletin*, **55**:20–31.

Campbell H, Armstrong JR, Byass P (1989) Indoor air pollution in developing countries and acute respiratory infection in children. *The Lancet*, **1**:1012.

Cerqueiro MC, Murtagh P, Halac A, Avila M, Weissenbacher M (1990) Epidemiologic risk factors for children with acute lower respiratory tract infection in Buenos Aires, Argentina: A matched case-control study. *Reviews of Infectious Diseases*, **12**:S1021–1028.

Chen BH, Hong CJ, Pandey MR, Smith KR (1990) Indoor air pollution in developing countries. *[World Health Statistics Quarterly] Rapport Trimestriel de Statistiques Sanitaires Mondiales*, **43**:127–138.

Collings DA, Sithole SD, Martin KS (1990) Indoor woodsmoke pollution causing lower respiratory disease in children. *Tropical Doctor*, **20**:151–155.

Dai XD, Lin CY, Sun XW, Shi YB, Lin YJ (1996) The etiology of lung cancer in nonsmoking females in Harbin, China. *Lung Cancer*, **14**:S85–91.

de Francisco A, Morris J, Hall AJ, Armstrong Schellenberg JR, Greenwood BM (1993) Risk factors for mortality from acute lower respiratory tract infections in young Gambian children. *International Journal of Epidemiology*, **22**:1174–1182.

Dennis RJ, Maldonado D, Norman S, Baena E, Martinez G (1996) Woodsmoke exposure and risk for obstructive airways disease among women. *Chest*, **109**:115–119.

Dhar SN, Pathania AGS (1991) Bronchitis due to biomass fuel burning in North India: "Guijjar lung" an extreme effect. *Seminars in Respiratory Medicine*, **12**: 69–74.

Døssing M, Khan J, al-Rabiah F (1994) Risk factors for chronic obstructive lung disease in Saudi Arabia. *Respiratory Medicine*, **88**:519–522.

Du E, Wang W, Zhang Q, Li Z, Wu Y, Zhao S (1988) Survey of indoor air quality and human health in rural Changchun, Jilin province, China. *Journal of Norman Bethune University of Medical Science*, **14**:88–91.

Du YX, Ou XL (1990) Indoor air pollution and women's lung cancer. *Proceedings of the Fifth International Conference on Indoor Air Quality and Climate, Toronto*, **1**:59–64.

Du YX, Cha Q, Chen XW et al. (1996) An epidemiological study of risk factors for lung cancer in Guangzhou, China. *Lung Cancer,* **14**:S9–37.

Dutt D, Srinivasa DK, Rotti SB, Sahai A, Konar D (1996) Effect of indoor air pollution on the respiratory system of women using different fuels for cooking in an urban slum of Pondicherry. *National Medical Journal of India*, **9**:113–117.

Ellegard A (1996) Cooking fuel smoke and respiratory symptoms among women in low-income areas in Maputo. *Environmental Health Perspectives*, **104**: 980–985.

Ezzati M (2002) Indoor air pollution from biomass stoves as a risk factor for acute respiratory infections in Kenya. *Proceedings of the Ninth International Conference on Indoor Air Quality and Climate, Monterey*, **4**:970–975.

Ezzati M, Kammen D (2001) Indoor air pollution from biomass combustion and acute respiratory infections in Kenya: an exposure-response study. *The Lancet*, **358**:619–624.

Ezzati M, Kammen DM (2002) Household energy, indoor air pollution and health in developing countries: knowledge base for effective interventions. *Annual Review of Energy and the Environment*, **27**:233–270.

FAO (1997a) *Review of wood energy data in RWEDP member countries*. (FAO Field Document No. 47.) Food and Agriculture Organization of the United Nations, Regional Wood Energy Development Program (RWEDP), Bangkok.

FAO (1997b) *The role of wood energy in Asia*. (FOPW/97/2.) Wood energy for today and tomorrow (WETT), Rome.

Finkelman RB, Belkin HE, Zheng B (1999) Health impacts of domestic coal use in China. *Proceedings of the National Academy of Sciences of the United States of America*, **96**:3427–3431.

Franco EL, Kowalski LP, Oliveira BV et al. (1989) Risk factors for oral cancer in Brazil: a case-control study. *International Journal of Cancer*, **43**:992–1000.

Fridley DG, Sinton JE, Lewis JI, Zhoui F, Li J (2001) *China energy databook—5th edition*. LBL-47832. Lawrence Berkeley National Laboratory and Beijing Energy Research Institute, Berkeley.

Gao YT, Blot WJ, Zheng W et al. (1987) Lung cancer among Chinese women. *International Journal of Cancer*, **40**:604–609.

García-Marcos L, Guillén JJ, Dinwiddie R, Guillén A, Barbero P (1999) The relative importance of socio-economic status, parental smoking and air pollution (SO2) on asthma symptoms, spirometry and bronchodilator response in 11-year-old children. *Pediatric Allergy and Immunology*, **10**:96–100.

Gold JA, Jagirdar J, Hay JG et al. (2000) Hut lung. A domestically acquired particulate lung disease. *Medicine*, **79**:310–317.

Goldemberg J, Reddy AKN, Smith KR, Williams RH (2000) Rural energy in developing countries. In: *World energy assessment*. Goldemberg JO, ed. United Nations Development Programme, New York.

Government of Botswana (1991) *Population and housing census*. Central Statistical Office, Gaborone.

Government of Brazil (1991) *Censo demografico: 1991; familias e domicilios. [Demographic census: 1991: families and households]* Instituto Brasileiro do Geografica e Estadistica (IBGE) [Brasilian Institute of National Statistics and Geography], Rio de Janeiro.

Government of Ecuador (1990a) *V censo de poblacion y iv censo de vivienda [Census of population and households]*. Instituto Nacional de Estadistica y Censo [National Institute of Statistics and Census], Ecuador.

Government of Ecuador (1990b) *V censo de poblacion, iv censo de vivienda, 1990. vivienda [Census of population and households, 1990]*. Instituto Nacional de Estatistica y Censo [National Institute of Statistics and Census], Ecuador.

Government of Ethiopia (1998) *1994 population and housing census of Ethiopia: results at the country level*. Volume 1. Statistical Office, Addis Ababa.

Government of India (1991a) *Census of India*. Office of the Registrar General, New Delhi.

Government of India (1991b) *Census of India: 1991*. Office of the Registrar General, New Delhi.

Government of Indonesia (1995) *Indonesia energy balance*. Central Bureau of Statistics, Jakarta.

Government of Indonesia (1996a) *Indonesia energy balance 1992–1996*. Central Bureau of Statistics, Jakarta.

Government of Indonesia (1996b) *Indonesia energy balance 1992–1996*. Central Bureau of Statistics, Jakarta.

Government of Mexico (1990a) *Censo general de poblacion y vivienda [General population and housing census]*. Instituto Nacional de Estadistica, Geografica e Informatica [National Institute of Statistics, Geographics and Informatics], Mexico.

Government of Mexico (1990b) *Censo general de poblacion y vivienda, 1990 [General population and household census, 1990]*. Instituto Nacional de Estatistica, Geografica e Informatica [National Institute of Statistics, Geographics and Informatics], Mexico.

Government of Nigeria (1990) *Social statistics in Nigeria, 1990*. Federal Office of Statistics, Lagos.

Government of Pakistan (1997) *Pakistan energy yearbook: 1996*. Hydrocarbon Development Institute of Pakistan, Ministry of Petroleum and Natural Resources, Islamabad.

Government of China (1996) *China energy statistical yearbook: 1991–1996*. Department of Industrial and Transportation Statistics, Chinese Statistical Publishing House, and Department of Industrial and Transportation Statistics, China.

Government of Zimbabwe (1992) *Census, Zimbabwe national report*. CSO, Harare.

Gupta BN, Mathur N (1997) A study of household environmental risk factors pertaining to respiratory diseases. *Energy Environment Monitor*, **13**: 61–67.

Huang Z-B (1999) A study on the risk factors and population attributable risk for primary lung cancer. *Journal of Guangxi Medical University* [in Chinese], **16**:447–450.

International Energy Agency (1999) *Energy statistics of non-OECD countries 1996–1997,* The Organization for Economic Cooperation and Development (OECD), Paris.

International Energy Agency (2002) *World energy outlook*. International Energy Agency, Paris.

ISAAC (1998) Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The international study of asthma and allergies in childhood (ISAAC) steering committee. *The Lancet*, **351**:1225–1232.

Johnson AW, Aderele WI (1992) The association of household pollutants and socio-economic risk factors with the short-term outcome of acute lower respiratory infections in hospitalized pre-school Nigerian children. *Annals of Tropical Paediatrics*, **12**:421–432.

Johnson DL, Swank PR, Baldwin CD, McCormick D (1999) Adult smoking in the home environment and children's IQ. *Psychological Reports*, **84**:149–154.

Jones G, Skeketee RW, Black RE, Bhutta ZA, Morris SS (2003) How many child deaths can we prevent this year? *The Lancet*, **362**:65–71.

Ko Y (2000) Chinese food cooking and lung cancer in women nonsmokers. *American Journal of Epidemiology*, **151**:140–147.

Ko YC, Lee CH, Chen MJ et al. (1997) Risk factors for primary lung cancer among non-smoking women in Taiwan. *International Journal of Epidemiology*, **26**:24–31.

Kossove D (1982) Smoke-filled rooms and lower respiratory disease in infants. *South Africa Medical Journal*, **61**:622–624.

Lan Q, Chapman RS, Schreinemachers DM, Tian L, He X (2002) Household stove improvement and risk of lung cancer in Xuanwei, China. *Journal of the National Cancer Institute*, **94**:826–835.

Lei YX, Cai WC, Chen YZ, Du YX (1996) Some lifestyle factors in human lung cancer: a case-control study of 792 lung cancer cases. *Lung Cancer*, **14**: S121–136.

Lin Z-G (1995) *[China's climate and its extremes] Zhong Guo De Qi Hou Ji Qi Ji Zhi.* The Commercial Press of China, Beijing.

Liu BQ, Peto R, Chen ZM et al. (1998) Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *British Medical Journal*, **317**:1411–1422.

Liu Q, Sasco AJ, Riboli E, Hu MX (1993) Indoor air pollution and lung cancer in Guangzhou, People's Republic of China. *American Journal of Epidemiology*, **137**:145–154.

Liu ZY, He XZ, Chapman RS (1991) Smoking and other risk factors for lung cancer in Xuanwei, China. *International Journal of Epidemiology,* **20**:26–31.

Loomis D, Castillejos M, Gold DR, McDonnell W, Borja-Aburto VH (1999) Air pollution and infant mortality in Mexico City. *Epidemiology*, **10**: 118–123.

Luo RX, Wu B, Yi YN, Huang ZW, Lin RT (1996) Indoor burning coal air pollution and lung cancer—a case-control study in Fuzhou, China. *Lung Cancer*, **14**:S113–119.

Mahalanabis D, Gupta S, Paul D et al. (2002) Risk factors for pneumonia in infants and young children and the role of solid fuel for cooking: a case–control study. *Epidemiology and Infection*, **29**:65–71.

Mahapatra R (2003) Up in smoke: National programme on improved chulhas put on inexplicable hold. *Down To Earth*, **11**:3–5.

Malik SK (1985) Exposure to domestic cooking fuels and chronic bronchitis. *Indian Journal of Chest Diseases and Allied Sciences*, **27**:171–174.

Mavalankar DV, Trivedi CR, Grah RH (1991) Levels and risk factors for perinatal mortality in Ahmedabad, India. *Bulletin of the World Health Organization*, **69**:435–442.

Mehta S, Smith KR (2002) *Exposure atlas for household energy and indoor air pollution modeling component: predicting household pollution levels*. World Bank Energy Sector Management Assistance Programme, Washington, DC.

Menezes AM, Victora CG, Rigatto M (1994) Prevalence and risk factors for chronic bronchitis in Pelotas, RS, Brazil: a population-based study. *Thorax*, **49**:1217–1221.

Mishra VK, Retherford RD, Smith KR (1999a) Biomass cooking fuels and prevalence of blindness in India. *Journal of Environmental Medicine*, **1**: 189–199.

Mishra VK, Retherford RD, Smith KR (1999b) Biomass cooking fuels and prevalence of tuberculosis in India. *International Journal of Infectious Diseases*, **3**:119–129.

Mohamed N, Ng'ang'a L, Odhiambo J, Nyamwaya J, Menzies R (1995) Home environment and asthma in Kenyan schoolchildren: a case-control study. *Thorax*, **50**:74–78.

Mohan M, Sperduto RD, Angra SK et al. (1989) India-US case-control study of age-related cataracts. India-US case-control study group. *Archives of Ophthalmology*, **107**:670–676.

Morris K, Morgenlander M, Coulehan JL, Gahagen S, Arena VC, Morganlander M (1990) Wood-burning stoves and lower respiratory tract infection in American Indian children. *American Journal of Diseases of Children*, **144**: 105–108.

Mtango FD, Neuvians D, Broome CV, Hightower AW, Pio A (1992) Risk factors for deaths in children under 5 years old in Bagamoyo district, Tanzania. *Tropical Medicine and Parasitology*, **43**:229–233.

Naeher LP, Leaderer BP, Smith KR (2000a) Particulate matter and carbon monoxide in highland Guatemala: indoor and outdoor levels from traditional and improved wood stoves and gas stoves. *Indoor Air*, **10**:200–205.

Naeher LP, Smith KR, Leaderer BP, Mage D, Grajeda R (2000b) Indoor and outdoor PM2.5 and CO in high- and low-density Guatemalan villages. *Journal of Exposure Analysis and Environmental Epidemiology*, **10**:544–551.

NCAER (2002) *Household survey to evaluate national program for improved Chula*. National Center for Applied Economic Research, Delhi.

NFHS (1995) *National family health survey (MCH and family planning): India, 1992–93*. International Institute for Population Sciences, Bombay.

NHLBI/WHO (2001) *Global initiative for chronic obstructive lung disease*. National Heart Lung Blood Institute/World Health Organization, Bethesda, MA.

Norboo T, Angchuk PT, Yahya M et al. (1991) Silicosis in a Himalayan village population: role of environmental dust. *Thorax*, **46**:341–343.

O'Dempsey T, McArdle TF, Morris J et al. (1996) A study of risk factors for pneumococcal disease among children in a rural area of West Africa. *International Journal of Epidemiology*, **25**:885–893.

Padmavati S, Pathak SN (1959) Chronic *cor pulmonale* in Delhi. *Circulation*, **20**:343–352.

Pandey MR (1984a) Domestic smoke pollution and chronic bronchitis in a rural community of the hill region of Nepal. *Thorax*, **39**:337–339.

Pandey MR (1984b) Prevalence of chronic bronchitis in a rural community of the hill region of Nepal. *Thorax*, **39**:331–336.

Pandey MR, Regmi HN, Neupane RP, Gautam A, Bhandari DP (1985) Domestic smoke pollution and respiratory function in rural Nepal. *Tokai Journal of Experimental and Clinical Medicine*, **10**:471–81.

Pandey MR, Basnyat B, Neupane RP (1988) *Chronic bronchitis and cor pulmonale in Nepal*. Mrigendra Medical Trust, Kathmandu.

Pandey MR, Boleij JS, Smith KR, Wafula EM (1989a) Indoor air pollution in developing countries and acute respiratory infection in children. *The Lancet*, **1**:427–429.

Pandey MR, Neupane RP, Gautam A, Shrestha I (1989b) Domestic smoke pollution and acute respiratory infections in a rural community of the hill region of Nepal. *Environment International*, **15**:337–340.

Pereira LA, Loomis D, Conceição GM et al. (1998) Association between air pollution and intrauterine mortality in São Paulo, Brazil. *Environmental Health Perspectives*, **106**:325–329.

Perez-Padilla R, Perez-Guzman C, Baez-Saldana R, Torres-Cruz A (2001) Cooking with biomass stoves and tuberculosis: a case control study. *International Journal of Tuberculosis and Lung Disease*, **5**:1–7.

Perez-Padilla R, Regalado J, Vedal S et al. (1996) Exposure to biomass smoke and chronic airway disease in Mexican women a case-control study. *American Journal of Respiratory Critical Care Medicine*, **154**:701–706.

Pintos J, Franco EL, Kowalski LP, Oliveira BV, Curado MP (1998) Use of wood stoves and risk of cancers of the upper aero-digestive tract: a case-control study. *International Journal of Epidemiology*, **27**:936–940.

Ponka A, Virtanen M (1996) Low-level air pollution and hospital admissions for cardiac and cerebrovascular diseases in Helsinki. *American Journal of Public Health*, **86**:1273–1280.

Pope CA 3rd, Schwartz J, Ransom MR (1992) Daily mortality and PM10 pollution in Utah valley. *Archives of Environmental Health*, **47**:211–217.

Prüss A, Mariotti S (2000) Preventing trachoma through environmental sanitation: a review of the evidence base. *Bulletin of the World Health Organization*, **78**:258–265.

Qureshi KA (1994) Domestic smoke pollution and prevalence of chronic bronchitis/asthma in a rural area of Kashmir. *Indian Journal of Chest Diseases and Allied Sciences*, **36**:61–72.

Ramage J, Roggli V, Bell D, Piantadosi C (1988) Interstitial lung disease and domestic wood burning. *American Review of Respiratory Disease*, **137**: 1229–1232.

Ramakrishna J, Durgaprasad MB, Smith KR (1989) Cooking in India: the impact of improved stoves on indoor air quality. *Environment International*, **15**: 341–352.

Rao C, Qin C, Robison W, Zigler J (1995) Effect of smoke condensate on the physiological integrity and morphology of organ cultured rate lenses. *Current Eye Research*, **14**:295–301.

Ritz B, Yu F (1999) The effect of ambient carbon monoxide on low birth weight among children born in southern California between 1989 and 1993. *Environmental Health Perspectives*, **107**:17–25.

Robin LF, Less PS, Winget M et al. (1996) Wood-burning stoves and lower respiratory illnesses in Navajo children. *Pediatric Infectious Disease Journal*, **15**:859–865.

Rossi G, Vigotti MA, Zanobetti A, Repetto F, Gianelle V, Schwartz J (1999) Air pollution and cause-specific mortality in Milan, Italy, 1980–1989. *Archives of Environmental Health*, **54**:158–164.

Sahlu T, Larson C (1992) The prevalence and environmental risk factors for moderate and severe trachoma in southern Ethiopia. *Journal of Tropical Medicine and Hygiene*, **95**:36–41.

Saiyed HN, Sharma YK, Sadhu HG et al. (1991) Non-occupational pneumoconiosis at high altitude villages in central Ladakh. *British Journal of Industrial Medicine,* **48**:825–829.

Saksena S, Prasad R, Pal RC, Joshi V (1992) Patterns of daily exposure to TSP and CO in the Garhwal, Himalaya. *Atmospheric Environment,* **26**A: 2125–2134.

Sandoval J, Salas J, Martinez-Guerra ML et al. (1993) Pulmonary arterial hypertension and *cor pulmonale* associated with chronic domestic wood smoke inhalation. *Chest*, **103**:12–20.

Schwartz J (1993) Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology*, **137**:1136–1147.

Schwartz J, Dockery DW (1992) Increased mortality in Philadelphia associated with daily air pollution concentrations. *American Review of Respiratory Disease*, **145**:600–604.

Schwartz J, Morris R (1995) Air pollution and hospital admissions for cardiovascular disease in Detroit, Michigan. *American Journal of Epidemiology*, **142**:23–35.

Scram R (1999) Impact of air pollution on reproductive health. *Environmental Health Perspectives*, **107**:A542–A543.

Shah N, Ramankutty V, Premila PG, Sathy N (1994) Risk factors for severe pneumonia in children in south Kerala: a hospital-based case-control study. *Journal of Tropical Pediatrics*, **40**:201–206.

Shalini V, Lothra M, Srinivas L (1994) Oxidative damage to the eye lens caused by cigarette smoke and fuel smoke condensates. *Indian Journal of Biochemistry and Biophysics*, **31**:261–266.

Shen XB, Wang GX, Huang YZ, Xiang LS, Wang XH (1996) Analysis and estimates of attributable risk factors for lung cancer in Nanjing, China. *Lung Cancer*, **14**:S107–112.

Sinton JE, Smith KR, Hu H, Liu J (1995) *Indoor air pollution database for China*. (WHO/EHG/95.8.) World Health Organization, United Nations Environment Programme, Geneva.

Smith K, Desai M (2002) The contribution of global environmental factors to ill-health. In: *Environmental change, climate, and health: issues and research methods*. Martens P, McMichael A, eds. Cambridge University Press, Cambridge.

Smith KR (1987) *Biofuels, air pollution, and health: a global review.* Plenum, New York.

Smith KR (1989) The dialectics of improved stoves. *Economic and Political Weekly*, **24**:517–522.

Smith KR (1993) Fuel combustion, air pollution exposure, and health: the situation in developing countries. *Annual Review of Energy and the Environment*, **18**:529–566.

Smith KR (2000) National burden of disease in India from indoor air pollution. *Proceedings of the National Academy of Sciences of the United States of America*, **97**:13286–13293.

Smith KR, Gu SH, Huang K, Qiu DX (1993) 100 million improved stoves in China: how was it done? *World Development*, **21**:941–961.

Smith KR, Liu Y (1994) Indoor air pollution in developing countries. In: *Epidemiology of lung cancer*. Samet J, ed. Dekker, New York.

Smith KR, Mehta S (2003) Burden of disease from indoor air pollution in developing countries: comparison of estimates. *International Journal of Hygiene and Environmental Health*, **206**:279–289.

Smith KR, Samet JM, Romieu I, Bruce N (2000) Indoor air pollution in developing countries and acute lower respiratory infections in children. *Thorax*, **55**:518–532.

Sobue T (1990) Association of indoor air pollution and lifestyle with lung cancer in Osaka, Japan. *International Journal of Epidemiology*, **19**:S62–66.

Spengler JD, Samet JM, McCarthy JF (2001) *Indoor air quality handbook,* McGraw-Hill, New York.

Steenland K, Sieber K, Etzel RA, Pechacek T, Maurer K (1998) Exposure to environmental tobacco smoke and risk factors for heart disease among never smokers in the third national health and nutrition examination survey. *American Journal of Epidemiology*, **147**:932–939.

Strachan DP, Cook DG (1998) Health effects of passive smoking. 6. Parental smoking and childhood asthma: longitudinal and case-control studies. *Thorax*, **53**:204–212.

Taylor H, West S (1989) Hygiene factors and increased risk of trachoma in central Tanzania. *Archives of Ophthalmology*, **107**:1821–1825.

Thomas P, Zelikoff J (1999) Air pollutants: moderators of pulmonary host resistance against infection. In: *Air pollution and health*. Holgate ST et al., eds. Academic Press, San Diego, CA.

UN (1993) *Energy statistics yearbook,* United Nations, New York.

UN (1998) *World population prospects*. United Nations, New York.

UNCHS (1996) *An urbanizing world: global report on human settlements*. United Nations Centre on Human Settlements (HABITAT), Oxford University Press, New York.

Velema JP, Ferrera A, Figueroa M et al. (2002) Burning wood in the kitchen increases the risk of cervical neoplasia in HPV-infected women in Honduras. *International Journal of Cancer*, **97**:536–541.

Verma BK, Thakur DK (1995) Effect of stressful environmental factors upon neonatal immune system. *Central European Journal of Public Health*, **3**:25–29.

Victora C, Fuchs S, Flores J, Fonseca W, Kirkwood B (1994) Risk factors for pneumonia among children in a Brazilian metropolitan area. *Pediatrics*, **93**:977–985.

Victora CG, Kirkwood BR, Ashworth A et al. (1999) Potential interventions for the prevention of childhood pneumonia in developing countries: improving nutrition. *American Journal of Clinical Nutrition*, **70**:309–320.

Walsh J (1993) Maternal and perinatal health. In: *Disease control priorities in developing countries*. Jamison D, ed. Oxford University Press, Oxford.

Wang TJ, Zhou BS, Shi JP (1996) Lung cancer in nonsmoking Chinese women: a case-control study. *Lung Cancer*, **14**:S93–98.

Wang X, Ding H, Ryan L, Xu X (1997) Association between air pollution and low birth weight: a community-based study. *Environmental Health Perspectives*, **105**:514–520.

West S (1992) Does smoke get in your eyes? *Journal of the American Medical Association*, **268**:1025–1026.

West S, Lynch M (1989) Water availability and trachoma. *Bulletin of the World Health Organization,* **67**:71–75.

Windham GC, Eaton A, Hopkins B (1999a) Evidence for an association between environmental tobacco smoke exposure and birthweight: a meta-analysis and new data. *Paediatric and Perinatal Epidemiology*, **13**:35–57.

Windham GC, Von Behren J, Waller K, Fenster L (1999b) Exposure to environmental and mainstream tobacco smoke and risk of spontaneous abortion. *American Journal of Epidemiology*, **149**:243–247.

Woodruff TJ, Grillo J, Schoendorf KG (1997) The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. *Environmental Health Perspectives*, **105**:608–612.

World Bank (1988) *Congo: issues and options in the energy sector*. 6420-COB. The World Bank Energy Sector Management Assistance Programme (ESMAP), Washington, DC.

World Bank (1998) *Chad household energy project staff appraisal report*. 17780-CD. The World Bank Energy Sector Management Assistance Programme (ESMAP), Washington, DC. (Also available at http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/1999/09/17/000009265_3980624143048/Rendered/PDF/multi_page.pdf.)

World Bank (2000) *Africa database CD-ROM*. World Bank, Washington, DC.

World Bank (2001) *World development indicators,* World Bank, Washington, DC.

World Energy Council (1999) *The challenge of rural energy poverty in developing countries,* World Energy Council, London.

World Resources Institute (2000) *World resources institute world resources report—2000–2001,* Oxford University Press, New York.

World Resources Institute (2003) *Earth-trends: the environmental information portal*. Available on the Internet at: http://earthtrends.wri.org/

Wu AH, Henderson BE, Pike MC, Yu MC (1985) Smoking and other risk factors for lung cancer in women. *Journal of the National Cancer Institute*, **74**:747–751.

Wu Y-L, Cao K-j, Ma G-S et al. (1999) A case-control study of the risk factors of male lung cancer in Guangzhou. *Cancer* [in Chinese], **18**:535–537.

Wu-Williams AH, Dai XD, Blot W et al. (1990) Lung cancer among women in north-east China. *British Journal of Cancer*, **62**:982–987.

Xu X, Ding H, Wang X (1995) Acute effects of total suspended particles and sulfur dioxides on preterm delivery: a community-based cohort study. *Archives of Environmental Health*, **50**:407–415.

Xu X, Niu T, Christian D (1996a) Occupational and environmental risk factors for asthma in rural communities in China. *International Journal of Occupational and Environmental Health*, **2**:172–176.

Xu ZY, Brown L, Pan GW et al. (1996b) Lifestyle, environmental pollution and lung cancer in cities of Liaoning in northeastern China. *Lung Cancer*, **14**: S149–160.

Yang X and Li YB (1990) A preliminary study of the impact of indoor air pollution on the health of the dwellers of Wuhan City. *Proceedings of the Fifth International Conference on Indoor Air Quality and Climate* (Toronto), **1**:83–88.

Zelikoff J (1994) Woodsmoke emissions: effects on host pulmonary immune defence. *CIAR Currents,* **1**:3.

Zelikoff JT, Chen LC, Cohen MD et al. (2003) Effects of inhaled ambient particulate matter on pulmonary antimicrobial immune defense. *Inhalation Toxicology*, **15**:131–150.

Zhang J, Smith KR (2003) Indoor air pollution: a global health concern. *British Medical Bulletin*, **67**:209–225.

Zhang J, Smith KR, Ma Y et al. (2000) Greenhouse gases and other airborne pollutants from household stoves in China: a database for emission factors. *Atmospheric Environment*, **34**:4537–4549.

Zhong L, Goldberg MS, Gao YT, Jin F (1999a) A case-control study of lung cancer and environmental tobacco smoke among nonsmoking women living in Shanghai, China. *Cancer Causes and Control*, **10**:607–616.

Zhong L, Goldberg MS, Gao YT, Jin F (1999b) Lung cancer and indoor air pollution arising from Chinese-style cooking among nonsmoking women living in Shanghai, China. *Epidemiology*, **10**:488–494.

Zodpey S, Ughade S (1999) Exposure to cheaper cooking fuels and risk of age-related cataract in women. *Indian Journal of Occupational and Environmental Medicine*, **3**:159–161.

Chapter 19

Lead exposure

Annette Prüss-Üstün, Lorna Fewtrell,
Philip J. Landrigan and José Luis Ayuso-Mateos

Summary

Exposure to lead causes a number of diseases, including mild mental retardation resulting from loss of IQ points, as well as increased blood pressure, anaemia, and gastrointestinal effects. Several other disease outcomes have been associated with exposure to lead, but evidence is considered insufficient at this time for a quantitative assessment of their impact on health to be made here.

The exposure variable used was the population distribution of blood-lead concentrations. We compiled data on concentrations of lead in blood from general population samples in countries around the world, as reported in more than 700 published studies. Only recent studies (published in or after 1995) were considered, because of changes in lead exposure that have taken place since the 1970s, mainly as a result of the implementation of lead reduction programmes (e.g. the phasing out of leaded petrol). Where current data were not available, we applied an adjustment of 39% reduction in blood-lead concentrations to allow for the effects of implementation of five-year lead reduction programmes. For countries for which data were not available, exposure reduction due to existing lead reduction programmes was accounted for by extrapolation.

Blood-lead concentrations of about 0.016 μg/dl have been measured in pre-industrial humans, indicating that the contribution of natural sources of lead to human exposure is minimal. Estimates published recently suggest that the theoretical-minimum-risk of health effects may occur at blood-lead concentrations as low as 0–1 μg/dl.

The association of increased blood-lead concentrations with loss of IQ points has been described in a meta-analysis by Schwartz (1994). Hazards for blood lead and blood pressure were from a meta-analysis by Schwartz (1994) and a published analysis of data from the second National Health and Nutrition Examination Survey (NHANES II).

Hazards for anaemia and gastrointestinal effects were based on a large review of toxicological and epidemiological data (ATSDR 1999). Based on the results of a study by Schwartz et al. (1990), as a consequence of individual variation we considered that only 20% of the people with blood-lead concentrations above those indicated by the Agency for Toxic Substances and Disease Registry (ATSDR) would actually develop symptoms.

The number of people with mild mental retardation as a result of IQ loss was determined on the basis of a standardized intelligence curve. To account for the higher prevalence of other mild mental retardation risk factors (e.g. malnutrition) in some subregions,[1] the prevalence of lead-induced mild mental retardation was adjusted for the known ratio of mental retardation caused by other factors. A number of health outcomes and social consequences of lead exposure (e.g. increased risk of violence and drug abuse) could not be quantified owing to insufficient evidence on hazard size.

In 2000, an estimated 120 million people around the world had blood-lead concentrations of between 5 and 10 μg/dl, and about the same number had concentrations of >10 μg/dl. Forty per cent of all children had blood-lead concentrations of >5 μg/dl and half of these children had blood-lead concentrations of >10 μg/dl; of these children, 97% were living in developing countries. The burden of disease caused by mild mental retardation attributable to exposure to lead resulted in 9.8 million disability-adjusted life years (DALYs), and the burden from cardiovascular diseases caused by elevated blood pressure resulted in 229 000 premature deaths and 3.1 million DALYs. In total, these two outcomes alone account for about 0.9% of the global burden of disease. Several health outcomes resulting from exposure to lead could not be quantified in this analysis, in particular, increased delinquent behaviour and its impact on injuries. Health impacts from anaemia and gastrointestinal effects caused by exposure to lead were relatively small. People affected by exposure to lead were concentrated mainly in developing countries. The burden of disease associated with exposure to lead could be virtually eliminated through interventions that have proven successful in developed countries, most importantly, the removal of lead from petrol.

1. Introduction

The toxic nature of lead has been recognized for millennia, with the earliest published reports dating back to 2000 BC (Needleman 1999). However, the range of health effects that exposure to lead can cause and the low concentrations of lead in blood at which these effects can occur is only now being fully appreciated. It is now understood that lead is toxic, especially to children, at levels that were previously thought to be safe.

Lead, due to its multiplicity of uses (e.g. leaded petrol, lead in paints, ceramics, food cans, make-up, traditional remedies, batteries), is present

in air, dust, soil and water to varying degrees. Each of these media can act as a route of human exposure, through ingestion or inhalation and, to a small degree for organic lead compounds, dermal absorption. Human exposure can be assessed directly, through body burden measurements (lead in blood, teeth or bone) or indirectly, by measuring levels of lead in the environment (air, dust, food or water).

Multiple health effects have been associated with lead exposure, including systemic effects (e.g. gastrointestinal effects, anaemia, hypertension and hearing loss), effects on the nervous system (e.g. on behaviour and cognition), on development, and on the reproductive system, as well as genotoxicity, carcinogenicity and social effects (ATSDR 1999). The strength of evidence supporting the association of these health effects with exposure to lead varies, and not all of these effects have been investigated sufficiently to permit quantification of their consequences in terms of disease burden.

1.1 Risk factor definition

In this analysis, exposure was characterized by the population mean and the population distribution of blood-lead concentrations. Occupational exposures or "hot spots" (i.e. areas of local relevance where lead levels are unusually high, such as around smelters) were excluded, unless they were assessed within general population samples.

1.2 Theoretical-minimum-risk exposure distribution

The definition of an elevated concentration of lead in the blood according to the Centers for Disease Control and Prevention (CDC 1991) is 10 μg/dl. However, evidence indicating that some health effects can occur below this threshold is accumulating. Recent analyses suggest that health effects may become apparent at concentrations of <5 μg/dl (Lanphear 2000) and, indeed, that no evidence exists for a threshold, even at 1 μg/dl (Schwartz 1994). For the purpose of this analysis, the concentration of blood-lead incurring the lowest population risk was considered to be 0–1 μg/dl, in the absence of scientific consensus and pending further investigation. The measurement of blood-lead concentrations in pre-industrial humans has shown that the contribution of natural sources of lead to human exposure is small; Flegal and Smith (1992) have estimated that pre-industrial humans had blood-lead concentrations of only 0.016 μg/dl.

2. Estimating risk factor levels

2.1 Choice of exposure variable

The concentration of lead in blood was chosen as the measure of exposure of the population because:

- it is an objective physiological measure, which can be measured accurately;
- it is directly related to health outcome and can be expected to reflect exposure more closely than estimates derived from measurement of the concentration of lead in air, soil, dust or food; and
- it is the only parameter for which measurements are available from many parts of the world, thus making it preferable to other physiological measures, such as the concentration of lead in bone.

Blood-lead concentrations are indicative of recent lead exposure (within the preceding few weeks) rather than of cumulative long-term exposure. However, as exposure to lead varies relatively little over a time span of a year (WHO 1998), this measure can also be a sign of longer-term exposure.

2.2 Data on blood-lead concentrations

Data sources

Exposure data were obtained from studies identified principally through Medline searches. The primary database used was compiled by the CDC (1999) and contained exposure data from over 700 articles published between 1965 and 1998. The search strategy used was the term "lead" paired with any of the following keywords: newborn, cord, adult, pregnancy, occupation, blood, tooth, hair, milk, placenta, urine, smelter, ceramics, pottery, petrol, cosmetics, kohl, surma, medicine, neurological deficit, cognitive function, pregnancy outcome, fertility and birth defect. Initial queries were followed up using the "related–articles" option of Medline and further searches on the basis of author. The reference list of each relevant article was also examined. Additionally, the authors conducted searches of the databases LILACS (Latin American and Caribbean Information System of Health Sciences), IMEMR (Index Medicus of the World Health Organization's [WHO] Regional Office for the Eastern Mediterranean—EMRO) and African Index Medicus, using the same keywords. Medline was also re-consulted to ensure coverage of the most recent publications (up to the end of 2000).

Compilation and presentation of data

Blood-lead concentrations in the population generally have a log-normal distribution, as reported in numerous countries and populations (Al-Saleh et al. 1999; Baghurst et al. 1995; Brody et al. 1994; CDC 2001; Harlan et al. 1985; Hense et al. 1992; Molla et al. 1997; Pocock et al. 1988; Schwartz 1991; Tong et al. 1998). Geometric means with standard deviations were therefore compiled to represent population exposures.

A few studies reported small differences (typically <10%) between blood-lead concentrations in males and females (e.g. CDC 2001; Nielsen et al. 1998; Omokhodion 1994; Paolielo et al. 1997). We therefore com-

bined exposure data for men and women in this analysis. Exposure data were compiled separately for children and adults where this information was available.

2.3 Country estimates and subregional aggregation

Adjustments for data from countries currently phasing out lead in petrol

Blood-lead concentrations can change dramatically over a few years in response to programmes to eliminate the use of lead in petrol. Therefore, data which were more than one year old from countries that had reduced exposure to lead required downward correction to avoid overestimating exposure.

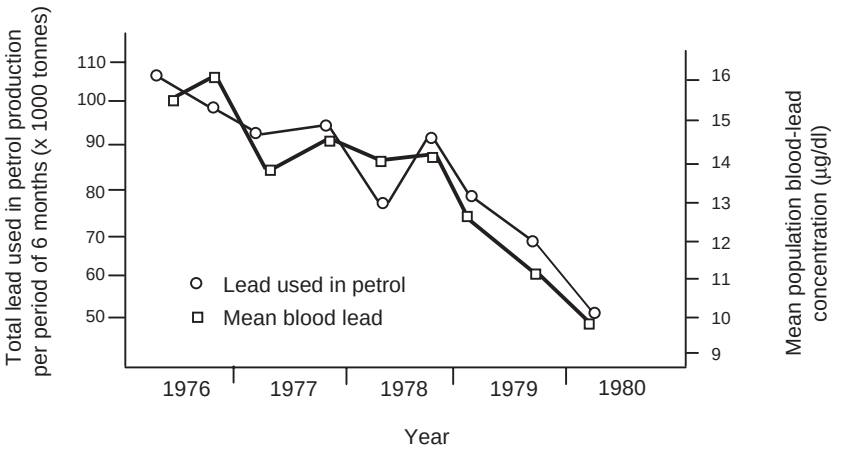
Decreases in blood-lead concentrations correlate well with the removal of lead from petrol (Thomas et al. 1999). Therefore, progress in phasing out lead was used to adjust exposure levels. Although leaded petrol is not the only source of lead in the environment, it is a good indicator of reduction in exposure to lead (Landrigan et al. 2000). Multiple studies have shown reductions in blood-lead concentrations in parallel with decreases in levels of lead in petrol (Thomas et al. 1999). Full implementation of lead reduction programmes has produced decreases in blood-lead concentrations in children of ≥90% over 25 years (CDC 1997, 2000).

Figure 19.1 shows changes in population blood-lead concentrations in the United States of America over a five-year period, during the early stages of a lead reduction programme (Annest 1983; Annest et al. 1983). During this period, the mean blood-lead concentration dropped by 37%. Other studies conducted in various countries have shown very similar results (Elinder et al. 1986; Schuhmacher et al. 1996; Wietlisbach et al. 1995), with decreases ranging from 30–48% over a five-year period. We chose the midpoint of 39% as a reduction factor for data that had been collected 5 years before the initiation of a leaded petrol phasing-out process or, for shorter periods, 7.8% decrease per year. Data which were >6 years old were not used. Only the most recent data were selected, provided that they were consistent with trends observed in older data.

Use of leaded petrol and the timing of any changes in the concentration of lead in petrol were assessed using data derived from the World Resources Institute (WRI 1998), the Earth Summit Watch (UNEP 2000), Car Lines (Walsh 2001) and M.P. Walsh (personal communication, 2002). A summary of the global situation with regard to reducing exposures to lead in petrol is provided in Figure 19.2 and below.

- *AFR-D and AFR-E*: With the exception of South Africa, countries in these subregions have not implemented lead reduction programmes.
- *AMR-A:* Blood-lead concentrations are likely to be significantly higher in Cuba than in Canada or the United States, as Cuba started lead reduction programmes more recently.

Figure 19.1 Decrease in mean population blood-lead concentrations in relation to reduction of lead in petrol,[a] in the United States



[a] The lead reduction programme started in 1974–1975.

Source: Annest (1983).

Figure 19.2 Sales of leaded petrol as a percentage of total petrol sales, by country, end of 2001

Source: Map based on data provided by M.P. Walsh.

- *AMR-B and AMR-D*: Most countries have now phased out leaded petrol. However, other sources, such as battery recycling and lead-glazed ceramics, are of importance in these subregions. Some countries, including Mexico and Peru, are major producers of lead.
- *EMR-B and EMR-D*: Some countries have now started a lead reduction programme. Egypt and Saudi Arabia are almost lead-free with regard to petrol.
- *EUR-A*: Most countries have phased out lead in petrol, and lead reduction programmes have been in place for a considerable period of time.
- *EUR-B and EUR-C*: While lead reduction programmes in some countries are relatively advanced, the great majority of countries have made little progress. EUR-C is less advanced than EUR-B in terms of phasing out leaded petrol.
- *SEAR-B and SEAR-D*: While Thailand has phased out lead in petrol, significant amounts of leaded petrol are still used in other countries. Sources other than leaded petrol (especially lead-containing cosmetics and local medications) greatly contribute to exposure in certain parts of these subregions.
- *WPR-A*: All countries have undertaken lead reduction programmes, most of which have been fully implemented. Japan began implementing lead reduction programmes at a very early stage.
- *WPR-B*: Some countries in this subregion, such as China and the Philippines, have made considerable efforts to phase out leaded petrol. Other countries have made little or no progress.

Exposure in rural populations

Where leaded petrol is still in use, blood-lead concentrations in rural populations are generally lower than those in urban populations (Nriagu et al. 1997a; Piomelli et al. 1980; Rhainds and Levallois 1993; Vasilios et al. 1997).

Thomas et al. (1999) showed that soon after the elimination of leaded petrol, population blood-lead concentrations tend to converge on an average of 3.1 µg/dl (SD 2.3 µg/dl). This value is similar to the mean, 3.0 µg/dl, of the data from rural areas available from countries in which lead has not been totally or partially phased out (3.4 µg/dl, Grobler et al. 1985; 2.3 µg/dl, Khwaja 2002; 3.8 µg/dl, Nriagu et al. 1997a; 3.4 µg/dl, Piomelli et al. 1980; 2.1 µg/dl, Vasilios et al. 1997). We therefore selected 3.1 µg/dl as the mean blood-lead concentration for rural populations in countries where concentrations of blood-lead in urban areas were higher than this value. In Latin America and the Caribbean, however, it was assumed that rural populations would have higher concentrations of blood-lead as other sources, such as ceramics and recy-

cling of batteries, contribute significantly to lead exposure in these areas (Romieu 2001a). The mean of the most recently-reported blood-lead concentrations in urban Latin American countries that have phased out lead is 4.3 μg/dl (Garcia and Mercer 2001; Sepulveda 2000), and we also used this value to represent rural blood-lead concentrations in this subregion.

After the complete phasing-out of leaded petrol, blood-lead concentrations continue to decrease, mainly as a result of additional efforts to reduce other sources of lead in the environment. Recent assessments from the United States have reported mean blood-lead concentrations of 1.6 μg/dl for the total population and 2.0 μg/dl for children aged <5 years (CDC 2001). In such cases, the same values were used to characterize urban and rural populations, as assuming higher exposures in urban environments did not seem justifiable.

Calculating subregional means

For each country with more than one source of data for blood-lead concentrations, geometric means were calculated (weighted by sample size) after the above adjustments. Means for urban and rural populations were estimated separately. Subregional means were calculated by weighting the mean for each country by the size of its urban population. For subregions with countries at different stages of lead phase-out (all subregions except EUR-A and AFR-D), urban means were estimated separately for countries which had made different degrees of progress in eliminating lead. In summary, we superimposed two or three log-normal distributions for each subregion to characterize the distribution of blood-lead concentrations in the population, each of which was weighted by the size of the population they represent. The urban/rural breakdown was based on data from UNDP (2000).

Fewer data on standard deviations were available than on mean blood-lead concentrations. In order to estimate standard deviations for areas for which data were sparse, we grouped subregions according to economic and lead-use patterns and calculated the average standard deviation for each grouping (AMR-A; EUR-A and WPR-A; AMR-B and D; remaining B and C subregions; remaining D and E subregions). Country averages (weighted by sample size) were estimated, and then averaged into the subregional standard deviations by weighting for the size of the urban population. For the grouping of D and E subregions, we did not weight by population size because the large countries contained within these two groupings were not representative of other countries. Table 19.1 shows the urban means for children and adults, the standard deviations, and the distribution of the population into categories of blood-lead concentration. As mentioned above, mean blood-lead concentrations in rural populations are assumed to be 3.1 μg/dl, or equal to blood-lead concentrations for urban populations in which levels have declined after lead reduction programmes.

Table 19.1 Blood-lead concentrations in children and adults, by subregion

Subregion	*AFR-D*[a,c]	*AFR-E*[a,c]	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
Mean blood-lead concentration, urban children (μg/dl)[b]	11.1	9.8	2.2	7.0	9.0	6.8	15.4	3.5	5.8	6.7	7.4	7.4	2.7	6.6
Mean blood-lead concentration, urban adults (μg/dl)[b]	11.6	10.4	1.7	8.5	10.8	6.8	15.4	3.7	9.2	6.7	7.4	9.8	2.7	3.6
Standard deviation (μg/dl)	5.6	5.6	2.9	3.9	3.9	3.9	5.6	1.9	3.0	3.0	3.0	5.6	1.9	3.0
Percentage of urban population	36	25	77	74	58	67	37	78	62	72	31	26	80	32

Countries with recent data	*Nigeria*[1]	*South Africa*[2]	*Canada*,[3] *USA*[4]	*Argentina*,[5] *Brazil*,[6] *Chile*,[7] *Jamaica*,[8] *Mexico*,[9] *Uruguay*,[10] *Venezuela*[11]	*Ecuador*,[12] *Nicaragua*,[13] *Peru*[14]	*Saudi Arabia*[15]	*Egypt*,[16] *Morocco*,[17] *Pakistan*[18]	*Denmark*,[19] *France*,[20] *Germany*,[21] *Greece*,[22] *Israel*,[23] *Sweden*[24]	*Poland*,[25] *Turkey*,[26] *Former Yugoslavia*[27]	*Hungary*,[28] *Russian Federation*[29]	*Indonesia*,[30] *Thailand*[31]	*Bangladesh*,[32] *India*[33]	*Australia*,[34] *Japan*,[35] *New Zealand*,[36] *Singapore*[37]	*China*,[38] *Micronesia*,[39] *Philippines*,[40] *Republic of Korea*[41]
Percentage of children with 5–10 μg/dl	18.6	19.1	12.4	21.2	23.2	23.3	18.1	22.7	22.7	23.6	21.8	19.2	14.1	21.8
Percentage of children with 10–20 μg/dl	10.0	8.9	4.7	16.3	16.4	15.7	10.1	5.1	13.8	16.3	11.2	8.8	2.9	10.9
Percentage of children with >20 μg/dl	13.9	9.5	1.9	16.7	17.2	11.4	17.2	0.5	8.9	11.9	6.5	8.3	0.3	5.8

continued

Table 19.1 Blood-lead concentrations in children and adults, by subregion (*continued*)

Countries with recent data	*Nigeria*[1]	*South Africa*[2]	*Canada*,[3] *USA*[4]	*Argentina*,[5] *Brazil*,[6] *Chile*,[7] *Jamaica*,[8] *Mexico*,[9] *Uruguay*,[10] *Venezuela*[11]	*Ecuador*,[12] *Nicaragua*,[13] *Peru*[14]	*Saudi Arabia*[15]	*Egypt*,[16] *Morocco*,[17] *Pakistan*[18]	*Denmark*,[19] *France*,[20] *Germany*,[21] *Greece*,[22] *Israel*,[23] *Sweden*[24]	*Poland*,[25] *Turkey*,[26] *Former Yugoslavia*[27]	*Hungary*,[28] *Russian Federation*[29]	*Indonesia*,[30] *Thailand*[31]	*Bangladesh*,[32] *India*[33]	*Australia*,[34] *Japan*,[35] *New Zealand*,[36] *Singapore*[37]	*China*,[38] *Micronesia*,[39] *Philippines*,[40] *Republic of Korea*[41]
Percentage of adults with 5–10 μg/dl	18.5	19.1	9.1	22.1	22.6	23.3	18.1	24.3	22.5	23.6	21.8	19.1	14.1	20.6
Percentage of adults with 10–20 μg/dl	10.0	8.9	3.2	17.2	16.6	15.7	10.1	5.7	16.7	16.3	11.2	9.0	2.1	8.4
Percentage of adults with >20 μg/dl	14.3	9.8	1.1	19.9	20.1	11.4	17.2	0.6	15.5	11.9	6.5	9.7	0.1	2.8

[a] Exposure data for these subregions were combined for this analysis.

[b] High and low means of urban blood-lead concentrations were used for subregions where countries were at different stages of the phasing-out of leaded petrol. The distribution is therefore a superposition of two or three log-normal distributions, and the mean and standard deviation do not necessarily reflect the distributions; therefore only one mean and standard deviation, as well as the distribution of people in exposure categories 5–10 μg/dl, 10–15 μg/dl and 15–20 μg/dl are displayed.

[c] Only data from South Africa, 1999, were used to represent countries with efforts in lead reduction in African subregions; older data were used to represent countries without such efforts.

Table references

1 Nriagu et al. (1997b); Omokhodion (1994).
2 Deveaux et al. (1986); Grobler et al. (1992); Karimi et al. (1999); Maresky and Grobler (1993); Nriagu et al. (1997a); von Schirnding et al. (2001); White et al. (1982).
3 Koren et al. (1990); Levallois et al. (1991); Rhainds and Levallois (1993); Smith and Rea (1995).
4 CDC (2000, 2001).
5 Cordeiro et al. (1996); Garcia and Mercer (2001).
6 dos Santos et al. (1994); Paolielo et al. (1997).
7 Sepulveda et al. (2000).
8 Matte et al. (1991).
9 Azcona-Cruz et al. (2000); Farias et al. (1998); Hernandez-Avila et al. (1996); Junco-Munoz et al. (1996); Lacasaña-Navarro et al. (1996); López-Lara et al. (2000); Romieu (2001b); Rothenberg et al. (1996).
10 Schutz et al. (1997).
11 Feo et al. (1993); Mujica (2001).
12 Counter et al. (1998).
13 Bonilla et al. (1998).
14 Jacoby (1998); Ramirez et al. (1997).
15 Al-Saleh (1995); Al-Saleh et al. (1995, 1999).
16 Kamal et al. (1991).
17 Khassouani et al. (1997).
18 Bashir et al. (1995); Hafeez and Malik (1996); Khan et al. (1994, 1995); Khwaja (2002); Sadaruddin et al. (1995).
19 Nielsen et al. (1998).
20 Flurin et al. (1998).
21 Jacob et al. (2000).
22 Vasilios et al. (1997).
23 Tepferberg and Almog (1999).
24 Bergdahl et al. (1997); Osterberg et al. (1997).
25 Dutkiewicz et al. (1993); Osman et al. (1999); Zejda et al. (1995, 1997).
26 Vural and Gulvendik (1988).
27 Blanusa et al. (1991); Factor-Litvak et al. (1996, 1998); Kostial et al. (1991).
28 Bitto et al. (1997).
29 Tepferberg and Almog (1999).
30 Heinze et al. (1998).
31 Wananukul et al. (1998).
32 Kaiser et al. (2001).
33 Awasthi et al. (1996); D'Souza et al. (1994); Gogte et al. (1991); Lal et al. (1991); Saxena et al. (1994); Shenoi et al. (1991); Wahid et al. (1997).
34 Australian Institute of Health and Welfare (1996).
35 Watanabe et al. (1996); Zhang et al. (1997).
36 Fawcett et al. (1996).
37 Chia et al. (1996, 1997); Neo et al. (2000).
38 Gao et al. (2001); Murata et al. (1995); Shen et al. (1996, 2001); Wan et al. (1996); Yan et al. (1999); Zhang et al. (1997).
40 Zhang et al. (1998).
41 Moon et al. (1995); Yang et al. (1996).

Based on these methods, we estimated that globally 120 million people had blood-lead concentrations of between 5 and 10 μg/dl in the year 2000, and about the same number of people had blood-lead concentrations of >10 μg/dl. Forty per cent of all children had blood-lead concentrations of >5 μg/dl, and 20% had concentrations of >10 μg/dl, and 97% of the latter were living in developing countries. Nine per cent of children had blood-lead concentrations of >20 μg/dl, and 99% of these children were living in developing countries.

3. Estimating risk factor–disease relationships

3.1 Health outcomes

Exposure to lead affects multiple health outcomes and physiological systems (ATSDR 1999), including the following: hypertension, the gastrointestinal system, anaemia, nephropathy, vitamin D metabolism, decreased growth, the immune system, the nervous system, behavioural/cognitive/IQ effects (and as a result, multiple social effects, including increased risk of violence and drug abuse), nerve conductive effects, hearing loss, effects on reproduction and development and death from encephalopathy.

Evidence relating exposure to lead and various health effects has been reviewed extensively (ATSDR 1988, 1993, 1999; International Programme on Chemical Safety 1977, 1995; National Research Council 1993; Pocock et al. 1994; Schwartz 1994). The most recent comprehensive review of the evidence on the risk factor–disease relationship was conducted by ATSDR in the United States (1999). Health effects considered in this review included systemic effects (e.g. raised blood pressure, gastrointestinal effects and anaemia) and nervous system effects (IQ reduction). Nephropathy and encephalopathy were not included as they rarely occur after environmental exposures but rather as a result of high-level exposure, such as ingestion of lead or lead salts, e.g. from local medication (Woolf 1990) or occupational exposure. A number of suggested health outcomes (e.g. developmental, reproductive and social effects) were not considered because of difficulties in quantifying the level of exposure at which a health outcome occurs, inadequate evidence of causality, or lack of information regarding baseline disease levels.

3.2 Evidence and exposure–risk relationships

General nervous system effects

The central and peripheral nervous systems are considered to be the principal targets affected by toxicity caused by the absorption of lead (Tsuchiya 1986; WHO 1996). Proposed mechanisms of toxicity (reviewed in ATSDR 1999) are based on the ability of lead to inhibit or mimic the action of calcium, and to interact with proteins. In terms of

general effects on the nervous system, the key mechanism is likely to be the substitution of lead for calcium as a "second messenger" in neurons. Lead blocks voltage-regulated calcium channels, inhibiting the influx of calcium and release of neurotransmitters and thus inhibiting synaptic transmission.

The most severe neurological effect of exposure to lead is encephalopathy. However, neurotoxic effects are apparent at much lower blood-lead concentrations than those that cause encephalopathy (i.e. ≤90 μg/dl for children, ≤140 μg/dl for adults, depending on the individual). Studies investigating occupational exposure to lead have reported symptoms such as loss of appetite, malaise, lethargy, headache, fatigue, forgetfulness and dizziness in workers with blood-lead concentrations of 40–120 μg/dl.

In the 1940s, Byers and Lord (1943) reported that children who had previously suffered from lead poisoning made poor progress at school, had a shorter attention span and exhibited behavioural disorders. Such observations were followed by epidemiological studies to determine the effects of low-level exposure to lead on children's intellectual abilities and behaviour. Low-level exposure to lead has been associated with failure to complete schooling, reading disability, longer reaction times, delinquent activity and other signs indicating effects on the central nervous system (Needleman et al. 1990, 1996). A large cohort study (Burns et al. 1999) showed that children exposed to relatively low levels of lead experienced an array of emotional and behavioural problems. Similarly, children who had experienced prenatal and postnatal exposure to lead were found to have an increased risk of cognitive deficit (Bellinger et al. 1990), problem behaviour and other dysfunctions, such as inappropriate approaches to tasks, or difficulty with simple directions and sequences of directions (Bellinger et al. 1994; Leviton et al. 1993). Nevin (2000) reported that long-term trends in population exposure to lead (indexed through use of leaded petrol and paint) were remarkably consistent with changes in violent crime; these findings are consistent with the reported link between IQ and social behaviour. Although it has been suggested that neurophysiological changes may be reversible (Ruff et al. 1993), the results of numerous studies indicate that this is unlikely (Schwartz et al. 2000b; Stokes 1998; Tong et al. 1998).

Effects on the nervous system other than loss of IQ and consequent mental retardation could not be quantified at the population level. This was either because of insufficient evidence linking effects to blood-lead concentrations, or because the outcome was not, in the strict sense, a quantifiable health effect (e.g. behavioural problems).

Loss of IQ points

Analyses of the body of evidence regarding the link between exposure to lead in early childhood and decrease in IQ score suggest that the relationship is causal (International Programme on Chemical Safety 1995;

Pocock et al. 1994; Schwartz 1994). In a meta-analysis, Schwartz (1994) estimated that a mean loss of 2.6 IQ points was associated with an increase in blood-lead concentration from 10 to 20 μg/dl. This result was robust to the inclusion or exclusion of the results of the strongest individual studies. This meta-analysis included eight cross-sectional and longitudinal studies, the largest longitudinal study being the Port Pirie cohort study in Australia (Baghurst et al. 1992) with about 500 participants and a follow-up of several years. Other meta-analyses report similar findings (International Programme on Chemical Safety 1995; Pocock et al. 1994). We selected the analysis by Schwartz (1994) to define the outcome as this study quantified the loss of IQ points and provided a point estimate with a confidence interval. At blood-lead concentrations of >20 μg/dl, we assumed a loss of 2–5 IQ points (midpoint of 3.5 IQ points) on the basis of the conclusions of the ATSDR report, which were derived from evidence from two studies (de la Burde and Choate 1972; Rummo et al. 1979).

Schwartz (1994) also reported that loss of IQ points is likely to be found between 5 and 10 μg/dl, with an even steeper relationship than in higher concentrations. For comparison, CDC currently defines blood-lead concentrations in excess of 10 μg/dl as elevated, although acknowledging Schwartz' analysis as evidence for subtle effects at lower concentrations (CDC 2000). The existence of such effects has recently been confirmed by Lanphear et al. (2000), whose analysis of data from about 5000 children showed that children with blood-lead concentrations of between 5 and 10 μg/dl had poorer cognitive skills.

In summary, to quantify IQ loss in the population, we assumed a linear relationship of 1.3 IQ points lost per 5 μg/dl increase in blood lead, for blood-lead concentrations of between 5 and 20 μg/dl (according to the analysis of Schwartz 1994, which showed a loss of 2.6 IQ points for an interval of 10 μg/dl). This linear relationship was divided into three segments, or increments, of 5 μg/dl, and the mean loss of IQ points in each increment was assigned to its mean blood-lead concentration—that is, 0.65 points for the increment 5–10 μg/dl with a mean of 7.5 μg/dl; 1.95 points for the increment 10–15 μg/dl with a mean of 12.5 μg/dl; and 3.25 points for the increment 15–20 μg/dl. The mean loss of IQ points is 0.65 (1.3/2) for the first increment, 1.95 (0.65+1.3) for the second, and 3.25 (1.95+1.3) for the third. This is illustrated in Figure 19.3. A loss of 3.5 IQ points was assumed for blood-lead concentrations of >20 μg/dl.

Mild mental retardation as a consequence of loss of IQ points

As loss of IQ points *per se* is not considered to be a disease by the international classification of disease (ICD) system, we converted IQ loss into mild mental retardation. Although loss of IQ potentially increases the risk for other diseases, injuries and adverse outcomes, such as violence (Needleman et al. 1996; Nevin 2000), no quantification could be made

Figure 19.3 Decrease in IQ points per increment increase in blood-lead concentration ("best estimate")

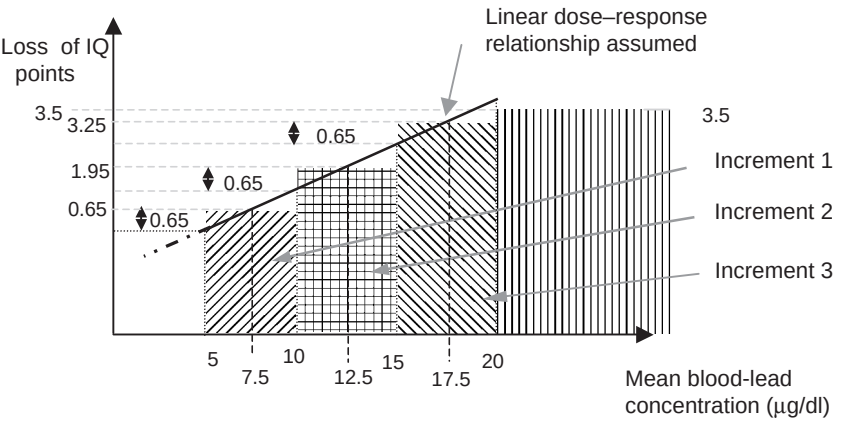
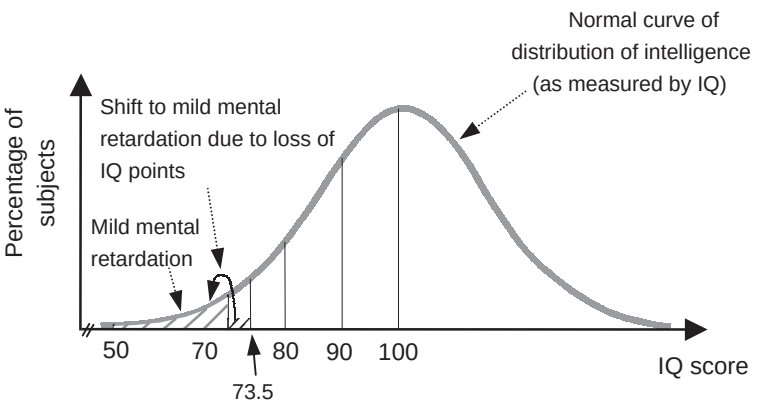


Figure 19.4 Loss of IQ points resulting in mild mental retardation



owing to the very heterogeneous nature of these relationships in different populations. For the purpose of this study, a reduction in IQ points was considered to be a disease burden when resulting in mild intellectual impairment, which was defined as having an IQ score of 50–69 points (see Figure 19.4). Intelligence in human populations approximates a normal distribution (Lezak 1995), except for an excess below IQ 50 (representing brain damage and disorder). To estimate the incidence of mild mental retardation resulting from IQ reduction attributable to lead exposure, we first estimated the proportion of children who would have

Table 19.2 Proportion of the population having an IQ score of between 70 and 73.50, assuming a normal distribution[a]

IQ (points)	*% of the population (assuming a normal IQ distribution)*
70–70.65	0.24
70–71.95	0.80
70–73.25	1.45
70–73.50	1.59

[a] Mean IQ score of 100, standard deviation of 15.

Source: Lezak (1995).

IQ scores close to the threshold defined and for whom the loss of a few IQ points would result in a total score of <70 points. This included the fractions of the child population with IQ scores of between 70 and 70.65, 71.95, 73.25 and 73.5 points (i.e. the intervals of interest defined by the loss of IQ points as specified above, representing increments of 0.65, 1.95, 3.25 and 3.5 points), assuming a normal distribution for IQ according to Lezak (mean IQ score of 100, standard deviation of 15; see Table 19.2).

The proportion of mild mental retardation attributable to exposure to lead was estimated as the proportion of children losing a number of IQ points (i.e. ratio of children with blood-lead concentrations within intervals 5–10 μg/dl, 10–15 μg/dl or 15–20 μg/dl; see Table 19.1), multiplied by the fraction of children within the interval $70+x$ IQ points (Table 19.2), for whom a loss of x points results in a final IQ score of <70 points (Figure 19.4). This is similar to the method used by INSERM (1999).

This standard IQ distribution does not include additional risk factors for IQ loss, which may be more common than exposure to lead. Several diseases that occur more frequently in developing countries result in cognitive impairment or mental retardation. The detailed estimates of the global burden of disease listed in the *World health report 2001* (WHO 2001) provide prevalences of cognitive impairment and mental retardation as a consequence of anaemia, meningitis and pertussis, Japanese encephalitis, ascaris, trichuriasis and infection with hookworm, as well as prevalences of cretinoidism and cretinism caused by iodine deficiency, for the 14 subregions studied here (Murray and Lopez 1996; WHO/EIP, unpublished data, 2001). The literature confirms that there are differences in the prevalence rates of mild mental retardation in developed countries compared to developing countries (Roeleveld et al. 1997), most of these differences being explained by noncongenital causes.

As the normal distributions of IQ scores were established on the basis of data from developed countries, the number of additional cases of mild

mental retardation that are likely to be observed in developing countries because of the additional risks mentioned above had to be estimated. This adjustment was based on the assumption that congenital causes of impaired cognitive function were separable and additive as compared to other risk factors. We thus assumed that the increase in the incidence of mild mental retardation (defined as an IQ score of 50–69 points) caused by factors other than exposure to lead was proportional to the increase in frequency of the IQ scores of slightly more than 70 points (i.e. between 70 and 73.5 points).

The prevalence of mild mental retardation and cognitive impairment resulting from known, noncongenital causes (see above) was estimated, values from developed and developing countries were compared, and an "adjustment ratio" to account for the increased risk of mental retardation in developing countries was estimated:

$$AR = \frac{P_R - P_{baseline} + P_{MR\ standard}}{P_{MR\ standard}}$$

AR	Adjustment ratio
MR	Mild mental retardation
P_R	Region-specific prevalence of MR from known causes from the Global Burden of Disease (GBD) database (WHO/EIP, unpublished data, 2001)
$P_{baseline}$	Prevalence of MR from known, noncongenital causes in developed countries
$P_{MR\ standard}$	Prevalence of MR according to the standard distribution of IQ score

The baseline prevalence of mild mental retardation caused by known, noncongenital factors of 420 per 100 000 population (typical in developed countries), and the total rate of mental retardation of 2270 per 100 000 population, as taken from the standard distribution, were used. The resulting subregion-specific adjustment ratios are summarized in Table 19.3. We assumed that the same adjustment ratio applied for the

Table 19.3 Adjustment ratios to account for excess prevalence rates of mild mental retardation caused by communicable diseases and iodine deficiency as compared to standardized rates

Subregion	*AFR-D*	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
Adjustment ratio	2.05	2.01	1.00	2.71	2.64	1.90	1.90	1.00	1.53	1.19	3.25	2.06	1.00	3.03

considered ranges of IQ (i.e. 70–73.5) in an attempt to take into account additional risks prevailing in certain developing subregions. It should be noted, however, that protein–energy malnutrition, potentially the most important risk factor for mild mental retardation, could not be taken into account here. The estimated prevalence rates per 1000 people affected by lead-induced mild mental retardation were multiplied by the subregion-specific adjustment ratios given in Table 19.3.

Lower mental ability has been associated with lower life expectancy. For example, a longitudinal cohort study of childhood IQ score and survival (Whalley and Deary 2000) found lower survival probabilities at age 76 for people who had a 15-point disadvantage in IQ score at age 11 years (RR = 0.76, 95% CI 0.75–0.84). Various mechanisms to explain this association were proposed, including childhood IQ as a record of bodily insult (e.g. antenatal care), as an indicator of "system integrity" (the availability of a cerebral reserve capacity to deal with other risks of cognitive decline), as a predictor of healthy behaviour and of entry into safer environments. All but the first mechanism would be relevant in relating the lead-induced loss of IQ points to additional adverse health effects or to reduced life expectancy, rather than considering mild mental retardation in isolation. Additional evidence, however, is needed to quantify this association.

Increased blood pressure

Schwartz (1995) conducted a meta-analysis examining the relationship between blood-lead concentration and systolic blood pressure in adult males (see Table 19.4). This analysis showed a significant association, with a reduction in blood-lead concentration from 10 μg/dl to 5 μg/dl being correlated with a decrease in blood pressure of 1.25 mmHg (95% CI 0.87–1.63). It has been suggested that lead exerts an influence on calcium metabolism, which is linked to modulation of blood pressure through vascular tone. *In vitro* studies have reported increased blood pressure in isolated tail arteries and increased responsiveness to alpha-adrenergic stimulation in response to exposure to lead. Raised blood pressure has been associated with increases in risk of cardiovascular and cerebrovascular disease.

A more recent meta-analysis by Nawrot et al. (2002) of data from 32 000 men estimated that a two-fold increase in blood-lead concentration was associated with a 1.2 mmHg increase in systolic blood pressure. While this analysis considered a two-fold increase in blood-lead concentration as the measure of exposure, most of the studies in the meta-analysis fell within the interval of 5–10 μg/dl considered by Schwartz (1995). The analysis of data from the second National Health and Nutrition Examination Survey (NHANES II) revealed decreases in blood pressure of 2 mmHg associated with a reduction in blood-lead from 20 μg/dl to 15 μg/dl and also from 15 μg/dl to 10 μg/dl (Pirkle et al. 1998; Schwartz 1988).

Table 19.4 The reduction in systolic blood pressure caused by a reduction in blood-lead concentration of 10 μg/dl in adult males

Reference	*Reduction in systolic blood pressure (mmHg)*	*Standard error*	*Age range (years)*	*Study type*
Orssaud et al. (1985)	1.74	0.73	24–55	Cross-sectional
Schwartz and Pitcher (1988)	2.24	0.86	20–74	Cross-sectional
Pocock et al. (1988)	1.45	0.49	40–59	Cross-sectional
Kromhout (1988)	3.15	1.20	57–76	Cross-sectional
Elwood et al. (1988, Wales)	0.25	0.49	18–64	Cross-sectional
Elwood et al. (1988, Caerphilly)	0.39	0.63	49–65	Cross-sectional
Neri et al. (1988)	1.05	0.70	NA	Cross-sectional and longitudinal
Moreau et al. (1988)	1.50	0.76	23–57	Cross-sectional
de Kort and Zwennis (1988)	0.90	0.39	25–60	Cross-sectional
Sharp et al. (1988))	0.80	1.25	28–64	Cross-sectional
Morris et al. (1990)	3.17	1.59	NA	Longitudinal
Egeland et al. (1992)	1.26	0.62	NA	Cross-sectional
Møller and Kristensen (1992)	1.86	0.63	40–51	Cross-sectional
Møller and Kristensen (1992)	0.90	0.74	40–51	Longitudinal
Hense et al. (1993)	1.45	0.51	28–67	Longitudinal

NA Not applicable.

Source: Adapted from Schwartz (1995).

As a conservative estimate, we used the same change of 1.25 mmHg in systolic blood pressure for all three intervals in blood-lead concentration. The relationship was assumed to be linear between 5 and 20 μg/dl, with a 1.25 mmHg rise in blood pressure for each incremental increase of 5 μg/dl. As with loss of IQ points, we converted the linear increase into three equal increments, using the midpoints of the increments with the corresponding increase in blood pressure.

In women, the association between systolic blood pressure and blood-lead concentrations is weaker and less well documented. The most recent and comprehensive estimate, using data from 24 000 women suggests that an increase of 0.8 mmHg in systolic blood pressure is associated with a doubling in blood-lead concentration (Nawrot et al. 2002). The association was not different from that of men in a statistically significant manner. We used an increase of 0.8 mmHg in systolic blood pressure for each 5 μg/dl increase in blood-lead concentration for women, for the interval between 5 and 20 μg/dl.

The disease burden caused by exposure to lead and mediated through increased blood pressure was based on the method used in chapter 6.

Anaemia

Absorbed lead inhibits the activity of a number of enzymes involved in haem biosynthesis. Several studies have shown that the activity of ALAD (δ-aminolevulinic acid dehydratase) is affected at very low blood-lead levels, with no apparent threshold (International Programme on Chemical Safety 1995). Typically, lead-induced anaemia arises from a combination of reduced haemoglobin formation (caused either by impaired haem synthesis or globin chain formation) and reduction in erythrocyte survival because of haemolysis (National Research Council 1993).

Adverse effects start to appear following decreases in concentrations of haemoglobin, which occurs at blood-lead concentrations of approximately 50 μg/dl in adults and 40 μg/dl in children, although there is increasing evidence that effects in children may occur at lower concentrations. Schwartz et al. (1990) studied the relationship between various levels of exposure to lead and anaemia in 579 children aged between 1 and 5 years living close to a primary lead smelter. The analysis related blood lead and hematocrit concentrations, and observed an increase in anaemia in children with blood-lead concentrations of >20 μg/dl. However, ATSDR (1999) defines children with blood-lead concentrations of ≥70 μg/dl and adults with ≥80 μg/dl to be at risk of anaemia, and we chose these thresholds for estimating disease burden. The number of studies looking at more severe health effects and issues of individual variability is relatively limited; however, results suggest that only a proportion of those exposed become ill. In the study conducted by Schwartz et al. (1990), 20% of children with a blood-lead concentration of ≥60 μg/dl exhibited signs of anaemia. This value was used in the present analysis to estimate the proportion of those exposed who became ill.

Gastrointestinal effects

Abdominal pain, constipation, cramps, nausea, vomiting, anorexia and weight loss, collectively known as colic, are early symptoms of lead poisoning in both adults and children. In adults, such symptoms occur at blood-lead concentrations of >80 μg/dl (International Programme on Chemical Safety 1995) and typically at concentrations of 100–200 μg/dl, while concentrations of 60–100 μg/dl are more typical for children (ATSDR 1999). No dose–response relationship for blood-lead and gastrointestinal effects has been published, so the same correction factor (20%) as that assumed for anaemia was used.

Other health effects

A number of health effects, such as nephropathy and encephalopathy, are associated with higher exposures to lead. These effects were not quantified as they occur in extreme cases for which population-based data from assumed distributions are highly uncertain.

Other health effects associated with lead exposure, such as hearing loss, cognitive deficits and reproductive effects, were not included in this

estimate. Exclusion was based on a number of factors, including difficulty in determining the threshold at which an effect could be expected to occur, inadequate causal evidence, or an outcome that fell outside of those for which the disease burden had been estimated in the GBD project.

Summary

Table 19.5 summarizes blood-lead concentrations at which the population is considered to be at risk of the health outcomes discussed here and the quantitative relationship between exposure and outcome. The values given in this table do not necessarily indicate the lowest levels at which lead exerts an effect.

Use of absolute versus relative risk ratios

The majority of studies investigating the relationship between lead exposure and disease have assessed incidence rates of disease for exposed and unexposed individuals in developed countries. These two rates are then generally combined into a relative risk. The exposure–risk relationships could therefore be applied to the populations of developing countries by transferring either the relative risks of disease, or the absolute disease

Table 19.5 Summary of health risks associated with blood-lead concentrations considered in this analysis

	Blood-lead concentration threshold (μg/dl)		
Outcome	*Children*	*Adults*	*Description of relationship*
IQ reduction[a]	5	NA	Linear relationship between 5 and 20 μg/dl (loss of 1.3 IQ points per 5 μg/dl BPb[c]); loss of 3.5 IQ points above 20 μg/dl
Increased systolic blood pressure[b]	NA	5	Linear relationship assumed between 5 and 20 μg/dl (increase of 1.25 mmHg per increase of 5 μg/dl BPb for males, and 0.8 mmHg for females), and increase of 3.75 mmHg above 20 μg/dl for males, and 2.4 for females
Gastrointestinal effects	60	NA	20% of people are assumed to be affected above these concentrations
Anaemia	70	80	20% of people are affected above these concentrations

NA Not applicable.

[a] Children aged 0–4 years.

[b] Applied to adults aged 20–79 years.

[c] BPb: blood-lead concentration.

rates for those exposed at equivalent levels. A transfer of relative risk rates to a country with higher baseline rates of the considered disease would result in a higher burden of disease for lead-induced illness than if absolute rates were transferred.

In the case of lead-induced outcomes occurring at high concentrations of blood-lead, including gastrointestinal symptoms and anaemia, it may be argued that lead poisoning acts independently of the baseline rates of disease and the presence of other risk factors in the population. We therefore determined incidences for gastrointestinal symptoms and anaemia on the basis of absolute risks rather than of relative risks. However, the risk of anaemia from exposure to lead may be magnified by other risk factors, and therefore a relative risk approach may be envisaged when more solid exposure–risk relationships become available.

With regard to the effects of lead on cognitive functions, data reported in the literature mainly provide mean decreases in IQ points, rather than relative risks for selected decreases. However, exposure to lead may interact with other risk factors, resulting in a magnifying effect, as previously mentioned.

As the exposure distribution relies on a database containing limited information concerning very high concentrations of blood-lead, the estimation of the percentage of the population that experiences extreme levels of exposure becomes very uncertain. Although the number of individuals with lead-induced nephropathy and encephalopathy could, in principle, be estimated, calculating disease burden would be misleading.

3.3 Estimation of the number of people affected by exposure to lead

To estimate the number of people whose health was affected by exposure to lead, the exposure–risk relationships described in section 3.2 were applied to the fraction of the population having the blood-lead concentrations at which these health effects occur. Figure 19.5 shows schematically how this was applied to the distribution of blood-lead concentrations in a population.

4. Sources of uncertainty

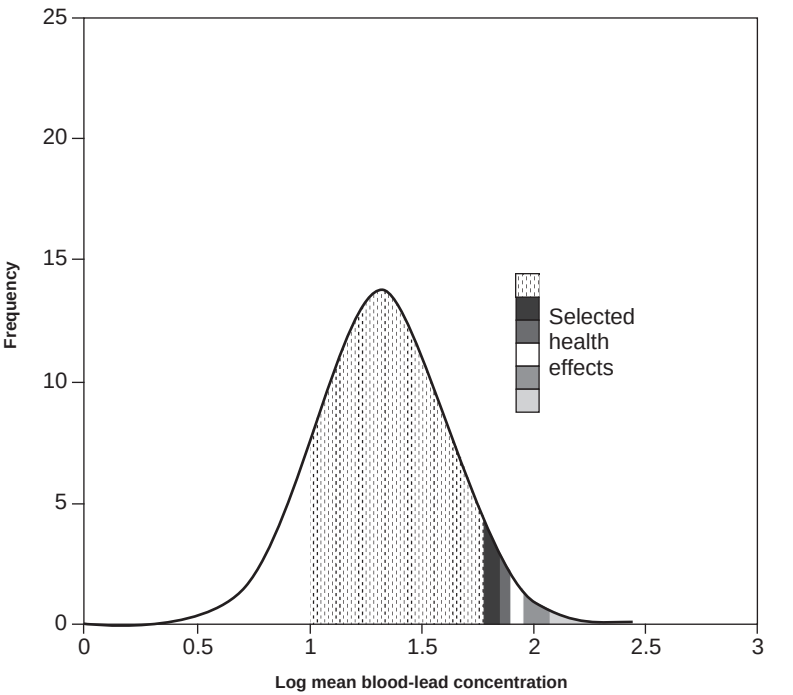
We estimated upper and lower uncertainty bounds for the best estimates by selecting upper and lower values for those parameters which were most likely to contribute significantly to uncertainty and which could be quantified.

4.1 Uncertainty in exposure assessment

Standard deviation of blood-lead concentrations

Fewer data are available for standard deviations of population blood-lead distributions than for mean blood-lead concentrations. We selected

Figure 19.5 Schematic diagram of the distribution of blood-lead concentrations in the population and the number of individuals who are at risk of selected health effects



upper and lower values for the standard deviation by recalculating subregional values after eliminating the upper and lower 20% of the reported standard deviations. The calculated values varied by approximately 13%.

Temporal change in blood-lead concentrations

Population blood-lead concentrations drop rapidly in countries that make substantial efforts to reduce exposure to lead, including the phasing-out of leaded petrol. Although we examined comprehensive sources documenting such efforts, and adjusted blood-lead concentrations to account for changes in exposure since the most recent assessment, uncertainty remained due to differences in the period of time needed to reduce the use of lead and the lack of data on timing of lead reduction programmes for some countries. In addition, lead reduction programmes other than the phasing-out of leaded petrol may have influenced blood-lead concentrations in the population since the year in

which the data were collected. The exposure level may therefore not always exactly correspond to the year 2000, but to a short time span around that year. To account for uncertainty in the timing of lead reduction, mean blood-lead concentrations were varied by ±16%, reflecting a two-year difference in progress in lead reduction programmes, assuming that blood-lead concentrations would change by 39% over a five-year period. Although not strictly applicable to the subregions in which leaded petrol has already been phased out for some time, we maintained this variation around the mean blood-lead concentration in these subregions to account for other factors that may have contributed to uncertainty.

Extrapolation to data-poor countries

Recent exposure information representative of parts of the general population was available for 41 countries. Many countries (or age groups) were, however, not represented. Exposure to lead in these countries was assumed to be similar to that in countries within the same subregion and which shared socioeconomic characteristics and similar implementation of lead reduction programmes.

Measurements of concentrations of blood lead

Uncertainty in the accuracy of measurement of blood-lead concentrations can be due to a deviation of the measured sample from the study population (i.e. bias), or to contamination problems during sampling or laboratory analysis. Also, the use of blood-lead measurements taken at a single point in time does not capture temporal variation in exposure to lead. Many sources of exposure, however, are likely to occur virtually continuously (e.g. lead in air, in drinking water, in certain foods or through use of leaded ceramics), limiting temporal variation other than that already accounted for.

4.2 Uncertainty in the exposure–risk relationships

Body burden of lead and associated health effects

Measurements of the concentration of lead in bone, which reflects long-term exposure, may be a better predictor of health effects than concentrations of lead in blood (Cheng et al. 2001; González-Cossío et al. 1997; Hu et al. 1996), but there are relatively few studies on which to base a global estimate. Also, the evidence on exposure–response relationships has not yet been quantified.

Health effects thresholds and individual variability

Health effects are likely to occur at lower concentrations of blood-lead than have been considered in this study. There may be no threshold for IQ/cognitive effects, and both renal and cardiac effects have recently been

reported to occur at low concentrations of blood-lead (Cheng et al. 1998; Payton et al. 1994). Health-effect thresholds are linked to individual variability, for which a number of factors are known to be important, including dietary factors (such as calcium; Harlan et al. 1985), general level of health, and genetic differences (e.g. Glen et al. 2001; Kelada et al. 2001; Schwartz et al. 2000a). Current knowledge does not allow the influence of such factors to be assessed precisely.

The relationship between exposure and effect for blood-lead concentrations and blood pressure

To quantify uncertainty regarding the effects of exposure to lead on blood pressure, the confidence interval around the risks of 30% reported by the meta-analysis (Schwartz 1995) was used.

Population-specific background mild mental retardation

A limitation of the approach used to estimate the risk of mild mental retardation attributable to lead exposure was the lack of studies examining the distribution of IQ scores in different populations. Additionally, population distributions have been found to change over time. In our estimate, the same normal distribution was applied worldwide. For estimating the uncertainty in mild mental retardation rates, we varied the mean IQ score (98 and 102, instead of 100). The lower bound for mild mental retardation was estimated by assuming that health effects occurred at >10 μg/dl instead of at >5 μg/dl as used in the best estimate.

4.3 Estimation of upper and lower uncertainty bounds

To derive an upper estimate for the proportion of people affected by lead exposure, the upper estimates for exposure assessment (i.e. upper estimates for standard deviation and mean of blood-lead concentrations) were multiplied by the upper rates in the risk estimates. A similar approach was used to obtain the lower estimates.

5. Results

5.1 Loss of IQ points and mild mental retardation

The incidence rates of mild mental retardation in children aged <5 years are summarized in Table 19.6. We assumed that loss of IQ and resulting mild mental retardation occurred only once, during the first year of life. Older age groups were assumed to have already experienced this health impact in previous years. Values presented in Table 19.6 are those estimated for the age group 0–1 year, but are divided by a factor of 5 (as the age group 0–4 years includes five 1-year cohorts of children).

The highest rates of mild mental retardation caused by exposure to lead occurred in developing countries, where the mean blood-lead concentrations were estimated to be many times higher than those in

Table 19.6 Proportion of children aged 0–1 year[a] affected by loss of IQ points caused by exposure to lead, and incidence rates of mild mental retardation caused by exposure to lead in children aged 0–1[a] year, in the year 2000

	AFR-D	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
	Proportion (number per 1000 children)													
IQ loss category														
0.65 IQ points	186	191	124	222	232	233	181	227	227	236	218	192	141	218
1.95 IQ points	66	61	33	104	105	102	66	41	92	106	76	61	23	75
3.25 IQ points	34	28	14	59	58	54	35	10	46	57	36	28	6	34
3.5 IQ points	139	95	21	167	172	114	172	5	89	119	65	83	3	58
Total	425	375	192	552	567	503	454	283	454	518	395	364	173	385
	Mean incidence rate (number per 1000 children)													
Mild mental retardation														
Best estimate	7.5	5.8	1.1	13.2	10.2	7.6	8.0	1.1	5.2	4.9	8.7	5.5	0.7	7.7
Lower estimate	4.2	3.0	0.5	7.0	5.3	3.6	4.6	0.3	2.4	2.3	3.9	2.8	0.2	3.4
Upper estimate	12.5	10.0	2.1	22.0	17.2	13.3	13.0	2.7	9.3	8.6	16.3	9.7	1.7	14.6

[a] GBD results were reported for children aged 0–4 years. As the effect is mostly irreversible, the entire cohort will remain affected.

developed countries. Latin American regions had relatively high incidence rates despite recent efforts to phase out lead.

5.2 Increased blood pressure

The proportion of adult men and women affected by increased blood pressure in age groups ranging from 20 to 79 years, are displayed in Table 19.7. To calculate the burden of disease for ischaemic heart disease, cerebrovascular disease, hypertensive disease and other cardiac diseases, these rates were converted into disease-specific relative risks, according to the methods used for the risk factor in chapter 6.

5.3 Anaemia and gastrointestinal symptoms

Tables 19.8 and 19.9 summarize the proportion of people affected by anaemia and gastrointestinal symptoms, assuming that these people are not removed from the source of lead or treated in order to reduce their blood-lead concentrations.

As anaemia and gastrointestinal effects were not included in the list of diseases for which baseline global data were available, they could not be quantified in terms of the attributable fraction of total disease or DALYs.

6. Discussion

This estimate of the global burden of disease caused by exposure to lead suggests that lead had a significant impact on health in the year 2000, mainly in developing countries where lead reduction programmes have not yet been fully implemented or, in some cases, initiated. In many subregions, relatively large fractions of the population had significantly elevated blood-lead concentrations. In particular, blood-lead concentrations in many developing countries in 2000 were comparable to, or even higher than, concentrations reported in the United States and Europe in the 1970s. The main disease end-points considered in this analysis included mild mental retardation caused by cognitive impairment and reduction of IQ, ischaemic heart disease, cerebrovascular disease, hypertensive disease and other cardiac diseases induced by increased blood pressure. Several additional disease outcomes associated with exposure to lead could not be considered in this analysis, either because the evidence was considered insufficient for a quantitative assessment at this point in time, or because baseline global data were not available. We estimated that 120 million people around the world had blood-lead concentrations of between 5 and 10 μg/dl in the year 2000, and about the same number had concentrations of >10 μg/dl. Forty per cent of all children had blood-lead concentrations of >5 μg/dl and 20% had concentrations of >10 μg/dl; 97% of these children were living in developing countries. These exposures resulted in a burden of disease of 9.8 million DALYs caused by mild mental retardation and 229 000

Table 19.7 Number of adult men and women[a] (per 1000) affected by increased systolic blood pressure with increased blood-lead concentrations caused by exposure to lead in the year 2000

Incremental increase in blood-lead concentration	*Number of adults (per 1000)*													
	AFR-D	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
0.625 mmHg in males, 0.4 mmHg in females	185	191	91	221	226	233	181	243	225	236	218	62	141	206
1.875 mmHg in males, 1.2 mmHg in females	66	61	23	108	106	102	66	46	106	106	76	29	17	61
3.125 mmHg in males, 2.0 mmHg in females	34	28	9	63	60	54	35	11	61	57	36	97	3	24
3.75 mmHg in males, 2.4 mmHg in females	143	98	11	199	201	114	172	6	155	119	65	97	1	28
Total	428	378	134	591	593	503	454	306	547	518	395	285	162	319

[a] Aged 20–79 years.

Table 19.8 Number of people (per 1000) affected by anaemia caused by exposure to lead in the year 2000

	Number of people (per 1000)													
	AFR-D	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
Children														
Best estimate	10	6	0	6	7	2	14	0	2	2	1	5	0	1
Lower estimate	4	3	0	2	2	0	6	0	0	0	0	2	0	0
Upper estimate	20	13	1	16	18	8	27	0	6	9	5	10	0	4
Adults														
Best estimate	9	6	0	7	8	2	13	0	3	2	1	6	0	0
Lower estimate	4	2	0	2	2	0	5	0	0	0	0	2	0	0
Upper estimate	18	12	0	18	21	7	25	0	12	7	4	12	0	1

Table 19.9 Number of children (per 1000 children) affected by gastrointestinal effects[a] caused by exposure to lead in the year 2000

	Number of children (per 1000 children)													
	AFR-D	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
Children														
Best estimate	12	7	0	8	9	3	16	0	2	3	2	6	0	1
Lower estimate	5	3	0	2	3	0	7	0	0	0	0	2	0	0
Upper estimate	22	14	1	20	22	11	30	0	8	11	6	11	0	5

[a] The basis for lower and upper estimates is outlined in section 4.

premature deaths and 3.1 million DALYs caused by cardiovascular disease. These two health outcomes alone account for about 0.9% of the global burden of disease.[2]

Assuming that each cohort of children aged 0–1 year is exposed to the same amount of lead year after year, the resulting prevalence of mild mental retardation attributable to lead exposure would be the following (obtained from Table 19.6):

- approximately 1–1.2% in AMR-B and AMR-D;
- approximately 0.5–0.8% for AFR, EMR, SEAR, WPR-B, EUR-B and EUR-C; and
- <0.1% in developed countries.

Prevalences of mild mental retardation reported for developed countries are generally about 1–3% (meta-analysis by Andersen et al. 1990; Baird and Sadovnick 1985; Murphy et al. 1995; Roeleveld et al. 1997; WHO 1985). Although it is commonly acknowledged that prevalence rates are higher in developing countries than in developed countries, data are scarce. Reported ranges vary greatly (0.4–9.5%, Roeleveld et al. 1997), and the median of available rates gives a prevalence of mild mental retardation of 6.5% (Durkin et al. 1998; Roeleveld et al. 1997). While the contribution of lead to the total incidence of mild mental retardation in developed countries is small, in developing countries as much as 15–20% of mild mental retardation could be caused by exposure to lead. A study from Australia (Wellesley et al. 1991) estimated that 40% of mental retardation was of genetic origin, 20% was caused by environmental factors and 40% was of unknown etiology. Also, the meta-analysis by Roeleveld et al. (1997) concluded that the high prevalence of mild mental retardation observed in developing countries points towards a role for partly avoidable exogenous influences.

An alternative approach for considering the effects of IQ loss was offered by the Dutch Burden of Disease study (Stouthard et al. 1997). This study used a severity weighting of 0.06 for any loss of IQ of between 1 and 4 points, whether this resulted in mental retardation or remained within the normal IQ range. Such an approach would magnify the effects of exposure to lead, but was not employed in this analysis, since incremental IQ loss is not considered to be a disease in the strict sense. At the same time, although in most instances a loss of IQ points does not lead to a recognizable health condition, it can affect physical functioning and life achievement (e.g. survival and earning potential). The Dutch approach is also supported by recent studies relating reduced mental ability to survival and to reduced lifetime earning capacity (Grosse et al. 2002; Korten et al. 1999; Whalley and Deary 2000).

The burden of ischaemic heart disease, cerebrovascular disease, hypertensive disease and other cardiac disorders caused by exposure to lead amounts globally to 3.1 million DALYs, which is about 2% of the total burden of cardiovascular disease. Worldwide, exposure to lead causes 229 000 deaths from these diseases. Ischaemic heart disease and cerebrovascular diseases are the two main contributors to the burden of disease in this group. Again, the burden is borne mainly by developing countries, owing to the higher exposures to lead in these areas. Together with mild mental retardation caused by exposure to lead, this brings the estimated total burden of disease in this analysis to 0.9%, in terms of DALYs. With quantification of additional outcomes discussed in this chapter, in particular, increased delinquent behaviour and its impact on injuries, the burden would most probably exceed 1% of the global total.

Lead did not contribute significantly to the global burden of anaemia, because other causes, such as iron deficiency, accounted for much higher prevalences of anaemia. The burden of gastrointestinal symptoms caused by lead was also relatively small as compared to that provoked by major risk factors such as unclean water, poor sanitation and hygiene, or unsafe food.

To improve the accuracy of these estimates, more population-representative blood-lead surveys from subregions for which little information has been published so far would be required. Also, additional information on the health impact of low lead levels would be needed in order to make estimates of the burden of disease caused by low doses of lead. The lack of quantitative information on health effects occurring at low exposure levels, the exclusion of data concerning exposure occurring around hot spots or at the workplace, combined with a number of conservative choices made throughout this study, all contribute to a probable underestimation of the burden of disease caused by lead. One particular health effect that could not be quantified but which has been associated in children with low levels of exposure to lead, and which may cause a significant disease burden, although indirectly, is violence.

Intentional injuries represent an important part of the burden of disease, a proportion of which may be attributable to low blood-lead concentrations encountered at high prevalences in many parts of the world.

In addition to the burden of disease, lead exposure may also contribute to socioeconomic burdens. Glotzer et al. (1995) estimated, for example, that in the United States, 45 000 cases of reading disability could be prevented by lead reduction programmes, saving more than US$ 900 million per year in overall costs of remedial education. This also has implications for inequalities in health, as exposure to lead tends to be higher in the lower socioeconomic groups of the population (Needleman 1994). These people often live in areas which are more exposed to industrial pollution or in degraded housing. Grosse et al. (2002) estimated that each IQ point raises worker productivity by 1.76–2.38%, resulting in an economic benefit of US$ 110–319 billion for each year's cohort of children.

All of the lead-induced disease burden is, in principle, preventable by phasing out the use of leaded petrol, reducing industrial emissions, removing lead from products such as ceramics, paint, "folk remedies" (traditional medicines) and food and drink cans, and replacing leaded pipes used for drinking-water. The phasing-out of leaded petrol is a particularly effective intervention, having the advantage of being a single action which permanently removes or reduces a health risk to current and future generations.

Although lead is one of the best-studied environmental pollutants, its full impact on population health is only now coming to light. The impacts of many other potentially harmful substances, such as heavy metals, pesticides or solvents, some of which are steadily accumulating in the environment, are as yet largely unknown.

7. Projections

Although exposure to lead can occur via a number of routes and from a range of sources, the level of lead in petrol is a key predictor of blood-lead concentrations at a country level. It has been shown that decreases in the use of leaded petrol are closely followed by parallel decreases in blood-lead concentrations (Annest 1983; Annest et al. 1983; Elinder et al. 1986; Schuhmacher et al. 1996; Thomas et al. 1999; Wietlisbach et al. 1995). We based the projected exposure estimates on predicted changes in transportation energy use by subregion (EIA 2001) and estimated completion dates of leaded petrol phase-out programmes (Walsh 2001). For countries that had not embarked upon lead reduction programmes, it was assumed that urban blood-lead concentrations would rise as a consequence of increases in transportation energy use, shown in Table 19.10. Where a lead reduction programme had been initiated, it was assumed that policy would not change and that the programme would be seen through to completion.

Table 19.10 Transportation energy use by subregion, 1990–2020

Subregion of analysis	*Energy Information Administration regional equivalent*	*Transportation energy consumption (million barrels of oil equivalent/day)*				*Average annual % change*
		1990	*1999*	*2010*	*2020*	*1999–2020*
AMR-A	North America	13	15	19	23	2.0
EUR-A	Western Europe	6	7	8	9	1.0
WPR-A	Industrialized Asia	3	3	3	3	1.0
EUR-B, EUR-C	Eastern Europe/ Former Soviet Union	3	2	3	4	2.8
SEAR-B, SEAR-D, WPR-B	Asia	3	6	10	16	5.1
EMR-B, EMR-D	Middle East	1	2	3	5	4.8
AFR-D, AFR-E	Africa	1	1	2	2	3.0
AMR-B, AMR-D	Central and South America	2	2	4	6	4.6

Source: Adapted from EIA (2001).

The starting point for the projection was the estimated mean blood-lead concentration for each subregion in the year 2000. In cases where a national lead reduction programme was due to be completed prior to 2010, urban blood-lead concentrations were assumed to converge at 3.1 μg/dl, according to the calculations of Thomas et al. (1999). The exceptions to this assumption were AMR-B and AMR-D. As countries in these subregions possess other important sources of lead, such as lead-glazed pottery, blood-lead concentrations were assumed to drop to 4.3 μg/dl, based on the mean of the most recently reported urban blood-lead concentrations in Latin American countries that have phased out leaded petrol (Garcia and Mercer 2001; Sepulveda et al. 2000). Energy use projections were available only until the year 2020; however, in the absence of other data, trends in 2000–2020 were assumed to continue until 2030. As urbanization is expected to increase steadily in most subregions (UN 1997), we also included a predicted change in urbanization in our projections of changes in blood-lead concentrations.

The effect of changes in energy use on blood-lead levels was calculated using the approach employed to adjust for the phasing-out of lead. Thus, as a 50% change in lead use (generally equivalent to the completion of a five-year leaded petrol reduction programme) is equal to a 39% change in blood-lead concentrations (see section 2.3), a 1% change in emissions would result in a 0.78% change in blood-lead concentrations. It was assumed that mean blood-lead concentrations would not increase beyond 30 μg/dl, as means exceeding this have rarely been reported.

Standard deviations were assumed to remain the same as in the year 2000. Projected blood-lead concentrations are presented in Table 19.11 for children and in Table 19.12 for adults. In subregions where many countries have not yet started to phase out lead (i.e. AFR, EMR, and EUR-B and C), it was estimated that blood-lead concentrations would increase steadily owing to the current widespread use of leaded petrol and lack of actions to reduce lead emissions. In most other subregions, lead emissions were predicted to decline gradually, with subsequent reductions in lead in the environment and in food. A large drop in blood-lead concentrations was projected to occur in many subregions between 2000 and 2010, as existing lead reduction programmes begin to take effect. After this period, rises in blood-lead concentrations would be observed due to continuing increases in the number of persons with elevated blood-lead concentrations living in countries and subregions which have not phased out lead.

Table 19.13 shows the projected incidence of mild mental retardation caused by exposure to lead for the years 2010, 2020 and 2030, using the methods described above and assuming constant prevalences of other diseases with cognitive impairment sequelae. As the proportion of people with elevated blood-lead concentrations will decrease in subregions where lead reduction programmes have recently been initiated (see Table 19.12), the incidence of mild mental retardation caused by exposure to lead is predicted to decline. Where no efforts to reduce exposure to lead are made, urbanization and increases in vehicle emissions are predicted to cause an increased incidence in lead-induced mild mental retardation. In the worst cases, in the African and Eastern Mediterranean subregions, exposure to lead could cause nearly 1.5% of cases of mild mental retardation.

The estimates for the incidence of anaemia and gastrointestinal symptoms caused by exposure to lead probably already carry the highest uncertainties, owing to the fact that these conditions appear at higher blood-lead concentrations, at which the distribution model is less accurate. Therefore projections for these outcomes have not been presented. In addition to the sources of uncertainty inherent in data collection and in the analysis for 2000, a number of additional uncertainties have been introduced into the projections we have made for future exposures, which are dominated by assumptions about policy and technology changes for lead reduction. Although it may well seem unacceptable that leaded petrol could still be in use in 10, 20 or 30 years' time, it must be remembered that lead has not been removed from the petrol supply of any country except by vigorous and concerted efforts by institutions concerned for the health of the public and especially of children.

For certain countries or subregions with significant additional sources of exposure to lead, the projections made here may be too optimistic. In general, however, we expect that appropriate lead-reduction measures will have been taken at a global level towards the end of the next decade.

Table 19.11 Projections of blood-lead concentrations in children for the years 2010, 2020 and 2030, by subregion

	Year	*AFR-D*	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
Subregional mean of urban blood-lead concentrations (μg/dl)	2010	14.1	9.7	2.2	4.7	4.3	7.7	17.0	3.1	6.3	8.3	3.2	3.7	2.7	3.1
	2020	18.1	12.8	2.2	4.9	4.3	10.5	20.9	3.1	7.7	10.5	3.3	3.8	2.7	3.2
	2030	23.5	17.4	2.2	5.1	4.3	14.7	20.6	3.1	9.2	13.1	3.5	3.7	2.7	3.1
Subregional mean of rural blood-lead concentrations (μg/dl)	2010–2030	3.1	3.1	2.2	4.3	4.3	3.1	3.1	3.1	3.1	3.1	3.1	3.1	2.7	3.1
Standard deviation (μg/dl)	2010–2030	5.6	5.6	2.9	3.8	3.8	3.0	5.6	1.9	3.0	3.0	3.0	5.6	1.9	3.0
Percentage of children with 5–10 μg/dl	2010	17.9	18.3	12.5	20.8	22.1	21.5	16.7	19.6	22.0	23.5	19.8	18.2	11.2	19.7
	2020	16.9	17.7	12.5	20.3	21.7	19.1	15.7	19.6	21.5	22.5	19.6	17.9	11.3	19.6
	2030	15.7	16.8	12.5	19.8	21.4	15.6	15.3	19.6	20.2	20.7	19.3	17.8	11.3	19.5
Percentage of children with 10–20 μg/dl	2010	10.8	9.5	4.8	13.9	13.3	16.5	10.3	3.8	14.9	18.5	8.1	8.2	1.2	8.1
	2020	11.4	10.1	4.8	13.9	13.4	17.4	10.6	3.8	16.5	20.5	8.5	8.5	1.2	8.3
	2030	11.8	10.4	4.8	13.9	13.4	16.2	11.0	3.8	17.1	21.6	8.7	8.5	1.2	8.5
Percentage of children with >20 μg/dl	2010	19.2	13.1	1.6	12.1	9.2	16.3	22.9	0.4	11.7	16.5	2.8	6.9	0.0	2.7
	2020	25.4	17.7	1.6	13.6	9.5	25.1	28.7	0.4	16.3	22.7	3.6	7.7	0.0	3.0
	2030	32.2	23.1	1.6	14.5	9.9	36.0	31.2	0.4	21.0	29.8	4.6	7.8	0.0	3.3

Table 19.12 Projections of blood-lead concentrations in adults for the years 2010, 2020 and 2030, by subregion

	Year	*AFR-D*	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
Subregional mean of urban blood-lead concentrations (μg/dl)	2010	14.8	10.1	1.7	4.8	4.3	7.7	17.0	3.1	8.6	8.3	3.2	3.8	2.4	2.7
	2020	19.0	13.3	1.7	5.0	4.3	10.5	20.9	3.1	10.7	10.5	3.3	3.8	2.4	2.7
	2030	24.6	18.1	1.7	5.1	4.3	14.7	20.6	3.1	12.7	13.1	3.5	3.7	2.4	2.7
Subregional mean of rural blood-lead concentrations (μg/dl)	2010–2030	3.1	3.1	2.2	4.3	4.3	3.1	3.1	3.1	3.1	3.1	3.1	3.1	2.7	3.1
Standard deviation (μg/dl)	2010–2030	5.6	5.6	3.0	3.8	3.8	3.0	5.6	1.9	3.0	3.0	3.0	5.6	1.9	3.0
Percentage of adults with 5–10 μg/dl	2010	17.8	18.3	9.3	20.8	22.1	21.5	16.7	19.6	20.4	23.5	19.8	18.2	12.4	19.8
	2020	16.8	17.6	9.3	20.2	21.7	19.1	15.7	19.6	18.9	22.5	19.6	17.8	12.4	19.6
	2030	15.6	16.7	9.3	19.8	21.4	15.6	15.3	19.6	17.1	20.7	19.3	17.8	12.4	19.5
Percentage of adults with 10–20 μg/dl	2010	10.8	9.5	3.2	13.9	13.3	16.5	10.3	3.8	16.0	18.5	8.1	8.2	1.4	8.0
	2020	11.4	10.0	3.2	13.9	13.4	17.4	10.6	3.8	16.6	20.5	8.5	8.5	1.4	8.3
	2030	11.7	10.4	3.2	13.9	13.4	16.2	11.0	3.8	16.1	21.6	8.7	8.5	1.4	8.5
Percentage of adults with >20 μg/dl	2010	19.7	13.3	0.9	12.3	9.2	16.3	22.9	0.4	18.3	16.5	2.8	6.9	0.1	2.6
	2020	26.0	18.0	0.9	13.8	9.5	25.1	28.7	0.4	24.3	22.7	3.6	7.8	0.1	2.9
	2030	32.8	23.5	0.9	14.5	9.9	36.0	31.2	0.4	29.6	29.8	4.6	7.8	0.1	3.2

Table 19.13 Projected incidence rates of mild mental retardation in children (aged 0–1 years) caused by exposure to lead in the years 2010, 2020 and 2030[a]

	Incidence rate (number per 1000)													
	AFR-D	*AFR-E*	*AMR-A*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-A*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-A*	*WPR-B*
2010	9.4	7.0	1.0	10.4	6.9	9.2	9.7	0.9	6.0	6.1	5.6	4.9	0.4	5.2
2020	11.6	8.6	1.0	11.1	7.0	12.0	11.5	0.9	7.4	7.5	6.1	5.2	0.4	5.4
2030	13.8	10.4	1.0	11.4	7.1	14.9	12.3	0.9	8.6	9.0	6.7	5.2	0.4	5.6

[a] GBD results reported for the 0–4 year age group.

In countries where mean population blood-lead concentrations are currently low, it is expected that health effects caused by lead will no longer be a concern for most people. However, exposure to lead is likely to remain a hazard for a minority of people, in particular, the children of the socially disadvantaged, including those living in houses containing leaded paint, or lead piping, or in areas affected by industrial contamination containing lead. Control of these sources will require continuing efforts.

Note

1 See preface for an explanation of this term.

2 Editorial note: The GBD mortality database includes a small number of deaths (approximately 5000) due to lead-induced mild mental retardation (MMR). These deaths are in fact deaths where MMR, regardless of being caused by lead or otherwise, has been specified as the underlying cause of death in the death registration data from some developed countries. The GBD has not attempted to make consistent estimates of MMR deaths for other regions, or to attribute some of these to lead. Because of the current GBD cause-of-death classification, these deaths are included in the Annex Tables (see the CD-ROM accompanying this book) and summary results in chapters 26 and 27. These deaths are however excluded from the results reported and discussed in this chapter.

References

Al-Saleh I (1995) Lead exposure in Saudi Arabia and its relationship to smoking. *Biometals*, **8**:243–245.

Al-Saleh I, Khalil MA, Taylor A (1995) Lead, erythrocyte protoporphyrin, and hematological parameters in normal maternal and umbilical cord blood from subjects of the Riyadh region, Saudi Arabia. *Archives of Environmental Health*, **50**:66–73.

Al-Saleh I, Nester M, DeVol E, Shinwari N, Al-Shahria S (1999) Determinants of blood lead levels in Saudi Arabian schoolgirls. *International Journal of Occupational and Environmental Health*, **5**:107–114.

Andersen E, Fledelius HC, Fons M, Haugsted R (1990) An epidemiological study of disability in 4-year-old children from a birth cohort in Frederiksborg County, Denmark. *Danish Medical Bulletin*, **37**:182–185.

Annest JL (1983) Trends in blood lead levels of the US population. In: *Lead versus health*. Rutter M, Jones RR, eds. John Wiley and Sons, New York.

Annest JL, Pirkle JL, Makuc D, Neese JW, Bayse DD, Kovar MG (1983) Chronological trend in blood lead levels between 1976 and 1980. *New England Journal of Medicine*, **308**:1373–1377.

ATSDR (1988) *The nature and extent of lead poisoning in children in the United States: a report to Congress*. (Agency for Toxic Substances and Disease Registry.) U.S. Department of Health and Human Services, Atlanta, GA.

ATSDR (1993) *Toxicological profile for lead*. (Agency for Toxic Substances and Disease Registry.) U.S. Department of Health and Human Services, Atlanta, GA.

ATSDR (1999) *Toxicological profile for lead* (update). (Agency for Toxic Substances and Disease Registry.) U.S. Department of Health and Human Services, Atlanta, GA.

Australian Institute of Health and Welfare (1996) *Lead in Australian children: summary of the National Survey of Lead*. Environment Protection Agency, Australia.

Awasthi S, Awasthi R, Pande VK, Srivastav RC, Frumkin H (1996) Blood lead in pregnant women in the urban slums of Lucknow, India. *Occupational and Environmental Medicine*, **53**:836–840.

Azcona-Cruz MI, Rothenberg SJ, Schnaas-Arrieta L, Romero-Placeres M, Perroni-Hernandez E (2000) Relationship of blood lead levels with visual-motor and equilibrium disturbances in children aged 8 to 10 years [in Spanish]. *Salúd Pública de México*, **42**:279–287.

Baghurst PA, McMichael AJ, Tong S, Wigg NR, Vimpani GV, Robertson EF (1995) Exposure to environmental lead and visual-motor integration at age 7 years. The Port Pirie cohort study. *Epidemiology*, **6**: 104–109.

Baghurst PA, McMichael AJ, Wigg NR et al. (1992) Environmental exposure to lead and children's intelligence at the age of seven years. The Port Pirie Cohort Study. *New England Journal of Medicine*, **327**:1279–1284.

Baird PA, Sadovnick AD (1985) Mental retardation in over half-a-million consecutive livebirths: an epidemiological study. *American Journal of Mental Deficiency*, **89**:323–330.

Bashir R, Khan DA, Saleem M, Zaman KU, Malik IA (1995) Blood lead levels and anemia in lead exposed workers. *Journal of the Pakistan Medical Association*, **45**:64–66.

Bellinger D, Leviton A, Allred E, Rabinowitz M (1994) Pre- and postnatal lead exposure and behavioural problems in school-aged children. *Environmental Research*, **66**:12–30.

Bellinger D, Leviton A, Sloman J (1990) Antecedents and correlates of improved cognitive performance in children exposed in utero to low levels of lead. *Environmental Health Perspectives*, **89**:5–11.

Bergdahl IA, Schutz A, Gerhardsson L, Jensen A, Skerfving S (1997) Lead concentrations in human plasma, urine and whole blood. *Scandinavian Journal of Work, Environment and Health*, **23**:359–363.

Bitto A, Horvath A, Sarkany E (1997) Monitoring of blood lead levels in Hungary. *Central European Journal of Public Health*, **5**:75–78.

Blanusa M, Telisman S, Hrsak J, Fugas M, Prpic-Majic D, Saric M (1991) Assessment of exposure to lead and cadmium through air and food in inhabitants of Zagreb. *Archives of Industrial Hygiene and Toxicology*, **42**:257–266.

Bonilla CM, Mauss L, Mauss EA (1998) A community-initiated study of blood lead levels of Nicaraguan children living near a battery factory. *American Journal of Public Health*, **88**:1843–1845.

Brody DJ, Pirkle JL, Kramer RA et al. (1994) Blood lead levels in the US population: phase 1 of the third National Health and Nutrition Examination Survey (NHANES III, 1988–1991). *Journal of the American Medical Association*, **272**:277–283.

Burns JM, Baghurst PA, Sawyer MG, McMichael AJ, Tong S (1999) Life-time low-level exposure to environmental lead and children's emotional and behavioural development at ages 11–13 years. The Port Pirie Cohort Study. *American Journal of Epidemiology*, **149**:740–749.

Byers RK, Lord EE (1943) Late effects of lead poisoning on mental development. *American Journal of Diseases in Children*, **66**:471–494.

CDC (1991) *Preventing lead poisoning in young children: a statement by the Centers for Disease Control*. Centers for Disease Control and Prevention. Department of Human Services, Atlanta, GA.

CDC (1997) Update: blood lead levels—United States, 1991–1994. *MMWR—Morbidity and Mortality Weekly Records*, **46**:141–146.

CDC (1999) Global lead literature database, Centers for Disease Control and Prevention, Atlanta, GA (unpublished).

CDC (2000) Blood lead levels in young children—United States and selected states, 1996–1999. *MMWR—Morbidity and Mortality Weekly Records*, **49**:1133–1137.

CDC (2001) *National report on human exposure to environmental chemicals*. Centers for Disease Control and Prevention, Atlanta, GA.

Cheng Y, Schwartz J, Sparrow D, Aro A, Weiss ST, Hu H (2001) Bone lead and blood lead levels in relation to baseline blood pressure and the prospective development of hypertension. *American Journal of Epidemiology*, **153**: 164–171.

Cheng Y, Schwartz J, Vokonas PS, Weiss St, Aro A, Hu H (1998) Electrocardiographic conduction disturbances in association with low level lead exposure. *American Journal of Cardiology*, **82**:594–599.

Chia SE, Chia HP, Ong CN, Jeyaratnam J (1996) Cumulative concentrations of blood lead and postural stability. *Occupational and Environmental Medicine*, **53**:264–268.

Chia SE, Chia HP, Ong CN, Jeyaratnam J (1997) Cumulative blood lead levels and neurobehavioral test performance. *Neurotoxicology*, **18**:793–803.

Cordeiro R, Lima Filho EC, Salgado PE (1996) Neurological disorders in workers with low levels of lead in the blood I: peripheral neuropathy [in Portuguese]. *Revista de Saude Pública*, **30**:248–255

Counter SA, Buchanan LH, Rosas HD, Ortega F (1998) Neurocognitive effects of chronic lead intoxication in Andean children. *Journal of the Neurological Sciences*, **160**(1):47–53.

de Kort WLAM, Zwennis WCM (1988) Blood lead and blood pressure: some implications for the situation in the Netherlands. *Environmental Health Perspectives*, **78**:6770.

de la Burde B, Choate Jr MS (1972) Does asymptomatic lead exposure in children have latent sequelae? *Journal of Pediatrics*, **81**:1088–1091.

Deveaux P, Kibel MA, Dempster WS, Pocock F, Formenti K (1986) Blood lead levels in pre-school children in Cape Town. *South African Medical Journal*, **69**:421–424.

dos Santos AC, Colacciopo S, Dal Bo CM, dos Santos NA (1994) Occupational exposure to lead, kidney function tests, and blood pressure. *American Journal of Industrial Medicine*, **26**:635–643.

D'Souza SJ, Narurkar LM, Narurkar MV (1994) Effect of environmental exposures to lead and cadmium on human lymphocytic detoxifying enzymes. *Bulletin of Environmental Contamination and Toxicology*, **53**:458–463.

Durkin MS, Hasan ZM, Hasan KZ (1998) Prevalence and correlates of mental retardation among children in Karachi, Pakistan. *American Journal of Epidemiology*, **147**:281–288.

Dutkiewicz T, Sokolowska D, Kulka E (1993) Health risk assessment in children exposed to lead compounds in the vicinity of mine-smelter plant "Orzel Bialy". *Polish Journal of Occupational Medicine and Environmental Health*, **6**:71–78.

Egeland GM, Burkhart GA, Schnorr TM et al. (1992) Effects of exposure to carbon disulfide on low density lipoprotein cholesterol concentration and diastolic blood pressure. *British Journal of Industrial Medicine*, **49**:287–293.

EIA (2001) Transportation energy use. In: *International energy outlook 2001*. Energy Information Administration, Washington DC.

Elinder CG, Friberg L, Lind B, Nilsson B, Svartengren M, Overmark I (1986) Decreased blood lead levels in residents of Stockholm for the period 1980–1984. *Scandinavian Journal of Work, Environment and Health*, **12**:114–120.

Elwood PC, Davey-Smith G, Oldham PD et al. (1988) Two Welsh surveys of blood lead and blood pressure. *Environmental Health Perspectives*, **78**:119–121.

Factor-Litvak P, Kline JK, Popovac D et al. (1996) Blood lead and blood pressure in young children. *Epidemiology*, **7**:633–637.

Factor-Litvak P, Slavkovich V, Liu X et al. (1998) Hyperproduction of erythropoietin in nonanemic lead-exposed children. *Environmental Health Perspectives*, **106**:361–364.

Farias P, Hu H, Rubenstein E et al. (1998) Determinants of bone and blood lead levels among teenagers living in urban areas with high lead exposure. *Environmental Health Perspectives*, **106**:733–737.

Fawcett JP Williams SM, Heydon JL, Walmsley TA, Menkes DB (1996) Distribution of blood lead levels in a birth cohort of New Zealanders at age 21. *Environmental Health Perspectives*, **104**:1332–1335.

Feo O, Fernandez M, Santaella N, Valera L (1993) Plumbemia en madres y sus hijos recien nacidos en el Hospital Central de Maracay [Lead in mothers and newborn children in the Maracay Central Hospital]. *Salud de los Trabajadores (Maracay)*, **1**:69–76.

Flegal AR, Smith DR (1992) Lead levels in preindustrial humans. *New England Journal of Medicine*, **326**:1293–1294.

Flurin V, Mauras Y, Le Bouil A, Krari N, Kerjan A, Allain P (1998) Lead blood levels in children under 6 years of age in the Le Mans region [in French]. *Presse Médicale*, **27**:57–59.

Gao W, Li Z, Kaufmann RB, Jones RL et al. (2001) Blood lead levels among children aged 1 to 5 years in Wuxi City, China. *Environmental Research*, **87**:11–19.

Garcia S, Mercer R (2001) *La problematica del plomo y salud de la ninez en la Argentina [The problems related to lead and health in children in Argentina]*. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment, May 07–10, WHO.) Lima, Peru.

Glen BS, Stewart WF, Schwartz BS, Bressler J (2001) Relation of alleles of the sodium-potassium adenosine triphosphate alpha 2 gene with blood pressure and lead exposure. *American Journal of Epidemiology*, **153**: 537–545.

Glotzer DE, Freedberg KA, Bauchner H (1995) Management of childhood lead poisoning: clinical impact and cost-effectiveness. *Medical Decision Making*, **15**:13–24.

Gogte ST, Basu N, Sinclair S, Ghai OP, Bhide NK (1991) Blood lead levels of children with pica and surma use. *Indian Journal of Pediatrics*, **58**:513–519.

González-Cossío T, Peterson KE, Sanin LH et al. (1997) Decrease in birth weight in relation to maternal bone-lead burden. *Pediatrics*, **100**:856–862.

Grobler SR, Maresky LS, Kotze TJ (1992) Lead reduction of petrol and blood lead concentrations of athletes. *Archives of Environmental Health*, **47**: 139–142.

Grobler SR, Rossouw RJ, Maresky LS (1985) Blood lead levels in remote, unpolluted rural area in South Africa. *South African Medical Journal*, **68**:323–324.

Grosse SD, Matte TD, Schwartz J, Jackson RJ (2002). Economic gains resulting from the reduction in children's exposure to lead in the United States. *Environmental Health Perspectives*, **110**:563–569.

Hafeez A, Malik QU (1996) Blood lead levels in pre-school children in Rawalpindi. *Journal of Pakistan Medical Association*, **46**:272–274.

Harlan WR, Landis JR, Schmouder RL, Goldstein NG, Harlan LC (1985) Blood lead and blood pressure. Relationship in the adolescent and adult

US population. *Journal of the American Medical Association*, **253**:530–534.

Heinze I, Gross R, Stehle P, Dillon D (1998) Assessment of lead exposure in schoolchildren from Jakarta. *Environmental Health Perspectives*, **106**: 499–501.

Hense HW, Filipiak B, Keil U (1993) The association of blood lead and blood pressure in population surveys. *Epidemiology*, **4**:173–179.

Hense HW, Filipiak B, Novak L, Stoeppler M (1992) Non occupational determinants of blood lead concentrations in a general population. *International Journal of Epidemiology*, **21**:753–762.

Hernandez-Avila M, González-Cossío T, Palazuelos E et al. (1996) Dietary and environmental determinants of blood and bone lead levels in lactating postpartum women living in Mexico City. *Environmental Health Perspectives*, **104**:1076–1082.

Hu H, Aro A, Payton M et al. (1996) The relationship of bone and blood lead to hypertension. The Normative Aging study. *Journal of the American Medical Association*, **275**:1171–1176.

INSERM (1999) *Plomb dans l'environnement. Quels risques pour la santé?* INSERM, Paris.

International Programme on Chemical Safety (1977) *Lead. Environmental health criteria 3*. World Health Organization, Geneva.

International Programme on Chemical Safety (1995) *Inorganic lead. Environmental health criteria 165*. World Health Organization, Geneva.

Jacob B, Ritz B, Heinrich J, Hoelscher BB, Wichmann HE (2000) The effect of low-level blood lead on hematologic parameters in children. *Environmental Research*, **82**:150–159.

Jacoby E (1998) Environmental lead is a problem in Lima, Peru. *Environmental Health Perspectives*, **10**:A170.

Junco-Munoz P, Ottman R, Lee JH, Barton SA, Rivas F, Cerda-Flores RM (1996) Blood lead concentrations and associated factors in residents of Monterrey, Mexico. *Archives of Medical Research*, **27**:547–551.

Kaiser R, Henderson AK, Daley WR et al. (2001) Blood lead levels of primary school children in Dhaka, Bangladesh. *Environmental Health Perspectives*, **109**:563–566.

Kamal AA, Eldamaty SE, Faris R (1991) Blood lead level of Cairo traffic policemen. *Science of the Total Environment*, **105**:165–170.

Karimi PG, Moodley J, Jinabhai CC, Nriagu, J (1999) Maternal and fetal blood lead levels. *South African Medical Journal*, **89**:676–679.

Kelada SN, Shelton E, Kaufmann RB, Khoury MJ (2001) Delta aminolevulinic acid dehydratase genotype and lead toxicity: a HuGE review. *American Journal of Epidemiology*, **154**:1–13.

Khan MH, Khan I, Shah SH, Rashid Q (1995) Lead poisoning-a hazard of traffic and industries in Pakistan. *Journal of Environmental Pathology, Toxicology and Oncology*, **14**:117–120.

Khan DA, Malik IA, Saleem M, Hashim R, Bashir R (1994) Screening for chronic lead poisoning in lead factory workers. *Journal of Pakistan Medical Association*, **44**:239–241.

Khassouani CE, Allain P, Soulaymani R (1997) Lead impregnation in inhabitants of the Rabat region [in French]. *Presse Médicale*, **26**:1714–1716.

Khwaja MA (2002) *Studies on blood lead levels in school children resulting from lead exposure due to lead petrol use and increasing road traffic in Pakistan*. (Presented at: Environmental threats to the health of children: hazards and vulnerability, March 3–7). Bangkok, Thailand.

Koren G, Chang N, Gonen R et al. (1990) Lead exposure among mothers and their newborns in Toronto. *Canadian Medical Association Journal*, **142**:1241–1244.

Korten AE, Jorm AF, Jiao Z, Leteneeur L, Jacomb PA, Henderson AS et al. (1999) Health, cognitive and psychosocial factors as predictors of mortality in an elderly community sample. *Journal of Epidemiology and Community Health*, **53**:83–8.

Kostial K, Dekanic D, Telisman S et al. (1991) Dietary calcium and blood lead levels in women. *Biological Trace Element Research*, **28**:181–185.

Kromhout D (1988) Blood lead and coronary heart disease risk among elderly men in Zutphen, the Netherlands. *Environmental Health Perspectives*, **78**:43–46.

Lacasaña-Navarro M, Romieu I, Sanìn-Aguirre LH, Palazuelos-Rendón E, Hernández-Avila M (1996) Consumo de calcio y plomo en sangre de mujeres en edad reproductiva [Consumption of calcium and blood lead level in women of reproductive age]. *Revista de Investgación Clinica*, **48**:425–430.

Lal M, Joseph D, Choudhury RK et al. (1991) Studies of blood lead levels in children by proton-induced X-ray emission (PIXE). *Science of the Total Environment*, **103**:209–214.

Landrigan PJ, Boffetta P, Apostoli P (2000) The reproductive toxicity and carcinogenicity of lead: a critical review. *American Journal of Industrial Medicine*, **38**:231–243.

Lanphear BP, Dietrich P, Auinger P, Cox C (2000) Subclinical lead toxicity in US children and adolescents. *Public Health Reports*, **115**:521–529.

Levallois P, Lavoie M, Goulet L, Nantel AJ, Gingras S (1991) Blood lead levels in children and pregnant women living near a lead-reclamation plant. *Canadian Medical Association Journal*, **144**:877–885.

Leviton A, Bellinger D, Allred EN, Rabinowitz M, Needleman H, Schoenbaum S (1993) Pre- and postnatal low-level lead exposure and children's dysfunction in school. *Environmental Research*, **60**:30–43.

Lezak MD (1995) *Neuropsychological assessment*. Oxford University Press, New York.

López Lara B, Cantú Martinez PC, Hernández Arizpe H, Gómez-Gúzman LG (2000) Niveles de plomo en sangre en recién nacidos y su relación con el peso al nacer [Blood lead levels and its relation to birth weight in newborns]. *Salud Pública y Nutrición*, **1**:1–7.

Maresky LS, Grobler SR (1993) Effect of the reduction of petrol lead on the blood lead levels of South Africans. *Science of the Total Environment*, **136**:43–48.

Matte TD, Figueroa JP, Ostrowski S, Burr G, Jackson-Hunt L, Baker EL (1991) Lead exposure from conventional and cottage lead smelting in Jamaica. *Archives of Environmental Contamination and Toxicology*, **21**: 65–71.

Molla AM, Akram S, Hussain A-K, Hussain NY (1997) Blood lead levels in Kuwaiti school children suspected of lead exposure. *Saudi Medical Journal*, **18**:290–293.

Møller L, Kristensen TS (1992) Blood lead as cardiovascular risk factor. *American Journal of Epidemiology*, **136**:1091–1100.

Moon CS, Zhang ZW, Shimbo S et al. (1995) Dietary intake of cadmium and lead among the general population in Korea. *Environmental Research*, **71**:46–54.

Moreau T, Hannaert G, Orssaud G et al. (1988) Influence of membrane sodium transport upon the relationship between blood lead and blood pressure in a general male population. *Environmental Health Perspectives*, **78**:47–52.

Morris C, McCarron DA, Bennet WM (1990) Low-level lead exposure, blood pressure, and calcium metabolism. *American Journal of Kidney Diseases*, **15**:568–574.

Mujica N. (2001) *Niveles de plomo en personas no expuestas ocupationalmente. Valores de referencia* [in Spanish]. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment, May 7–10, WHO.) Lima, Peru.

Murata K, Araki S, Yokoyama K et al. (1995) Autonomic and central nervous system effects of lead in female glass workers in China. *American Journal of Industrial Medicine*, **28**:233–244.

Murphy CC, Yeargin-Allsopp M, Decoufle P, Drews CD (1995) The administrative prevalence of mental retardation in 10-year-old children in metropolitan Atlanta, 1985 through 1987. *American Journal of Public Health*, **85**:319–323.

Murray CJL, Lopez AD, eds. (1996) *Global health statistics: a compendium of incidence, prevalence and mortality estimates for over 200 conditions*. Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Nawrot TS, Thijs L, Den Hond EM, Roels HA, Staessen JA (2002) An epidemiological re-appraisal of the association between blood pressure and blood lead: a meta-analysis. *Journal of Human Hypertension*, **16**:123–131.

Needleman HL (1994) Preventing childhood lead poisoning. *Preventive Medicine*, **23**:634–637.

Needleman HL (1999) History of lead poisoning in the world. In: *Lead poisoning prevention and treatment: implementing a national program in developing countries*. George AM ed. The George Foundation, Bangalore.

Needleman HL, Riess JA, Tobin MJ, Biesecker GE, Greenhouse JB (1996) Bone lead levels and delinquent behaviour. *Journal of the American Medical Association*, **275**:363–369.

Needleman HL, Schell A, Bellinger D, Leviton A, Allred EN (1990) The long-term effects of exposure to low doses of lead in childhood. An 11-year follow-up report. *New England Journal of Medicine*, **322**:83–88.

Neo KS, Goh KT, Sam CT (2000) Blood lead levels of a population group not occupationally exposed to lead in Singapore. *Southeast Asian Journal of Tropical Medicine and Public Health*, **31**:295–300.

Neri LC, Hewitt D, Orser B (1988) Blood lead and blood pressure: analysis of cross-sectional and longitudinal data from Canada. *Environmental Health Perspectives*, **78**:123–126.

Nevin R (2000) How lead exposure relates to temporal changes in IQ, violent crime, and unwed pregnancy. *Environmental Research*, **83**:1–22.

Nielsen JB, Grandjean P, Jorgensen PJ (1998) Predictors of blood lead concentrations in the lead-free gasoline era. *Scandinavian Journal of Work, Environment and Health*, **24**:153–156.

National Research Council (1993) *Measuring lead exposure in infants, children, and other sensitive populations*. National Academy Press, Washington, DC.

Nriagu J, Jinabhai CC, Naidoo R, Coutsoudis A (1997a) Lead poisoning of children in Africa, II Kwaulu/Natal, South Africa. *Science of the Total Environment*, **197**:1–11.

Nriagu J, Oleru NT, Cudjoe C, Chine A (1997b) Lead poisoning of children in Africa, III Kaduna, Nigeria. *Science of the Total Environment*, **197**: 13–19.

Omokhodion FO (1994) Blood lead and tap water lead levels in Ibadan, Nigeria. *Science of the Total Environment*, **151**:187–190.

Orssaud G, Claude JR, Moreau T et al. (1985) Blood lead concentration and blood pressure. *British Medical Journal*, **290**:244.

Osman K, Elinder C, Schutz A, Grubb A (1999) Biomarkers of nephrotoxicity in children environmentally exposed to lead in Poland. *Journal of Environmental Medicine*, **1**:33–38.

Osterberg K, Borjesson J, Gerhardsson L, Schutz A, Skerfving S (1997) A neurobehavioural study of long-term occupational inorganic lead exposure. *Science of the Total Environment*, **201**:39–51.

Paolielo MM, Gutierrez PR, Turini CA et al. (1997) Reference values for lead levels in blood for the urban population [in Portuguese]. *Revista de Saude Publica*, **31**:144–148.

Payton M, Hu H, Sparrow D, Weiss ST (1994) Low-level lead exposure and renal function in the normative aging study. *American Journal of Epidemiology*, **140**:821–829.

Piomelli S, Corash L, Corash MB et al. (1980) Blood lead concentrations in a remote Himalayan population. *Science*, **210**:1135–1137.

Pirkle JL, Kaufmann RB, Brody DJ, Hickman T, Gunter EW, Paschal DC (1998) Exposure of the US population to lead, 1991–1994. *Environmental Health Perspectives*, **106**:745–750.

Pocock SJ, Shaper AG, Ashby D, Delves HT, Clayton BE (1988) The relationship between blood lead, blood pressure, stroke, and heart attacks in middle aged men. *Environmental Health Perspectives*, **78**:23–30.

Pocock SJ, Smith M, Baghurst P (1994) Environmental lead and children's intelligence: a systematic review of the epidemiological evidence. *British Medical Journal*, **309**:1189–1197.

Ramirez AV, Paucar JC, Medina JM (1997) Blood lead in the inhabitants of four Peruvian localities [in Spanish]. *Revista Panamericana de Salud Pública*, **1**:344–348.

Rhainds M, Levallois P (1993) Umbilical cord blood lead levels in the Quebec City area. *Archives of Environmental Health*, **48**:421–427.

Roeleveld N, Zielhuis GA, Gabreels F (1997) The prevalence of mental retardation: a critical review of recent literature. *Developmental Medicine and Child Neurology*, **39**:125–132.

Romieu I (2001b) *Estudios epidemiológicos: diseño, interpretación y uso de datos de plomo [Epidemiological studies: design, interpretation and use of lead data]*. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment, May 7–10, WHO.) Lima, Peru.

Romieu I (2001a) *Plomo en America Latina y el Caribe [Lead in Latin America and the Caribbean]*. (Presented at: Latin American situation: lead poisoning in children: evaluation, prevention and treatment. May 7–10, WHO.) Lima, Peru.

Rothenberg SJ, Karchmer S, Schnaas L et al. (1996) Maternal influences on cord blood lead levels. *Journal of Exposure Analysis and Environmental Epidemiology*, **6**:211–227.

Ruff HA, Bijur PE, Markowitz M, Ma YC, Rosen JF (1993) Declining blood lead and cognitive changes in moderately lead-poisoned children. *Journal of the American Medical Association*, **269**:1641–1646.

Rummo JH, Routh DK, Rummo NJ (1979) Behavioural and neurological effects of symptomatic and asymptomatic lead exposure in children. *Archives of Environmental Health*, **34**:120–125.

Sadaruddin A, Agha F, Khatoon N, Sultana K (1995) Blood lead levels in young children in Chakshahzad, Islamabad. *Journal of Pakistan Medical Association*, **45**:215–218.

Saxena DK, Singh C, Murthy RC, Mathur N, Chandra SV (1994) Blood and placental lead levels in an Indian city: a preliminary report. *Archives of Environmental Health*, **49**:106–110.

Schuhmacher M, Belles M, Rico A, Domingo JL, Corbella J (1996) Impact of reduction of lead in gasoline on the blood and hair lead levels in the population of Tarragona Province, Spain, 1990–1995. *Science of the Total Environment*, **184**:203–209.

Schutz A, Barregard L, Sallsten G et al. (1997) Blood lead in Uruguayan children and possible source of exposure. *Environmental Research*, **74**:17–23.

Schwartz BS, Stewart WF, Bolla KI et al. (2000b) Past adult exposure is associated with longitudinal decline in cognitive decline. *Neurology*, **55**:1144–1150.

Schwartz BS, Stewart WF, Kelsey KT et al. (2000a) Associations of tibial lead levels with *Bsm*I polymorphisms in the vitamin D receptor in former organolead manufacturing workers. *Environmental Health Perspectives*, **108**: 199–203.

Schwartz J (1988) The relationship between blood lead and blood pressure in the NHANES II survey. *Environmental Health Perspectives*, **78**:15–22.

Schwartz J, Pitcher H (1988) The relationship between blood lead in the United States. *Environmental Health Perspectives*, **78**:23–30.

Schwartz J (1991) Lead, blood pressure and cardiovascular disease in men and women. *Environmental Health Perspectives*, **91**:71–75.

Schwartz J (1994) Low-level lead exposure and children's IQ: a meta-analysis and search for a threshold. *Environmental Research*, **65**:42–55.

Schwartz J (1995) Lead, blood pressure, and cardiovascular disease in men. *Archives of Environmental Health*, **50**:31–37.

Schwartz J, Landrigan PJ, Baker EL Jr, Orenstein WA, von Lindern IH (1990) Lead-induced anemia: dose–response relationships and evidence for a threshold. *American Journal of Public Health*, **80**:165–168.

Sepulveda AV, Vega Morales J, Delgado BI (2000) Childhood environmental lead exposure in Antofagasta, Chile [in Spanish]. *Revista de Médica de Chile*, **128**:221–232.

Sharp DS, Osterloh J, Becker CE et al. (1988) Blood pressure and blood lead concentrations in bus drivers. *Environmental Health Perspectives*, **78**: 131–137.

Shen XM, Rosen JF, Guo D, Wu SM (1996). Childhood lead poisoning in China. *The Science of the Total Environment*, **181**:101–109.

Shen XM, Sheng SH, Yan CH (2001) Impacts of low-level lead exposure on development of children: recent studies in China. *Clinica Chimica Acta*, **313**:217–220.

Shenoi RP, Khandekar RN, Jaykar AV, Raghunath R (1991) Sources of lead exposure in urban slum school children. *Indian Journal of Pediatrics*, **28**:1021–1027.

Smith LF, Rea E (1995) Low blood lead levels in Northern Ontario—what now? *Canadian Journal of Public Health*, **86**:373–376.

Stokes L (1998) Neurotoxicity in young adults 20 years after childhood exposure to lead: the Bunker Hill experience. *Occupational and Environmental Medicine*, **55**:507–516.

Stouthard MEA, Essink-Bot ML, Bonsle GJ et al. (1997) *Disability weights for diseases in the Netherlands*. Department of Public Health, Erasmus University, Rotterdam, the Netherlands.

Tepferberg M, Almog S (1999) Prenatal lead exposure in Israel: an international comparison. *Israel Medical Association Journal*, **1**:250–253.

Thomas VM, Socolow RH, Fanelli JJ, Spiro TG (1999) Effects of reducing lead in gasoline: an analysis of the international experience. *Environmental Science and Technology*, **33**:3942–3948.

Tong S, Baghurst PA, Sawyer MG, Burns J, McMichael AJ (1998) Declining blood lead levels and changes in cognitive function during childhood: the Port Pirie cohort study. *Journal of the American Medical Association*, **280**: 1915–1919.

Tsuchiya K (1986) Lead. In: *Handbook on the toxicology of metals*. Volume II. Specific metals. Friberg L, Nordberg GF, Vouk VB, eds. Elsevier, Amsterdam.

UN (1997) *Urban and rural areas*. (Publication ST/ESA/SER.A/166.) United Nations, New York.

UNDP (2000) *Human development report 2000*. United Nations Development Programme. Oxford University Press, New York.

UNEP (2000) *The global phaseout of leaded gasoline: a successful initiative*. UNEP Earth Summit Watch. Available at http://www.earthsummitwatch.org/gasoline.html.

Vasilios D, Theodor S, Konstantinos S, Evangelos PE, Fotini K, Dimitrios L (1997) Lead concentrations in maternal and umbilical cord blood in areas with high and low air pollution. *Clinical and Experimental Obstetrics and Gynecology*, **24**:187–189.

von Schirnding Y, Mathee A, Robertson P, Strauss N, Kibel M (2001) Distribution of blood lead levels in schoolchildren in selected Cape Peninsula suburbs subsequent to reductions in petrol lead. *South African Medical Journal*, **91**:870–872

Vural N, Gulvendik G (1988) Blood lead level distribution by age group in inhabitants of Ankara. *Biological Trace Element Research*, **18**:85–93.

Wahid A, Koul PA, Shah SU, Khan AR, Bhat MS, Malik MA (1997) Lead exposure in papier maché workers. *Human and Experimental Toxicology*, **16**:281–283.

Walsh MP (2001) Car Lines newsletter, 2001(4). (Unpublished document available at http://www.walshcarlines.com

Wan BJ, Zhang Y, Tian CY, Cai Y, Jiang HB (1996) Blood lead dynamics of lead-exposed pregnant women and its effects on fetus development. *Biomedical and Environmental Sciences*, **9**:41–45.

Wananukul W, Sirivarasai J, Sriapha C et al. (1998) Lead exposure and accumulation in healthy Thais: assessed by lead levels, EDTA mobilization and heme synthesis-related parameters. *Journal of the Medical Association of Thailand*, **81**:110–116.

Watanabe T, Nakatsuka H, Shimbo S et al. (1996) Reduced cadmium and lead burden in Japan in the past 10 years. *International Archives of Occupational and Environmental Health*, **68**:305–314.

Wellesley D, Hockey A, Stanley F (1991) The aetiology of intellectual disability in Western Australia: a community-based study. *Developmental Medicine and Child Neurology*, **33**:963–973.

Whalley LJ, Deary IJ (2000) Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal*, **322**:819–822.

White NW, Dempster WS, Pocock F, Kibel MA (1982) Lead absorption in Cape children: a preliminary report. *South African Medical Journal*, **62**:799–802.

WHO (1985) *Mental retardation: meeting the challenge*. (WHO Offset Publication No. 86.) World Health Organization Geneva.

WHO (1996) Lead. In: *Guidelines for drinking-water quality. Volume 2: Health criteria and other supporting information*. World Health Organization, Geneva.

WHO (1998) *Air management information system* (AMIS) 2.0. World Health Organization, Geneva.

WHO (2001) *World health report 2001*. World Health Organization, Geneva.

Wietlisbach V, Rickenbach M, Berode M, Guillemin M (1995) Time trend and determinants of blood lead levels in a Swiss population over a transition period (1984–1993) from leaded to unleaded gasoline use. *Environmental Research*, **68**:82–90.

Woolf MA (1990) Aetiology of acute lead encephalopathy in Omani infants. *Journal of Tropical Pediatrics*, **36**:328–330.

World Resources Institute (1998) *Resources 1998–1999. A guide to the global environment*. WRI, United Nations Environment Programme, United Nations Development Programme/World Bank, New York.

Yan CH, Shen XM, Zhang YW et al. (1999) A study on relationship between blood level and physical growth and development of babies and young children in Shanghai. *Chinese Journal of Preventive Medicine*, **33**:269–271.

Yang JS, Kang SK, Park IJ, Rhee KY, Moon YH, Sohn DH (1996) Lead concentrations in blood among the general population of Korea. *International Archives of Occupational and Environmental Health*, **68**:199–202.

Zejda JE, Grabecki J, Krol B, Panasiuk Z, Jedrzejczak A, Jarkowski M (1997) Blood lead levels in urban children of Katowice Voivodship, Poland: results of the population-based biomonitoring and surveillance program. *Central European Journal of Public Health*, **5**:60–64.

Zejda JE, Sokal A, Grabecki J, Panasiuk Z, Jarkowski M, Skiba M (1995) Blood lead concentrations in school children of Upper Silesian Industrial Zone, Poland. *Central European Journal of Public Health*, **3**:92–96.

Zhang ZW, Moon CS, Watanabe T et al. (1997) Background exposure of urban populations to lead and cadmium: comparison between China and Japan. *International Archives of Occupational and Environmental Health*, **69**: 273–281.

Zhang ZW, Subida RD, Agetano MG et al. (1998) Non-occupational exposure of adult women in Manila, the Philippines, to lead and cadmium. *Science of the Total Environment*, **215**:157–165.

Chapter 20

Global climate change

Anthony J. McMichael,
Diarmid Campbell-Lendrum, Sari Kovats,
Sally Edwards, Paul Wilkinson, Theresa Wilson,
Robert Nicholls, Simon Hales, Frank Tanser,
David Le Sueur, Michael Schlesinger and
Natasha Andronova

Summary

Accumulating evidence suggests that the global climate (i.e. conditions measured over 30 years or longer) is now changing as a result of human activities—most importantly, those which cause the release of greenhouse gases from fossil fuels. The most recent report (2001) from the United Nations' Intergovernmental Panel on Climate Change (IPCC) estimates that the global average land and sea surface temperature has increased by 0.6±0.2 °C since the mid-19th century, with most change occurring since 1976. Patterns of precipitation have also changed: arid and semi-arid regions are becoming drier, while other areas, especially mid-to-high latitudes, are becoming wetter. Where precipitation has increased, there has been a disproportionate increase in the frequency of the heaviest precipitation events. Based on a range of alternative development scenarios and model parameterizations, the IPCC concluded that if no specific actions were taken to reduce greenhouse gas emissions, global temperatures would be likely to rise between 1.4 and 5.8 °C from 1990 to 2100. Predictions for precipitation and wind speed were less consistent, but also suggested significant changes.

Risks to human health from climate change would arise through a variety of mechanisms. In this chapter, we have used existing or new models that describe observed relationships between climate variations, either over short time periods or between locations, and a series of health outcomes. These climate–health relationships were linked to alternative projections of climate change, related to unmitigated future emissions of greenhouse gases, and two alternative scenarios for greenhouse gas emissions. Average climate conditions during the period 1961–1990 were used as a baseline, as anthropogenic effects on climate are considered more significant after this period. The resulting models give estimates of the likely future effects of climate change on exposures to thermal extremes and weather disasters (deaths and injuries associated with

floods), the distribution and incidence of malaria, the incidence of diarrhoea, and malnutrition (via effects on yields of agricultural crops). As there is considerable debate over the extent to which such short-term relationships will hold true under the longer-term processes of climate change, we made adjustments for possible changes in vulnerability, either through biological or socioeconomic adaptation. Estimates of future effects were interpolated back to give an approximate measure of the effects of the climate change that have occurred since 1990 on the burden of disease in 2000.

The effects considered here represent only a subset of the ways in which climate change may affect health. Other potential consequences include influences of changing temperature and precipitation on other infectious diseases (including the possible emergence of new pathogens), the distribution and abundance of agricultural pests and pathogens, destruction of public health infrastructure, and the production of photochemical air pollutants, spores and pollens. Rising sea levels may cause salination of coastal lands and freshwater supplies, resulting in population displacements. Changes in the availability and distribution of natural resources, especially water, may increase risk of drought, famine and conflict.

Our analyses suggested that climate change will bring some health benefits, such as lower cold-related mortality and greater crop yields in temperate zones, but these will be greatly outweighed by increased rates of other diseases, particularly infectious diseases and malnutrition in developing regions. We estimated a small proportional decrease in cardiovascular and respiratory disease mortality attributable to climate extremes in tropical regions, and a slightly larger benefit in temperate regions, caused by warmer winter temperatures. As there is evidence that some temperature-attributable mortality represents small displacements of deaths that would occur soon in any case, no assessment was made of the associated increase or decrease in disease burden. Climate change was estimated to increase the relative risk of diarrhoea in regions made up mainly of developing countries to approximately 1.01–1.02 in 2000, and 1.08–1.09 in 2030. Richer countries (gross domestic product [GDP] >US$ 6000/year), either now or in the future, were assumed to suffer little or no additional risk of diarrhoea. This modest change in relative risk relates to a major cause of ill-health, so that the estimated associated disease burden in 2000 is relatively large (47 000 deaths and 1.5 million disability-adjusted life years [DALYs]). Effects on malnutrition varied markedly even across developing subregions,[1] from large increases in SEAR-D (RR=1.05 in 2000, and 1.17 in 2030) to no change or an eventual small decrease in WPR-B. Again, these are small relative changes to a large disease burden, giving an estimated 77 000 deaths and 2.8 million DALYs in 2000. We calculated much larger proportional changes in the numbers of people killed in coastal floods (RR in EUR-B of up to 1.8 in 2000, and 6.3 in 2030), and inland floods (RR in AMR-

A of up to 3.0 in 2000, and 8.0 in 2030). Although the proportional change is much larger than for other health outcomes, the baseline disease burden is much lower. The aggregate health effect in 2000 is therefore comparatively small (2000 deaths and 193 000 DALYs). We estimated relatively large changes in the relative risk of falciparum malaria in countries at the edge of the current distribution. However, most of the estimated attributable disease burden (27 000 deaths and 1 million DALYs) is associated with small proportional changes in regions that are already highly endemic, principally in Africa.

Overall, the effects of global climate change are predicted to be heavily concentrated in poorer populations at low latitudes, where the most important climate-sensitive health outcomes (malnutrition, diarrhoea and malaria) are already common, and where vulnerability to climate effects is greatest. These diseases mainly affect younger age groups, so that the total burden of disease due to climate change appears to be borne mainly by children in developing countries.

Considerable uncertainties surround these estimates. These stem partly from the complexity of climate models, partly from gaps in reliable data on which to base climate–health relationships, and, most importantly, from uncertainties around the degree to which current climate–health relationships will be modified by biological and socioeconomic adaptation in the future. These uncertainties could be reduced in subsequent studies by (i) applying projections from several climate models; (ii) relating climate and disease data from a wider range of climatic and socioeconomic environments; (iii) more careful validation against patterns in the present or recent past; and (iv) more detailed longitudinal studies of the interaction of climatic and non-climatic influences on health.

1. Introduction

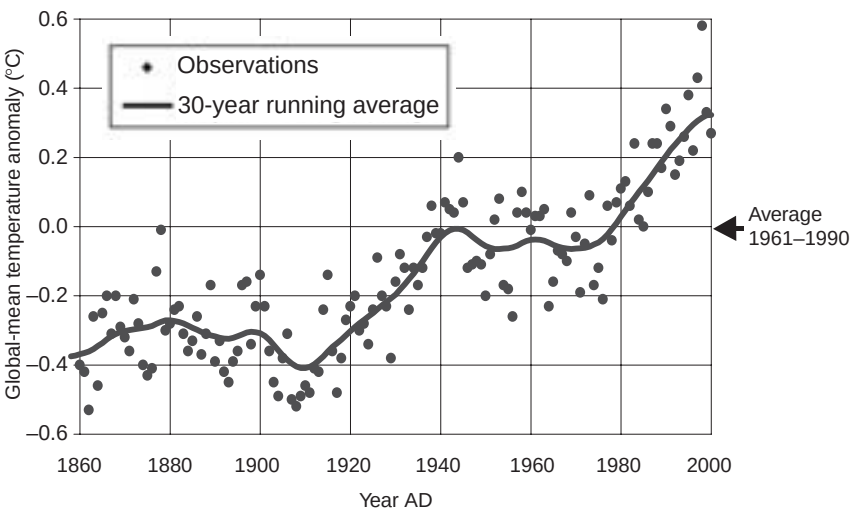
1.1 Evidence for climate change in the recent past and predictions for the future

Humans are accustomed to climatic conditions that vary on daily, seasonal and inter-annual time-scales. Accumulating evidence suggests that in addition to this natural climate variability, average climatic conditions measured over extended time periods (conventionally 30 years or longer) are also changing, over and above the natural variation observed on decadal or century time-scales. The causes of this climate change are increasingly well understood. Climatologists have compared climate model simulations of the effects of greenhouse gas (GHG) emissions against observed climate variations in the past, and evaluated possible natural influences such as solar and volcanic activity. They concluded that "... there is new and stronger evidence that most of the warming observed over the last 50 years is likely to be attributable to human activities" (IPCC 2001b).

The Third Assessment Report of the IPCC (IPCC 2001b) estimates that globally the average land and sea surface temperature has increased by 0.6±0.2 °C since the mid-19th century, with much of the change occurring since 1976 (Figure 20.1). Warming has been observed in all continents, with the greatest temperature changes occurring at middle and high latitudes in the Northern Hemisphere. Patterns of precipitation have also changed: arid and semi-arid regions are apparently becoming drier, while other areas, especially mid-to-high latitudes, are becoming wetter. Where precipitation has increased, there has also been a disproportionate increase in the frequency of the heaviest precipitation events (Karl and Knight 1998; Mason et al. 1999). The small amount of climatic change that has occurred so far has already had demonstrable effects on a wide variety of natural ecosystems (Walther et al. 2002).

Climate model simulations have been used to estimate the effects of past, present and likely future GHG emissions on climate changes. These models are primarily based on data on the heat-retaining properties of gases released into the atmosphere from natural and anthropogenic (man-made) sources, as well as the measured climatic effects of other natural phenomena, as described above. The models used by the IPCC have been validated by "back-casting"—that is, testing their ability to explain climate variations that already occurred in the past. In general, the models are able to give good approximations of past patterns only

Figure 20.1 Observed global average land and sea surface temperatures from 1860 to 2000



Source: Climatic Research Unit, Norwich, England.

when anthropogenic emissions of non-GHG air pollutants (particulates, dust, oxides of sulfur, etc.) are included along with natural phenomena (IPCC 2001b). This emphasizes that (i) the models represent a good approximation of the climate system; (ii) natural variations are important contributors to climatic variations, but cannot adequately explain past trends on their own; and (iii) anthropogenic GHG emissions are an important contributor to climate patterns, and are likely to remain so in the future.

Considering a range of alternative economic development scenarios and model parameterizations, the IPCC concluded that if no specific actions were taken to reduce GHG emissions, global temperatures would rise between 1.4 and 5.8 °C from 1990 to 2100. The projections for precipitation and wind speed are less consistent in terms of magnitude and geographical distribution, but also suggest significant changes in both mean conditions and in the frequency and intensity of extreme events (Table 20.1).

1.2 Estimating the effects of climate change on health

Human health is sensitive to temporal and geographical variations in weather (short-term fluctuations in meteorological conditions) and climate (longer-term averages of weather conditions). Weather has not historically been considered as subject to alteration by human actions, although its effects may be lessened by adaptation measures (e.g. Kovats et al. 2000b). While adaptation is also a very important determinant of the health consequences of climate change, the effect of anthropogenic GHG emissions on climate means that climate change can in principle be considered a risk factor that could potentially be altered by human intervention, with associated effects on the burden of disease.

The effects of GHG emissions on human health differ somewhat from the effects of other risk factors in that they are mediated by a diversity of causal pathways (e.g. Figure 20.2; McMichael et al. 1996; Patz et al. 2000; Reiter 2000) and eventual outcomes, typically long delays between cause and effect, and great difficulties in eliminating or substantially reducing the risk factors. An additional challenge is that climate change occurs against a background of substantial natural climate variability, and its health effects are confounded by simultaneous changes in many other influences on population health (Kovats et al. 2001; Reiter 2001; Woodward et al. 1998). Empirical observation of the health consequences of long-term climate change, followed by formulation, testing and then modification of hypotheses would therefore require long time-series (probably several decades) of careful monitoring. While this process may accord with the canons of empirical science, it would not provide the timely information needed to inform current policy decisions on GHG emission abatement, so as to offset possible health consequences in the future. Nor would it allow early implementation of policies for adaptation to climate changes, which are inevitable owing

Table 20.1 Estimates of confidence in observed and projected changes in extreme weather and climate events

Changes in phenomenon	*Confidence in observed changes (latter half of 1900s)*	*Confidence in projected changes (during the 21st century)*
Higher maximum temperatures and more hot days over nearly all land areas	Likely[a]	Very likely[a]
Higher minimum temperatures, fewer cold days and frost days over nearly all land areas	Very likely[a]	Very likely[a]
Reduced diurnal temperature range over most land areas	Very likely[a]	Very likely[a]
Increase of heat index[b] over land areas	Likely,[a] over many areas	Very likely,[a] over most areas
More intense precipitation events[a]	Likely,[a] over many northern hemisphere mid- to high-latitude land areas	Very likely,[a] over many areas
Increased summer continental drying and associated risk of drought	Likely,[a] in a few areas	Likely,[a] over most mid-latitude continental interiors. (Lack of consistent projections in other areas)
Increase in tropical cyclone peak wind intensities[c]	Not observed in the few analyses available	Likely,[a] over some areas
Increase in tropical cyclone mean and peak precipitation intensities[d]	Insufficient data for assessment	Likely,[a] over some areas

[a] Judgement estimates for confidence: *virtually certain* (greater than 99% chance that the result is true); *very likely* (90–99% chance); *likely* (66–90% chance); *medium likelihood* (33–66% chance); *unlikely* (10–33% chance); *very unlikely* (1–10% chance); *exceptionally unlikely* (less than 1% chance).

[b] Past and future changes in tropical cyclone location and frequency are uncertain.

[c] For other areas, there are either insufficient data or conflicting analyses.

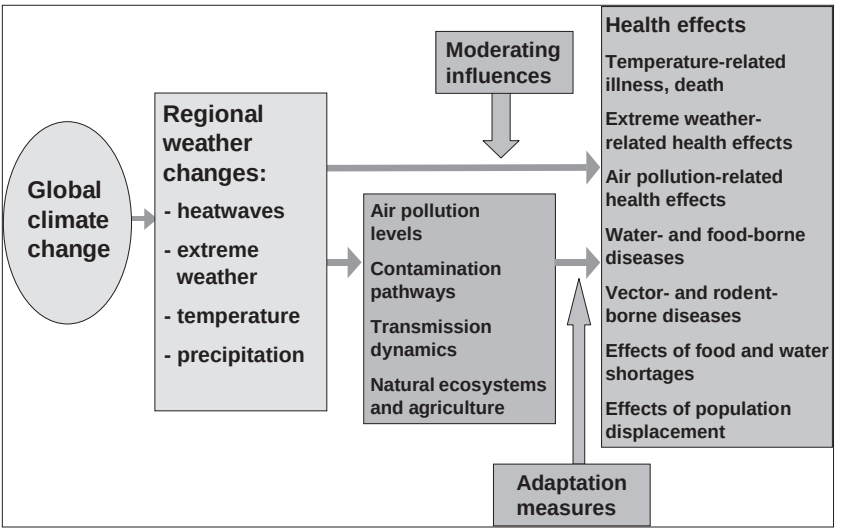
[d] Based on warm season temperature and humidity.

Source: Adapted from IPCC (2001b).

to both natural variations and past GHG emissions. Therefore, the best estimation of the future health effects of climate change will necessarily come from modelling based on current understanding of the effects of climate (not weather) variation on health from observations made in the present and recent past, acknowledging the influence of a large range of mediating factors.

Since the early 1990s, IPCC Working Group II has collated some of the accumulating predictions of climate effects on health (IPCC 2001a). The health effects of climate variability and change have also been reviewed by a national committee in the United States of America

Figure 20.2 Pathways through which climate change may affect health



Source: Adapted from Patz et al. (2000).

(National Research Council 2001a) and, more recently, in a book by the World Health Organization (WHO) (McMichael et al. 2003). As yet, however, there has been no concerted attempt to integrate these various research findings into a single standardized estimate of the likely net health effects of climate change, nor to estimate the possible health gains associated with different mitigation and amelioration strategies. In addition to the uncertainties of future climate projections, there are several obstacles to achieving this aim.

- Not all of the probable health outcomes have been modelled, often due to lack of parameterization data and the complexity of causal pathways. Modelling efforts so far have tended to concentrate on those causal relationships that can more easily be modelled, rather than those with the potentially greatest effects (e.g. extreme temperatures on cardiovascular disease mortality, rather than sea-level rise on the health of displaced populations).
- Little emphasis has been given to the validation of models relating climate change to health. Validation would provide a basis for making uncertainty estimates around projections, and would afford an objective criterion for choosing between different models or modelling approaches.

- Adaptations to climate change (i.e. autonomous or planned responses that reduce the vulnerability of populations to the consequences of climate change) are often not addressed.
- Interactions between the effects of climate change and other changes to human populations (e.g. investment in health infrastructure, level and equity of distribution of wealth) are seldom explicitly estimated.
- The various disease-specific models invariably generate outputs in different units, which may be only indirectly related to disease burden (e.g. populations at risk of disease transmission, rather than disease incidence). This hampers estimation of the aggregated health impacts of different scenarios.
- Little effort has previously been directed to describing and understanding the geographical variations in likely impacts.

The first four obstacles are likely to be at least partially addressed in the future, as disease-specific models become more sophisticated and, perhaps more importantly, through the accumulation of greater quantities of reliable data for model parameterization and testing. In this chapter, we have estimated the relative risk of a series of health outcomes under a range of scenarios of climate change, variously mitigated by reducing GHG emissions. In all cases, care was taken to describe explicitly the scientific basis for our estimation, the assumptions that were built into the quantitative models, and to give realistic uncertainty estimates around projections. Later sections describe ways in which specific disease models may be improved.

2. Risk factor definition and measurement

2.1 Definitions of risk factor and exposure scenarios

The risk factor was defined as current and future changes in global climate attributable to increasing atmospheric concentrations of greenhouse gases (GHGs).

Composite climate scenarios are adopted instead of the (more preferable) continuous measurements of individual climate variables because (i) climate is a multivariate phenomenon, including temperature, precipitation, wind speed, etc., and therefore cannot be measured on a single scale; (ii) climate changes will vary significantly with geography and time: these are not fully captured in global averages of climate variables; and (iii) all aspects of climate are likely to be altered by GHG levels in the atmosphere.

The exposure categories considered here are global climate scenarios resulting at specified points in time over the coming half-century from:

1. unmitigated emissions trends, that is, approximately following the IPCC IS92a scenario;

2. emissions reduction, resulting in stabilization at 750 ppm CO_2 equivalent by the year 2210 (s750);
3. more severe emissions reduction, resulting in stabilization at 550 ppm CO_2 equivalent by the year 2170 (s550);
4. average climate conditions for 1961–1990, the World Meteorological Organization (WMO) climate normal (baseline).

Although future GHG emissions are inherently uncertain, the unmitigated emissions scenario adopted here was, until recently, the IPCC mid-range projection, and was very widely used in climate impact modelling. The stabilization categories used here represent plausible, though economically and technically challenging, projections that are dependent on there being major efforts to curtail emissions. Estimated changes in CO_2 concentrations, and associated changes in global temperature and sea level, are shown in Table 20.2 and Figure 20.3. Although alternative emissions scenarios for climate stabilization are available, they have not been applied to a wide range of impact models.

We do not attempt to estimate all health outcomes of specific policy/development pathways through which these, or other, GHG levels could be achieved: for example, compliance with the Kyoto Protocol of the United Nations Framework Convention on Climate Change (UNFCCC), or of the world following one or other of the IPCC Special Report on Emissions Scenarios (SRES)—both of which also include descriptions of alternative future global socioeconomic development scenarios. The costs or additional benefits of specific interventions to

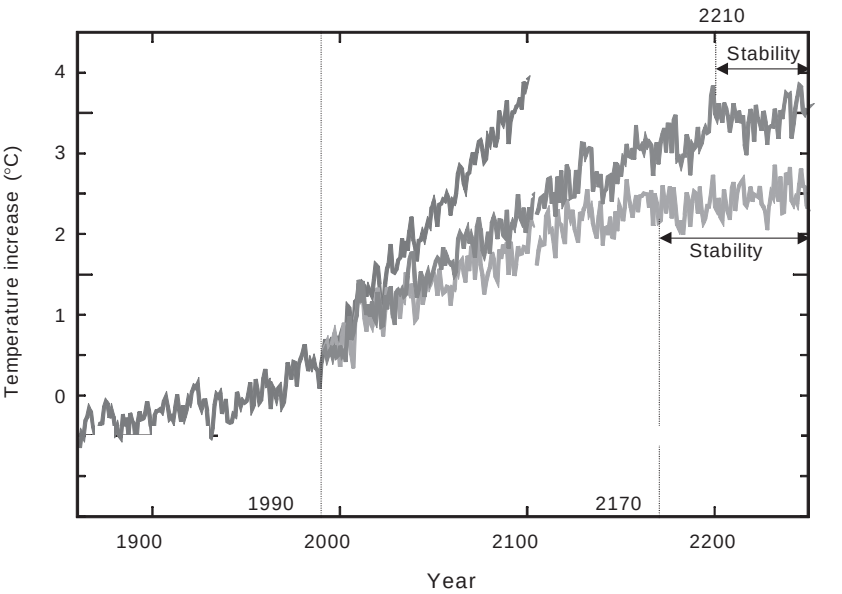
Table 20.2 Successive measured and modelled CO_2 concentrations, global mean temperature and sea-level rise associated with alternative emissions scenarios

	1961–90	*1990s*	*2020s*	*2050s*
Carbon dioxide concentration (ppm) by volume				
HadCM2 unmitigated emissions	334	354	441	565
S750	334	354	424	501
S550	334	354	410	458
Temperature (°C change)				
HadCM2 unmitigated emissions	0	0.3	1.2	2.1
S750	0	0.3	0.9	1.4
S550	0	0.3	0.8	1.1
Sea-level (cm change)				
HadCM2 unmitigated emissions	0	—	12	25
S750	0	—	11	20
S550	0	—	10	18

— No data.

Source: McMichael et al. (2000a).

Figure 20.3 The global average temperature rise predicted from the unmitigated emissions scenario (upper trace), and emission scenario which stabilizes CO_2 concentrations at 750 ppm (middle trace) and at 550 ppm (lower trace)



Note: All values are relative to mean values for the period 1961–1990, and may therefore be either positive or negative.

Source: Hadley Centre (1999).

achieve this reduction are artificially separated from the resulting health benefits. For such a distal risk factor, this separation of intervention from exposure and resulting health consequences may potentially introduce inconsistencies: for example, the economic changes necessary to achieve GHG stabilization are more consistent with some projections of levels and distribution of population and GDP than others. These socioeconomic factors are themselves likely to effect disease rates, potentially in interaction with climate. Integrated assessment of all effects of interventions would be conceptually more consistent, but this has not been attempted here, since it would introduce an additional layer of uncertainty and assumptions into the models and has previously only been explored for a few health outcomes (Tol and Dowlatabadi 2001).

The choice of baseline or "theoretical minimum" exposure follows the WMO and IPCC practice of using the observed global climate normal (i.e. averages) for 1961–1990 (New et al. 1999) as a reference point. Alternative baselines, such as pre-industrial climate, are not used,

because of the absence of a published consensus among climatologists on definitions of an appropriate time period and the relative roles of anthropogenic and natural influences before 1961–1990. This choice of a 1961–1990 baseline will therefore generate relatively conservative estimates of change in exposure and associated disease outcomes, as it does not address any human-induced climate change that occurred before this period. Indeed, as explained in section 2.6 below, the further choice of 1990 as the actual baseline year for linear-regression based estimates at current and future years heightens the conservative nature of these estimates.

The approach here treats climate change as a slowly evolving and continuous exposure, with the majority of disease models linked to those changes for which climate models make the most consistent predictions: gradual changes in temperature and, to a lesser extent, precipitation. This is again a limited approach. As shown in Table 20.1, it is very likely that climate change will also increase the frequency of extreme conditions, with likely effects on health. However, quantitative estimates of increased frequency have only very recently become available for some measures of extremes (e.g. wet winters; Palmer and Ralsanen 2002), and are not yet available for different GHG scenarios. The consequences of such changes are modelled here only in the context of inland flooding, but they could potentially be applied to other health end-points in the future.

Finally, there is some concern that disruption of the climate system may pass critical thresholds, resulting in abrupt rather than gradual changes (Broecker 1997; National Research Council 2001a) and associated rapid impacts on health. There is no consensus on the probability of such events, and they have therefore not been included in any published health outcome assessment studies. However, they should be borne in mind as a "worst-case" scenario.

2.2 Methods for estimating risk factor levels

Projections of the extent and geographical distribution of climate change were generated by applying the various emissions scenarios described above to the HadCM2 global climate model (GCM) of the Hadley Centre in the United Kingdom of Great Britain and Northern Ireland (Tett et al. 1997). This is one of several alternative GCMs used by the IPCC; it generates projections of changes in temperature and other climate properties which have been verified by back-casting (Johns et al. 2001), and which lie approximately in the middle of the range generated by alternative models.

The HadCM2 model generates estimates of the principal characteristics of climate, including temperature, precipitation and absolute humidity, for each cell of a global grid at resolution 3.75° longitude by 2.5° latitude. As for most climate change models, HadCM2 generates daily projections representing both long-term trends and the degree of natural

climate variability, but not necessarily its specific temporal pattern (i.e. the models do not accurately predict the climate of specific days or months). In order to account for such natural variability, the outputs that are most commonly used for modelling consequences of climate change are monthly means for average 30-year periods centred on the 2020s, 2050s, 2080s, etc. The baseline climate (1961–1990) describes the same properties for the land surface of the world at $0.5° \times 0.5°$ resolution.

The climate model projections describe forecast changes in global climate conditions. Therefore, we did not attempt to estimate the "exposure prevalence": the entire world population was assumed to be exposed to one or other global climate scenario (i.e. exposure prevalence = 100%). However, it should be noted that the climate scenarios incorporate geographical variation in both current climate (e.g. the lower temperatures in higher latitudes) and projected climate change (e.g. more rapid and intense warming is predicted in high northern latitudes than elsewhere). Different populations will therefore experience different climate conditions under any one climate-change scenario.

2.3 Uncertainties in risk factor levels

Two major sources of uncertainty surround the forecasting of future climate scenarios: (i) uncertainty in changes in factors such as population, economic growth, energy policies and practices on GHG emissions; and (ii) uncertainties over the accuracy of any climate model in predicting the effects of specified emissions scenarios on future climate in specific locations, against the background of substantial natural climate variability over time, and in space (i.e. downscaling). Climatologists have only very recently provided probabilistic measurements of uncertainty incorporating one or both of these sources (Knutti et al. 2002; Stott and Kettleborough 2002), and there is still debate over the reliability and utility of such measures (Schneider 2002). They have not previously been applied in impact studies (Katz 2002), and were therefore not used in this assessment. More importantly, however, there remains considerable uncertainty over the accuracy with which any single model can predict future climate. This is usually addressed by using outputs from a range of models, from independent groups. This was not possible here, as the particular GHG stabilization scenarios have only been applied, and fed through to estimates of likely consequences, for the HadCM2 model.

For this analysis, we did not address the first and last sources of uncertainty explicitly, and assumed that it is incorporated in the various alternative exposure scenarios, reflecting different trajectories of GHG emissions. The second source of uncertainty was partially addressed by using 30-year averages of climate conditions, which helps to "smooth out" the effects of natural climate variability. Further, the single model used was run with slight variations in initial conditions, allowing the calculation of an "ensemble mean". Four runs were used to generate an ensemble mean for the unmitigated emissions scenario. However, only

single climate runs were available for the stabilization scenarios. Therefore, the climate scenarios associated with those emissions scenarios are more uncertain.

Although it was not feasible to generate formal uncertainty ranges and feed these through the disease models, the approximate degree of uncertainty is illustrated in Appendix A. Here, the stabilization scenarios were run on a suite of 14 simple climate models, making different plausible assumptions about climate sensitivity to GHG emissions using the simplified COSMIC climate model described by Schlesinger and Williams (1997). This illustrates the degree of variation between the projections for future temperature and precipitation patterns. (Note that projections for precipitation vary more between models than does temperature—therefore models that rely on precipitation estimates from single scenarios have an additional component of uncertainty.)

3. Risk factor–disease relationship

3.1 Outcomes included

The health outcomes addressed here were selected on the basis of observed sensitivity to temporal and geographical climate variation, importance in terms of mortality and/or burden of disease (Longstreth 1999; McMichael et al. 1996; Patz et al. 2000) and availability of quantitative global models (or feasibility of constructing informative models in the time available) (Table 20.3). More detail on evidence for causality and quantitative estimation methods for each health outcome is given below.

Climate change is by and large a relatively distal risk factor for ill-health, often acting through complex causal pathways which result in heterogeneous effects across populations. There is, therefore, a series of additional likely outcomes that have not yet been formally modelled. They include the potential health consequences of climate change on:

Table 20.3 Health outcomes considered in this analysis

Outcome class	*Incidence/ prevalence*	*Outcome*
Direct effects of heat and cold	Incidence	Cardiovascular disease deaths
Foodborne and waterborne diseases	Incidence	Diarrhoea episodes
Vector-borne diseases	Incidence	Malaria cases
Natural disasters[a]	Incidence	Deaths due to unintentional injuries
	Incidence	Other unintentional injuries (non-fatal)
Risk of malnutrition	Prevalence	Non-availability of recommended daily calorie intake

[a] All natural disaster outcomes are separately attributed to coastal floods and inland floods/landslides.

- changes in pollution and aeroallergen levels;
- the rate of recovery of the ozone hole, affecting exposure to ultraviolet radiation (Shindell et al. 1998);
- changes in the distribution and transmission of other infectious diseases, particularly other vector-borne diseases, geohelminths and rodent-borne diseases, and possible emergence of new pathogens;
- the distribution and abundance of plant and livestock pests and diseases, affecting agricultural production (Baker et al. 2000; Rosenzweig et al. 2001);
- the probability of crop failure through prolonged dry weather and famine, depending on location and crisis management;
- population displacement due to natural disasters, crop failure, water shortages; and
- destruction of health infrastructure in natural disasters.

3.2 Methods for estimating risk factor–disease relationships

Various methods have been developed for the quantitative estimation of health outcomes of climatic change (reviewed by Martens and McMichael 2002; McMichael and Kovats 2000). It is not yet feasible to base future projections on observed long-term climate trends, for three reasons: (i) the lack of standardized long-term monitoring of climate-sensitive diseases in many regions; (ii) methodological difficulties in measuring and controlling for non-climatic influences on long-term health trends; and (iii) the small (but significant) climate changes that have occurred so far are an inadequate proxy for the larger changes that are forecast for coming decades (Campbell-Lendrum et al. 2002).

Instead, estimates are based on observations of the effects of shorter-term climate variation in the recent past (e.g. the effects of daily or inter-annual climate variability on specific health outcomes) or the present (e.g. climate as a determinant of current disease distribution), or on specific processes that may influence health states (e.g. parasite and vector population dynamics in the laboratory, determining the transmission of infectious diseases). These quantitative relationships were then applied to future climate scenarios (Figure 20.3). Such an approach makes the important assumption that such associations will be maintained in the future, despite changes in mediating factors such as socioeconomic variables, infrastructure and technology. This introduces significant uncertainty, and possibly bias, in the estimates.

The extent and type of modelling applied to different health effects vary considerably. Consequently, several outcomes can only be estimated by crude adaptation of the outputs of available models. For example, some of the predictive models generate health-relevant outputs that do

not correspond directly to categories of disease states used in the Global Burden of Disease (GBD) study. These include the incidence of deaths and injuries due to, specifically, floods (rather than injuries due to all causes), or populations at risk of hunger or malaria infection (rather than prevalence of malnutrition, or incidence of clinical malaria). Currently, there are spatial resolution differences between models, which is not ideal. These relate to how the models account for geographical variation in: (i) baseline climate (potentially differentiated to 0.5° globally, or higher resolutions for some regions); (ii) climate change (usually at the level of the GCM projections: 3.75° longitude by 2.5° latitude); and (iii) aggregation of the final results (occasionally according to regions other than subregions, depending on the purpose of the original model). Levels of spatial resolution for each disease model are described in section 3.6 and onwards. All of the above considerations are represented in the descriptions of strength of evidence and quantitative estimates of uncertainty for specific health outcomes.

Assumptions

Simplifying assumptions have been made to facilitate clear definition of scenarios and associated consequences.

Different mechanisms for reducing GHG emissions

The alternative GHG emissions scenarios outlined above could be achieved through an almost infinite variety of changes to economic and social development and energy use policies. As outlined in section 2.1, we did not attempt to estimate the secondary effects of GHG mitigation policies on health. These effects are potentially large. They include relatively direct mechanisms which may be *negative*, such as the potential negative effects of GHG emissions policies on economic development, personal wealth and vulnerability to disease (Tol and Dowlatabadi 2001), or *positive*, such as reduction in the levels of ozone and other outdoor (Kunzli et al. 2000) and indoor air pollutants (Wang and Smith 1999). They may also act through more complex routes, for example, avoiding production of aeroallergens in CO_2-enriched environments (Wayne et al. 2002; Ziska and Caulfield 2000).

Population growth

The models described below estimated the relative per capita incidence of specific health outcomes under the different climate scenarios. The size and distribution of current and future populations therefore affect the relative risk estimates either (i) where the climate hazard is not evenly distributed geographically throughout the region, or (ii) where population is an integral part of the model—for example, in the risk-of-hunger model, where population size has an influence on food availability per capita. In adjusting the relative risks for these population effects, the distribution of future populations was estimated by applying the World

Bank mid-range estimate of population growth either at the national level (for malnutrition), or to a 1°×1° resolution grid map of population distribution (Bos et al. 1994) for all other outcomes.

Modifying factors: adaptation and vulnerability

Factors such as physiological adaptation, technological and institutional innovation and individual and community wealth will influence not only the exposure of individuals and populations to climate hazards, but also the associated hazards (e.g. IPCC 2001a; Reiter 2000; Woodward et al. 1998). For some assessments, simpler modifying factors are integrated into the models for both present and future effects. For example, estimates of changes in the number of people at risk of hunger incorporate continental projections of economic growth, affecting capacity to buy food (Parry et al. 1999). Other models incorporate the effects of existing modifiers when defining current climate–disease relationships, such as estimates of the global distribution of malaria based on current climate associations (Rogers and Randolph 2000). Such models implicitly capture the *current* modifying effects of socioeconomic and other influences on climate effects, but do not attempt to model *future* changes in these modifiers. Finally, some models make no estimate of such modifying influences in either the present or future; for example, models that estimate future changes in the geographical range which is climatically suitable for malaria transmission, and associated populations at risk (Martens et al. 1999). To generate consistent estimates in this analysis:

- we attempted to account for current geographical variation in vulnerability to climate, where not already incorporated into the predictive models.
- we attempted to account for future changes in disease rates due to other factors (e.g. decreasing rates of infectious disease due to technological advances/improving socioeconomic status), and for changes in population size and age structure (e.g. potentially greater proportion of older people at higher risk of mortality related to cardiovascular disease in response to thermal extremes). This was addressed by calculating only relative risks under alternative climate change scenarios, which should be applied to GBD projections of disease rates and population size and age structure. The GBD projections take into account the effects of changing GDP, "human capital" (as measured by average years of female education), and time (to account for trends such as technological development) (Murray and Lopez 1996) on the overall "envelope" of cause-specific mortality and morbidity for diseases affected by climate change.
- all quantitative estimates of the health effects of climate change were based on observed effects of climate variations either over short time

periods, or between locations. They therefore made the important assumption that these relationships are also relevant to long-term climate change. To avoid unrealistic extrapolation of short-term relationships, we included consideration of mechanisms by which climate-health relationships may alter over time (i.e. adaptation). We considered whether each disease in turn was likely to be significantly affected by biological adaptation (i.e. either behavioural, immunological or physiological) and/or by generally improving socioeconomic conditions (i.e. increasing GDP) (see Table 20.4). In each case we defined appropriate adjustments to the relative risk estimates, in line with published studies. The different factors for each health outcome were then applied to the same projections of future GDP (WHO/EIP/GPE, unpublished data, 2001) and changing climate (from our models), to adjust the relative risks over the time course of the assessment. There is, however, substantial uncertainty over the most likely degree of adaptation under different conditions. This was reflected in

Table 20.4 Assumptions on adaptation and vulnerability

	Biological[a] *adaptation affecting RRs*	*Socioeconomic adaptation affecting RRs*
Direct physiological effects of heat and cold	Yes. Temperature associated with lowest mortality was assumed to change directly with temperature increases driven by climate change	None
Diarrhoea	None	Assumed RR = 1 if GDP per capita rises above US$ 6000/year
Malnutrition	None	Food-trade model assumed future increases in crop yields from technological advances, increased liberalization of trade, and increased GDP[b]
Disasters: coastal floods	None	Model assumed the RR of deaths in floods decreases with GDP, following Yohe and Tol (2002)
Disasters: inland floods and landslides	None	Model assumed the RR of deaths in floods decreases with GDP, following Yohe and Tol (2002)
Vector-borne diseases: malaria	None	None (for RR)

[a] Physiological, immunological and behavioural.

[b] GDP scenarios are developed from EMF14 (Energy Modeling Forum 1995).

the uncertainty estimates for the relative risks for each disease. Quoted uncertainty estimates therefore describe uncertainties around climate change predictions, about current exposure–response relationships, and around the degree to which these are likely to be maintained in the future. Since, to date, these have not been formally modelled, they were generated here by qualitative assessment in collaboration with the original modelling group. The uncertainty estimates should therefore be interpreted with caution.

- we ignored more complex aspects of future vulnerability. Whereas projected trends for average income (which is included in estimating baseline rates and, where possible, relative risks) are broadly positive, other factors may have opposite effects. These include income distribution, maintenance of disease surveillance, control and eradication programmes, technological change and secondary or threshold effects, such as the protective effect of forests in reducing the frequency and intensity of flooding (e.g. Fitzpatrick and Knox 2000).
- we made no attempt to estimate the effects of actions taken specifically to adapt to the effects of climate change (e.g. the upgrading of flood defences specifically to cope with sea-level rise attributable to climate change). Therefore, our estimates represent a "business-as-usual" scenario of the health effects associated with global climate change.

3.3 Estimation for different time points

As stated above, climate model outputs are usually presented as averages over 30-year periods, for example, centred on 2025 and 2055. In order to generate estimates for any specific year, we defined 1990 (i.e. the last year of the baseline climate period) as year 0. Central, lower and upper estimates of relative risks were calculated for the 2020s (i.e. centred on 2025) and the 2050s, using models described elsewhere. Quoted estimates for the years 2001, 2005, 2010 and 2020 were estimated by linear regression against time between 1990 and 2025. Estimates for 2030 were estimated by linear regression between 2025 and 2055. Using this method, our choice of 1990 as the baseline year rather than the middle of the 1961–1990 period led to conservative estimates of health consequences, particularly in the near future.

3.4 Risk reversibility

In the context of this assessment, risk reversibility describes the proportion of the health consequences that would be avoided if the population were shifted to a different exposure scenario. Given the definition of exposure scenarios, complete avoidance of climate change (i.e. populations exposed to baseline climate conditions rather than unmitigated climate change) would avoid all of the health consequences that we have described. Risk reversibility would therefore be 100% in this case.

3.5 Criteria for identifying relevant studies and for estimating strength of evidence

In identifying relevant studies, we have considered all publications reviewed in the IPCC Third Assessment Report (IPCC 2001a), as well as those found in more recent literature searches (Appendix B). However, as the field is relatively new and expanding rapidly, we have also used some studies that are either in-press or submitted for publication (material available on request). As these may not be easily accessible to readers of this report, the major underlying assumptions are described.

There are still relatively few models that link climate change models to quantitative global estimates of health or health-relevant outcomes (e.g. numbers of people flooded or at risk of hunger). Where global models did not exist, we have extrapolated from models that make local or regional projections. Methods for extrapolating to the subregions are described separately for each disease. Where there was more than one published model for a particular outcome, decisions on model selection were based as much as possible on validation against historical/geographical patterns. We excluded models: (i) that have been shown to be significantly less accurate than equivalent models in predicting historical or geographical trends; or (ii) that have been superseded by later models by the same group; or (iii) that are based on unrealistic biological assumptions; or (iv) that cannot be plausibly extrapolated to a wider area. Given the very limited means for model validation, these results involved choices made in this work among models and resulted in very large uncertainty.

As climate change impact assessment is, at this stage, predominantly a model-based exercise, the assessment of strength of evidence for causality is necessarily indirect. It is based on two considerations: (i) the strength of evidence for the current role of climate variability in affecting the health outcome, and (ii) the likelihood that this relationship between climate and health will be maintained through the process of long-term climate change.

Several of the uncertainties involved in this exercise, particularly those relating to climate modelling and those around the quantitative relationships between climate and health, should decrease through improvements in modelling, and more importantly, through better empirical research. However, estimation of climate change effects is currently a predictive exercise, based on some form of indirect modelling (e.g. analysis of temporal or geographical variation in climate), rather than direct experience (i.e. of a change process that has previously occurred). It is therefore possible that some of the health outcomes listed above may not respond in the predicted manner. On the other hand, we do not yet have direct experience of the full range of health outcomes that may be associated with exposure.

The analyses described here were specifically to estimate the future disease burden, which may be avoided by climate change mitigation policies. However, it must be emphasized that these burdens may also be reduced by adaptation interventions to reduce the vulnerability of populations. Given the "exposure commitment" (unavoidable climate change due to past GHG emissions), and the large gap between even plausible mitigation scenarios and the baseline climate, adaptation strategies are essential to the goal of reducing adverse health effects. These include early warning systems and defences for protection from natural disasters, and improved health infrastructure (e.g. water and sanitation, control programmes for vector-borne diseases) to reduce the baseline incidence of climate-sensitive diseases.

3.6 Direct effects of heat and cold on mortality

An association between weather and daily mortality has been shown consistently in many studies in diverse populations. The effect of extreme temperature events (heat-waves and cold spells) on mortality has also been well described in developed countries. However, temperature-attributable mortality has also been found at "moderate" temperatures (Curriero et al. 2000; Kunst et al. 1993).

Causality is supported by physiological studies of the effects of very high or very low temperatures in healthy volunteers. High temperatures cause some well described clinical syndromes such as heat-stroke (see review by Kilbourne 1989). Very few deaths are reported as directly attributed to "heat" (International Classification of Diseases, ninth revision [ICD-9 code 992.0]) in most countries. Exposure to *high* temperatures increases blood viscosity and it is therefore plausible that heat stress may trigger a vascular event such as heart attack or stroke (Keatinge et al. 1986a). Studies have also shown that elderly people have impaired temperature regulation (Drinkwater and Horvath 1979; Kilbourne 1992; Mackenbach et al. 1997; Vassallo et al. 1995). Clinical and laboratory studies indicate that exposure to *low* temperatures causes changes in haemostasis, blood viscosity, lipids, vasoconstriction and the sympathetic nervous system (e.g. Keatinge et al. 1986b, 1989; Khaw 1995; Woodhouse et al. 1993, 1994). The strongest physiological evidence is therefore for cardiovascular disease.

Population-based studies also provide evidence that environmental temperature affects mortality due to both cardiovascular and respiratory disease. The best epidemiological evidence is provided by time-series studies of daily mortality. These methods are considered sufficiently rigorous to assess short-term associations (days, weeks) between environmental exposures and mortality, if adjustment is made for longer-term patterns in the data series, particularly the seasonal cycle and any long-term trends, such as gradually decreasing mortality rates over decades (Schwartz et al. 1996).

The effect of a "hot" day is apparent for only a few days in the mortality series; in contrast, a "cold" day has an effect that lasts up to two weeks. Further, in many temperate countries, mortality rates in winter are 10–25% higher than death rates in summer. The causes of this winter excess are not well understood (Curwen and Devis 1988). It is therefore plausible that different mechanisms are involved and that cold-related mortality in temperate countries is related in some part to the occurrence of seasonal respiratory infections.

Although the physiological evidence for causality of an effect of temperature on mortality is greatest for cardiovascular, followed by respiratory, diseases, temperature has been shown to affect all-cause mortality in areas where cardiovascular disease rates are relatively low and infectious disease mortality is relatively high. Many studies report the seasonal patterns of infectious disease in developing countries, but the role of temperature is not well described. It is likely that seasonal rains influence the seasonal transmission of many infectious diseases. High temperatures encourage the growth of pathogens and are associated with an increased risk of diarrhoeal disease in poorer populations (see section 3.7).

Studies have also described mortality and morbidity during extreme temperature events (heat-waves). However, these events are, by definition, rare. It is therefore difficult to compare heat-waves in different populations and for different intensities. The studies that have been used to describe the effects of heat-waves (episode analyses) also use different methods for estimating the "expected" mortality, which makes comparison difficult (Whitman et al. 1997). An assessment of the health consequences of climate change on thermal stress requires estimation of past and future probabilities of extreme temperature events. The current methods of assessment use 30-year averages of monthly data, and few scenarios consider change in the frequency or magnitude of extreme events. This is because: (i) suitable methods have not been developed, and (ii) climate model output at the appropriate spatial and temporal resolution is not readily available (Goodess et al. 2001).

There is little published evidence of an association between weather conditions and measures of morbidity such as hospital admissions or primary care consultations (Barer et al. 1984; Ebi et al. 2001; Fleming et al. 1991; McGregor et al. 1999; Rothwell et al. 1996; Schwartz et al. 2001). A study of general practitioner consultations among the elderly in Greater London found that temperature affected the rate of consultation for respiratory diseases but not that for cardiovascular diseases (Hajat et al. 2001). However, it is not clear how these end-points relate to quantitative measures of health burden.

Estimating the temperature–mortality relationship

A review of the literature was undertaken to identify studies that report relationships between daily temperature and mortality (see Appendix B).

The following criteria were used to select studies for deriving the modelled estimates.

- A study that uses daily time-series methods to analyse the relationship between daily mean temperature and mortality.
- A study that reports a coefficient from linear regression which estimates the percentage changes in mortality per degree centigrade changes in temperature above and below a reported threshold temperature.

Several studies have estimated future temperature-related mortality for a range of climate scenarios (e.g. Guest et al. 1999; Kalkstein and Greene 1997; Langford and Bentham 1995; Martens 1998a). These methods were not considered appropriate for this project, as described in Appendix B.

The best characterized temperature–mortality relationships are those for total mortality in temperate countries. Fewer studies have also looked at the particular causes of death for which physiological evidence is strongest: cardiovascular disease, and to a lesser extent respiratory disease. In this study, we used the specific relationships for cardiovascular disease where these were available (temperate and cold-climate zones), and the general relationships for all-cause mortality for climatic regions where such disease-specific relationships could not be found in the literature review (all tropical populations) (Table 20.5).

As outlined in section 2.2, it was assumed that everybody is exposed to the ambient temperatures prevailing under the different climate scenarios. However, populations differ in their responses to temperature variability, which is partly explained by location or climate.

The global population distribution was divided into five climate zones (Table 20.6), according to definitions developed for urban areas by the

Table 20.5 Temperature-related mortality: summary of exposure–response relationships, derived from the literature[a]

		Medical all-cause mortality[b]		*Cardiovascular mortality*	
Climate zone	*Threshold (T_{cutoff})*	*Heat*	*Cold*	*Heat*	*Cold*
Hot and dry	23	3.0	1.4	—	—
Warm humid	29	5.5	5.7	—	—
Temperate	16	NA	NA	2.6	2.9
Cold	16	NA	NA	1.1	0.5

— No data.

NA Not applicable.

[a] Change in mortality per 1 °C change in mean daily temperature (%).

[b] Excludes external causes (deaths by injury and poisoning).

Table 20.6 Climate zones

Zone	*Climate definition*	*% of world population in zone (1990s)*	*City from which representative daily temperature distribution was derived*	*Mean temperature (°C) (5th–95th percentile)*
Hot/dry	Temperature of warmest month >30 °C	17	Delhi	25.0 (13.5–35.2)
Warm/humid	Temperature of the coldest month >18 °C, warmest month <30 °C	21	Chiang Mai	26.3 (21.6–29.5)
Temperate	Average temperature of the coldest month <18 °C and >–3 °C, and average temperature of warmest month >10 °C	44	Amsterdam	9.6 (2.0–17.8)
Cold	Average temperature of warmest month >10 °C and that of coldest month <–3 °C	14	Oslo	5 (–6.3–16.5)
Polar	Average temperature of the warmest month <10 °C	0.2	NA	NA

NA Not applicable.

Australian Bureau of Meteorology (BOM 2001). The population in the polar zone is small (0.2% of world population) and was excluded.

It was necessary to estimate daily temperature distributions in order to calculate the number of attributable deaths. Daily temperature distributions clearly vary a great deal, even between localities within the same country. However, it was not feasible within this assessment to obtain sufficient meteorological data to estimate daily temperature distributions throughout the world at a fine spatial resolution. Therefore, a single distribution was chosen to represent each climate zone. New daily temperature distributions were then estimated for each climate scenario, by shifting the currently observed temperature distributions by the projected change in mean temperatures for each month, and of the variability of daily temperatures as well as changes in the mean.

Estimating temperature-attributable mortality

An exposure–response relationship and threshold temperature (T_{cutoff}) was applied within each climate zone (Table 20.7). The average temperature difference above (hot days) and below (cold days) this temperature was calculated for baseline climate and each of the climate scenarios.

The short-term relationships between daily temperature and mortality (Table 20.5) were used to estimate the annual attributable fraction of deaths due to hot days and cold days for each of the climate

Table 20.7 Threshold T_{cutoff} for each scenario in each climate zone (original $T_{cutoff} + Dt_{summer}$ rounded to integers)

Climate zone	*Baseline*	*BaU 2020s*	*BaU 2050s*	*S550 2020s*	*S550 2050s*	*S750 2020s*	*S750 2050s*
Cold	16	17	18	17	17	17	17
Temperate	16	17	19	17	17	17	18
Warm/humid	29	29	31	29	29	29	30
Hot/dry	23	25	26	24	25	24	25

BaU Business-as-usual scenario.

scenarios (i.e. annual temperature distributions based on averages over 30 years). Deaths attributable to climate change were calculated as the change in proportion of temperature-attributable deaths (i.e. heat-attributable deaths plus cold-attributable deaths) for each climate scenario compared to the baseline climate. The $1° \times 1°$ resolution grid map of population distribution (Bos et al. 1994) was then overlaid on the maps of climate zones in a geographical information system (GIS), to estimate the proportion of the population in each subregion who live in each climate zone. The proportional changes in temperature-attributable deaths were therefore calculated by taking the average of the changes in each climate zone represented in the subregion, weighted by the proportion of the subregion's population living within that climate zone.

Adaptation

Acclimatization includes autonomous adaptation in the individual (physiological adaptation, changes in behaviour) and autonomous and planned population-level adaptations (public health interventions and changes in built environment). Acclimatization to warmer climate regimes is likely to occur in individuals and populations, given the rate of change in *mean* climate conditions currently projected by climate models. However, it is uncertain whether populations are able to adapt to non-linear increases in the frequency or intensity of daily temperature extremes (heat-waves). Even small increases in average temperature can result in large shifts in the frequency of extremes (IPCC 2001b; Katz and Brown 1992).

Few studies have attempted to incorporate acclimatization into future projections of temperature-related mortality (Kalkstein and Greene 1997), but all studies report that acclimatization would reduce potential increases in heat-related mortality. Our estimates incorporated an assumption regarding acclimatization of the populations to the changing climate that describes this reduced effect. We assumed that the threshold temperature (T_{cutoff}) is increased as populations adapt to a new

climate regime, reflecting physiological and behavioural acclimatization that can take place over the time-scale of decades. Changes in T_{cutoff} are region and scenario specific, as they reflect the rate of warming experienced. Therefore, they were assumed to be proportional to the projected change in average summer temperature (Dt_{summer}, equal to the mean of the three hottest months) from the climate scenario. The temperature–mortality relationships were assumed not to change over time; that is, populations biologically adapt to their new average temperatures, but remain equally vulnerable to departures from these conditions. We made no explicit adjustment for an effect of socioeconomic development and technological change on temperature-related mortality. The resulting relative risks are given in Table 20.8.

Short-term mortality displacement

Evidence suggests that the increase in mortality caused by high temperatures is partially offset by decreased deaths in a subsequent "rebound" period (Braga et al. 2001). This indicates that some of the observed increase in heat-related mortality may be displacement of deaths among those with pre-existing illness, which would have occurred soon in any case. However, this effect has not been quantified for temperature exposures and was not included in the model. The estimates are therefore used to calculate only attributable deaths, but not DALYs, as the estimate of attributable years of life lost was highly uncertain.

In subregions with predominantly temperate and cold climates, reductions in cold-related mortality are likely to be greater than increases in heat-related mortality. Therefore, all climate scenarios show a net benefit on mortality in these subregions, consistent with the IPCC conclusions described above. The effect of the adaptation assumption is to reduce relative risks and therefore net mortality.

Uncertainty estimates

The principal uncertainty in these estimates, and for all other health effects of climate change, relates to the extrapolation of a short-term climate-health relationship to the long-term effects of climate change. The degree to which this is a reasonable extrapolation relates to the degree to which populations will adapt to changing temperatures, both in terms of reducing the additional mortality attributable to heat, and the possible benefits of avoiding cold deaths. This outcome is unusual, in that the predicted health effects of climate change are negative in some regions, but positive in others. We therefore use slightly different terminology to describe the range of uncertainty around the estimates. The mid-range estimate was given by applying the model described above (i.e. making an adjustment for biological adaptation). The "high-impact" estimate assumes that there was no physiological or behavioural adaptation, and therefore no change in the dose–response relationship over time. This maximizes both positive and negative effects. The low-impact

Table 20.8 Central, low and high estimates of the relative risk of cardiovascular disease (all ages) for alternative climate scenarios relative to baseline climate[a]

		2000			*2001*			*2005*			*2010*			*2020*			*2030*		
Subregion	*Climate*[b]	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
AFR-D	2	1.001	1.000	1.002	1.001	1.000	1.002	1.002	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.006	1.004	1.000	1.008
	3	1.001	1.000	1.003	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005	1.004	1.000	1.008	1.005	1.000	1.009
	4	1.002	1.000	1.004	1.002	1.000	1.004	1.003	1.000	1.005	1.004	1.000	1.007	1.005	1.000	1.011	1.007	1.000	1.013
AFR-E	2	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.002	1.002	1.000	1.003	1.002	1.000	1.005	1.003	1.000	1.006
	3	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.006	1.003	1.000	1.007
	4	1.001	1.000	1.003	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005	1.004	1.000	1.008	1.005	1.000	1.010
AMR-A	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.999
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.999
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.999
AMR-B	2	1.001	1.000	1.001	1.001	1.000	1.001	1.001	1.000	1.002	1.001	1.000	1.002	1.002	1.000	1.003	1.002	1.000	1.004
	3	1.001	1.000	1.001	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005
	4	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.006	1.004	1.000	1.007
AMR-D	2	1.001	1.000	1.001	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005
	3	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005	1.003	1.000	1.007
	4	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.002	1.000	1.005	1.004	1.000	1.007	1.005	1.000	1.009
EMR-B	2	1.000	1.000	1.001	1.001	1.000	1.001	1.001	1.000	1.001	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004
	3	1.001	1.000	1.001	1.001	1.000	1.001	1.001	1.000	1.002	1.001	1.000	1.002	1.002	1.000	1.004	1.002	1.000	1.004
	4	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005	1.003	1.000	1.007
EMR-D	2	1.000	1.000	1.001	1.001	1.000	1.001	1.001	1.000	1.001	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004
	3	1.001	1.000	1.001	1.001	1.000	1.001	1.001	1.000	1.002	1.001	1.000	1.002	1.002	1.000	1.004	1.002	1.000	1.005
	4	1.001	1.000	1.002	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005	1.003	1.000	1.007

EUR-A	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.999
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.999
EUR-B	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.999
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.999
	4	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.998
EUR-C	2	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.998	0.999	1.000	0.998
	3	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.998	0.999	1.000	0.998
	4	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.998	0.999	1.000	0.997	0.998	1.000	0.997
SEAR-B	2	1.001	1.000	1.002	1.001	1.000	1.003	1.002	1.000	1.004	1.002	1.000	1.005	1.004	1.000	1.007	1.004	1.000	1.009
	3	1.001	1.000	1.003	1.002	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.006	1.004	1.000	1.009	1.005	1.000	1.011
	4	1.002	1.000	1.004	1.002	1.000	1.004	1.003	1.000	1.006	1.004	1.000	1.008	1.006	1.000	1.011	1.007	1.000	1.014
SEAR-D	2	1.001	1.000	1.002	1.001	1.000	1.002	1.002	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.006	1.004	1.000	1.008
	3	1.001	1.000	1.003	1.001	1.000	1.003	1.002	1.000	1.004	1.003	1.000	1.005	1.004	1.000	1.008	1.005	1.000	1.009
	4	1.002	1.000	1.004	1.002	1.000	1.004	1.003	1.000	1.005	1.004	1.000	1.007	1.005	1.000	1.011	1.007	1.000	1.013
WPR-A	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.999
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.999	0.999	1.000	0.999	0.999	1.000	0.999
WPR-B	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

[a] Note that for SEAR-B, there are no suitable studies of the relationship between temperature and cardiovascular disease-specific mortality; they are therefore based on relationships with all-cause mortality.

[b] 2 = s550, 3 = s750, 4 = unmitigated emissions.

estimate assumes complete adaptation to a changing climate, and therefore no change in the relative risk as the climate changes.

There is therefore a need for further time-series studies, applying a standard approach to populations living in as wide a range of climates as possible. Previous analyses have focused mainly on temperate zones, where the winter effects are greatest, potentially causing over-estimation of the reduced burden attributable to climate change (e.g. Martens 1998a). More analyses of temperature–mortality relationships are therefore required in tropical developing countries.

Such studies should attempt to estimate formally the degree to which adaptation may decrease mortality, and to which observed associations between climate and mortality reflect displaced rather than additional deaths. Finally, there is a need for greater investigation of the health burden of morbidity associated with temperature extremes, including, for example, inability to work in extreme temperatures.

3.7 Diarrhoeal disease

Diarrhoeal diseases are highly sensitive to climate, showing seasonal variations in numerous sites (Drasar et al. 1978). This observation is supported by regression analyses of the effects of seasonal and longer-term variation in a limited number of sites (Checkley et al. 2000; Singh et al. 2001). The results of a literature search on the associations between diarrhoeal disease and climate are shown in Appendix B, Table B.2. The climate-sensitivity of diarrhoeal disease is consistent with observations of the direct effects of climate variables on the causative agents. Temperature and relative humidity have a direct influence on the rate of replication of bacterial and protozoan pathogens, and on the survival of enteroviruses in the environment (Blaser et al. 1995). Rainfall may affect the frequency and level of contamination of drinking water (Curriero et al. 2001).

Quantitative relationships can be defined between climate variations and incidence, which can in turn be directly linked to the outputs of global climate change models. There are, however, challenges and uncertainties in estimating the magnitude of effects.

- The sites from which these relationships were defined cover only a small part of the spectrum of global climate variation. Different relationships may apply at higher or lower temperatures.
- The relative importance of different pathogens and modes of transmission (e.g. via water, food, insects or human–human contact) varies between locations, and is heavily influenced by level of sanitation (Black and Lanata 1995). As the pathogens are known to vary in their response to climate (e.g. Cook et al. 1990; Chaudhury et al. 1996), this will cause uncertainty in extrapolating temperature relationships from local studies to other regions with different levels of development.

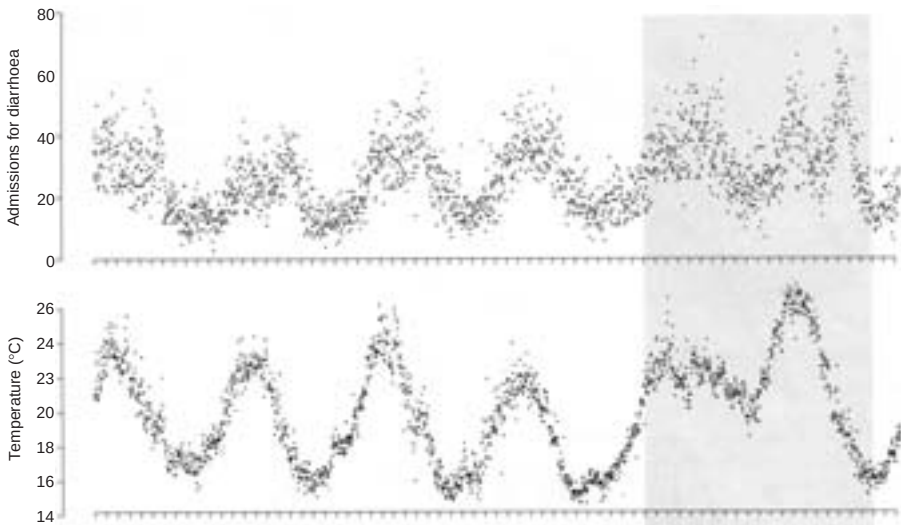
- Pathogens vary in the severity of clinical symptoms, and the likelihood that they will be reported to health services (e.g. Wheeler et al. 1999). Therefore, climate–disease relationships derived only from passive reporting may differ from those based on other methods of surveillance.
- While several studies describe climate effects on particular diarrhoea pathogens (e.g. Eberhard et al. 1999; Konno et al. 1983; Purohit et al. 1998), these cannot be used directly to estimate effects on diarrhoeal disease without information on: (i) their relative contribution to overall disease incidence, and (ii) equivalent data on climate sensitivity and relative prevalence for all other diarrhoea pathogens.
- Despite convincing evidence on the effect of extreme rainfall on waterborne outbreaks of diarrhoea, even in highly developed countries (Curriero et al. 2001), this cannot easily be generalized to the total burden of diarrhoeal disease without information on the relative contribution of such outbreaks to overall diarrhoea incidence.
- Rainfall effects on overall diarrhoea (where observed) are non-linear, and cannot easily be extrapolated to other regions.

In order to minimize these uncertainties, we restricted our estimates to the effect of increasing temperatures on the incidence of all-cause diarrhoea reported to health services (i.e. without attempting to make separate estimates for different pathogens, transmission routes, severities, or for the more complex associations with rainfall). There are important residual uncertainties related to the extrapolation of temperature relationships from specific study sites to others with different temperature regimes and levels of development, and climate effects on reported diarrhoea compared to the true burden of disease. Projections of increasing frequency of extreme wet seasons are very large (e.g. two to five times increase in the regions analysed; Palmer and Ralsanen 2002), and extreme precipitation is associated with increased diarrhoea in both developed (Curriero et al. 2001) and developing (Singh et al. 2001) countries.

Graded maps of 3.75° longitude by 2.5° latitude resolution showing the change in temperature under the alternative scenarios were overlaid with 0.5°×0.5° resolution maps of predicted population distributions for the 2020s and 2050s in a GIS. The GIS was used to calculate the average change in exposure (temperature) for each population grid-cell.

Although seasonality of diarrhoeal disease is well recognized, the quantitative relationship between climate and overall diarrhoea incidence has only been explicitly measured in two studies. Both studies described relationships with all-cause diarrhoea, that is, specific pathogens were not differentiated. Checkley et al. (2000) used time-series analysis to correlate measurements of temperature and relative humidity against daily hospital admissions at a single paediatric diarrhoeal

Figure 20.4 Hospitalizations for diarrhoea (upper line) correspond closely with temperature (lower line) at a clinic in Lima, Peru



Note: Shaded region corresponds to the 1997–1998 El Niño event.

Source: Checkley et al. (2000).

disease clinic in Lima, Peru (Figure 20.4) for just under 6 years. A total of 57 331 admissions were recorded during the study. The analysis showed a 4% (95% CI 2–5%) increase in admissions for each degree centigrade increase in temperature during the hotter months, and a 12% (95% CI 10–14%) per degree centigrade increase in the cooler months, averaging at an 8% (95% CI 7–9%) per degree centigrade increase over the course of the study. During the 1997–1998 El Niño period there was an additional increase in admissions above that expected on the basis of pre-El Niño temperature relationships, but no association with relative humidity independent of temperature. No rainfall during that period.

In the Checkley et al. study, exposure (climate) data were recorded at local meteorological stations, and can be considered to have negligible measurement error at the population level. The analysis independently controlled for seasonal variations and long-term trends, imparting high confidence to the observed effect of temperature on the outcome recorded. The positive correlation with temperature is also biologically plausible, as a high proportion of diarrhoea cases in many tropical developing countries are caused by bacteria, entamoeba and protozoa (Black

and Lanata 1995), which are favoured by high temperatures. The principle limitations of the study by Checkley et al. are that the outcome recorded may not be representative of climate effects on: (i) less severe disease (i.e. not requiring hospitalization) or more severe disease (diarrhoeal deaths), or (ii) disease in adults rather than children.

Singh et al. (2001) used similar time-series analyses to correlate monthly reported incidence of diarrhoea throughout Fiji against variations in temperature and rainfall, after allowing for the effects of seasonal variation and long-term trend. The study covered the period between 1978 and 1998, with an average of some 1000 reported cases for each of the 228 study months. Reported incidence increased by approximately 3% (95% CI 1.2–5.0%) for each degree centigrade increase in temperature, by 2% (95% CI 1.5–2.3%) per unit increase in rainfall above average rainfall conditions (5×10^{-5} kg/m^2 per min), and by 8% per unit decrease below average conditions. The pattern is supported by a positive geographical correlation between temperature and incidence in 18 Pacific Island countries (Singh et al. 2001). Climate measurements were from a 2.5°×2.5° cell of a global gridded data set corresponding to Fiji. The use of monthly averages of climate data from this large geographical area may not have reflected the full range of climate exposures of the population; this would have introduced random error and decreased sensitivity. Low rainfall may force use of contaminated water, while high rainfall may contaminate water through flooding. The major limitations of this study are lack of a clear clinical or laboratory definition for diarrhoea, and lack of information on the age distribution of cases.

We are not aware of any similar studies of climate effects on all-cause diarrhoea in developed regions, although studies have been carried out on some subsets of total incidence. Bentham (1997) showed that the incidence of food poisoning, usually caused by bacteria, increased by approximately 9% per degree centigrade in England and Wales. Konno et al. (1983) demonstrated a non-linear inverse relationship between rotavirus infection and temperature in Japan. The relative importance of pathogens that thrive at lower temperatures appears to be greater in populations with higher standards of living, who have access to clean water and sanitation and for whom there is no clear and consistent evidence for peaks in all-cause diarrhoea in warmer months. This is in contrast to the situation of less well-off populations, where diarrhoea is usually more common in warmer, wetter months, as well demonstrated by clear summer peaks of diarrhoea in black, but not white, infants in Johannesburg during the 1970s (Robins-Browne 1984).

For this assessment we defined developing countries as those with per capita incomes lower than the richer of the two study countries (Fiji) in the year 2000—approximately US$ 6000/year in 1990 US dollars. For such countries, we applied a dose–response relationship of 5% increase in diarrhoea incidence per degree centigrade increase in temperature, to

both sexes and all age groups. This is consistent with the relationships derived from the two studies described above. We chose 5% rather than the arithmetic mean of the constants from the two studies (5.5%) for two reasons: (i) to avoid giving a false impression of precision based on only two estimates, each with their own confidence intervals, and (ii) in order to be conservative.

Although the confidence intervals around the estimates from the individual studies are relatively small, these clearly cover only a small range of climatic and socioeconomic environments. As described above, Checkley et al. (2000) showed that even in a single socioeconomic setting the temperature dependence of diarrhoeal disease may vary across the temperature range or, less plausibly, with non-climatic seasonal variations. This introduces uncertainty when extrapolating such a single relationship. There is also potential bias: if the temperature-responsiveness is indeed greater at low temperatures, extrapolation of an average value will tend to underestimate effects in areas that are on average colder, and overestimate in hotter regions. However, the average annual temperatures experienced by the residents of the areas to which we extrapolated this relationship was 20.3 °C, as calculated by averaging the temperatures in each 1 °×1 °grid-cell weighted by the population. This is in the mid-range of the temperatures experienced throughout the year in Lima (16–26 °C), and cooler than those in the capital of Fiji (23–27 °C) (World Climate 2002). Our extrapolation tended towards being conservative, but with significant uncertainty. Therefore, we place a wide uncertainty range (0–10%) on this value.

As there are no such studies published for developed regions, we made the assumption that overall diarrhoea incidence in richer countries is insensitive to climate change, that is, 0% (–5 to +5%) change per degree centigrade temperature change. Relative risks for each country under each scenario were calculated by multiplying the projected increase in temperature by the relevant exposure–response value. The quoted estimate for each subregion is the population-weighted average of the relative risks for each country in the subregion (Table 20.9). In addition to changes in baseline diarrhoeal disease over time, we assumed that the climate sensitivity of diarrhoea in developing countries will decrease as they become better off. For projections of relative risks for years after 2000, we used projections of future changes in GDP (WHO/EIP/GPE, unpublished data, 2000) to apply the relationship used above—that is, overall diarrhoea incidence does not respond to temperature in any country that attains a per capita GDP of at least US$ 6000/year. Relative risks for each time point were then calculated as above.

The quantitative estimates in this analysis were highly sensitive to the exposure–response relationship, around which there is substantial uncertainty due to the very small number of analysed time-series. The estimates could be rapidly improved by analysis of climate exposure–response relationships from sites from a wider climatic and

Table 20.9 Central, low and high estimates of the relative risk of diarrhoea for alternative climate scenarios relative to baseline climate

		2000			2001			2005			2010			2020			2030		
Subregion	*Climate*[a]	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
AFR-D	2	1.01	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.04	1.00	1.08	1.05	0.99	1.10
	3	1.02	1.00	1.04	1.02	1.00	1.04	1.03	1.00	1.05	1.04	1.00	1.07	1.05	1.00	1.11	1.06	0.99	1.13
	4	1.02	1.00	1.05	1.02	1.00	1.05	1.03	1.00	1.07	1.05	1.00	1.09	1.07	1.00	1.14	1.08	0.99	1.16
AFR-E	2	1.01	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.04	0.99	1.08	1.05	0.99	1.11
	3	1.02	1.00	1.04	1.02	1.00	1.04	1.02	1.00	1.05	1.03	1.00	1.07	1.05	0.99	1.10	1.06	0.99	1.13
	4	1.02	1.00	1.05	1.03	1.00	1.05	1.03	1.00	1.07	1.04	0.99	1.09	1.06	0.99	1.13	1.08	0.99	1.16
AMR-A	2	1.00	0.99	1.02	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.05	1.00	0.95	1.06
	3	1.00	0.99	1.02	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.05	1.00	0.94	1.06
	4	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.04	1.00	0.94	1.06	1.00	0.93	1.08
AMR-B	2	1.00	0.99	1.02	1.00	0.99	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.04	1.00	0.95	1.05
	3	1.00	0.99	1.02	1.00	0.99	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.05	1.00	0.94	1.06
	4	1.00	0.98	1.02	1.00	0.98	1.03	1.00	0.97	1.03	1.00	0.96	1.04	1.00	0.94	1.06	1.00	0.92	1.08
AMR-D	2	1.01	1.00	1.03	1.01	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.05	1.02	0.98	1.06	1.02	0.96	1.07
	3	1.02	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.05	1.03	1.00	1.06	1.02	0.97	1.07	1.02	0.96	1.08
	4	1.02	1.00	1.04	1.02	1.00	1.05	1.03	1.00	1.06	1.04	1.00	1.08	1.03	0.97	1.09	1.02	0.95	1.10
EMR-B	2	1.01	1.00	1.03	1.01	1.00	1.03	1.02	0.99	1.04	1.02	0.99	1.06	1.00	0.95	1.05	1.00	0.94	1.06
	3	1.01	1.00	1.03	1.01	1.00	1.03	1.02	0.99	1.04	1.02	0.99	1.05	1.00	0.95	1.05	1.00	0.94	1.06
	4	1.02	0.99	1.04	1.02	0.99	1.05	1.03	0.99	1.06	1.04	0.99	1.08	1.00	0.93	1.08	1.00	0.91	1.09
EMR-D	2	1.02	1.00	1.03	1.02	1.00	1.04	1.02	1.00	1.05	1.03	1.00	1.06	1.05	1.00	1.10	1.06	1.00	1.12
	3	1.02	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.05	1.03	1.00	1.07	1.05	1.00	1.10	1.06	1.00	1.13
	4	1.03	1.00	1.05	1.03	1.00	1.06	1.04	1.00	1.08	1.05	1.00	1.10	1.08	1.00	1.15	1.09	1.00	1.19

continued

Table 20.9 Central, low and high estimates of the relative risk of diarrhoea for alternative climate scenarios relative to baseline climate (*continued*)

		2000			2001			2005			2010			2020			2030		
Subregion	*Climate*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
EUR-A	2	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.95	1.05	1.00	0.94	1.06
	3	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.97	1.03	1.00	0.95	1.05	1.00	0.94	1.06
	4	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.04	1.00	0.94	1.06	1.00	0.92	1.08
EUR-B	2	1.01	0.99	1.03	1.01	0.99	1.03	1.01	0.99	1.04	1.02	0.98	1.05	1.02	0.97	1.07	1.01	0.94	1.07
	3	1.01	0.99	1.03	1.01	0.99	1.03	1.01	0.99	1.04	1.02	0.98	1.05	1.02	0.97	1.07	1.01	0.94	1.08
	4	1.01	0.99	1.03	1.01	0.99	1.04	1.02	0.98	1.05	1.02	0.98	1.06	1.02	0.96	1.09	1.01	0.93	1.09
EUR-C	2	1.02	1.00	1.03	1.02	1.00	1.04	1.01	0.98	1.03	1.01	0.98	1.04	1.01	0.96	1.06	1.00	0.94	1.07
	3	1.02	1.00	1.03	1.02	1.00	1.04	1.01	0.98	1.03	1.01	0.98	1.04	1.01	0.96	1.06	1.00	0.94	1.07
	4	1.02	1.00	1.04	1.02	1.00	1.05	1.01	0.98	1.04	1.01	0.97	1.06	1.01	0.95	1.08	1.00	0.92	1.08
SEAR-B	2	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.00	1.03	1.00	0.98	1.03	1.00	0.97	1.04	1.00	0.95	1.05
	3	1.01	1.00	1.03	1.01	1.00	1.03	1.02	0.99	1.04	1.00	0.97	1.04	1.00	0.95	1.05	1.00	0.94	1.06
	4	1.02	1.00	1.04	1.02	1.00	1.04	1.02	0.99	1.05	1.00	0.96	1.04	1.00	0.94	1.07	1.00	0.92	1.08
SEAR-D	2	1.02	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.05	1.03	1.00	1.07	1.05	1.00	1.10	1.06	1.00	1.13
	3	1.02	1.00	1.04	1.02	1.00	1.04	1.03	1.00	1.06	1.04	1.00	1.08	1.06	1.00	1.12	1.07	1.00	1.15
	4	1.03	1.00	1.05	1.03	1.00	1.06	1.04	1.00	1.08	1.05	1.00	1.10	1.07	1.00	1.15	1.09	1.00	1.19
WPR-A	2	1.00	0.99	1.01	1.00	0.99	1.01	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.04	1.00	0.95	1.05
	3	1.00	0.99	1.01	1.00	0.99	1.01	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.96	1.04	1.00	0.95	1.05
	4	1.00	0.98	1.02	1.00	0.98	1.02	1.00	0.97	1.03	1.00	0.96	1.04	1.00	0.94	1.06	1.00	0.93	1.07
WPR-B	2	1.01	1.00	1.03	1.01	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.00	0.96	1.05	1.00	0.95	1.06
	3	1.01	1.00	1.03	1.01	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.00	0.96	1.05	1.00	0.95	1.06
	4	1.02	1.00	1.05	1.03	1.00	1.05	1.03	1.00	1.07	1.05	1.00	1.09	1.00	0.93	1.08	1.01	0.92	1.09

[a] 2 = s550, 3 = s750, 4 = unmitigated emissions.

socioeconomic spectrum. Future studies should also explicitly measure the degree to which economic development and improved levels of sanitation influence vulnerability to the effects of climate variation on diarrhoeal disease.

3.8 Malnutrition

Multiple biological and social factors affect the incidence of malnutrition, but one of the fundamental determinants is the availability of staple foods. Climate change may affect this through the balance of the (broadly negative) effects of changes in temperature and precipitation, and (broadly positive) effects of higher CO_2 levels on yields of major food crops (e.g. IPCC 1996; Rosenzweig and Parry 1994; see also Appendix B, Table B.4). These effects are likely to vary markedly with geography: productivity is projected to increase in higher-latitude producers such as Canada and the United States, but to decrease closer to the equator. The global food trade system may be able to absorb these effects at the global level. However, climate change can be expected to have significant effects on food poverty in some regions, owing to variation in both productivity and in economic capacity to cope (Parry et al. 1999).

Crop models have been validated at 124 sites in 18 countries over a wide range of environments (e.g. Otter-Nacke et al. 1986). However, estimation of changes in food availability (and by inference malnutrition) requires additional analyses of the geographical variation in effects on food production, and of food trade patterns. Only one modelling group (Parry et al. 1999) has integrated a basic physiological model of climate change effects on region-specific crop production with food trade models, in order to make projections of the numbers of people actually at risk of hunger. All results presented here were based on the model described in Parry et al. (1999) and other work by the same group.

As for other potential effects of climate change, there is considerable uncertainty over the degree to which current relationships, such as those between climate and crops, and food trade systems will remain constant over time. The most important uncertainties probably relate to the ability of the world food trade system to adapt to changes in production (Dyson 1999; Waterlow et al. 1998). Although these are the most complete models currently available, they do not describe the likely effect of climate change on more complex pathways, such as animal husbandry, or the relative importance of fruit and vegetable production. These in turn may affect micronutrient (e.g. vitamin A, iodine, iron and zinc) deficiency.

Global distribution of temperature, rainfall and CO_2 were mapped for each of the alternative scenarios, as described above. Climate dose–response relationships have been defined for yields of major grain cereals and soybean, which account for 85% of world cereal exports.

Effects of temperature and precipitation, and the beneficial effects of higher CO_2 levels, have been defined using the IBSNAT-ICASA dynamic crop growth models (IBSNAT 1989). The exposure distributions described above were applied to these crop growth models, and the derived yield functions were extrapolated to other crops and regions on the basis of agro-climatic similarity.

These crop yield estimates are used as inputs for the Basic Linked System world food trade model. This consists of a linked series of 40 national and regional food models, representing food production, the effects of market forces and government policies on prices and trade, and trends in agricultural, economic and technological conditions over time (see Fischer et al. 1988 for a full description).

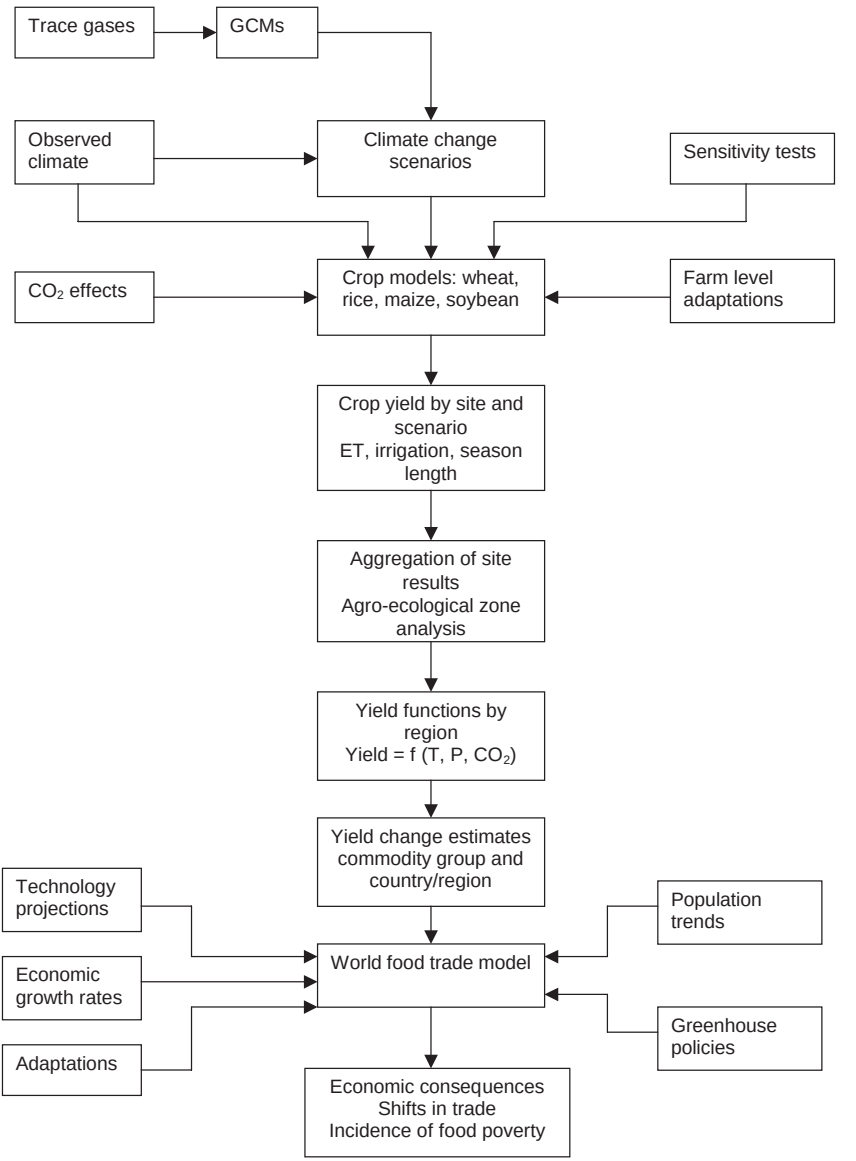
The model is represented schematically in Figure 20.5. Principal characteristics of this model are:

- no major changes in the political or economic context of world food trade or in food production technology;
- population growth to occur following the World Bank mid-range estimate (World Bank 1994) i.e. 10.7 billion by the 2080s;
- GDP to accumulate as projected by EMF14 (Energy Modeling Forum 1995);
- a 50% trade liberalization in agriculture is introduced gradually by 2020.

The model results in an estimation of national food availability. This is used to generate an estimate of per capita food availability in each country, assuming that this food is distributed among the population following a skewed (beta) distribution. The final model output is the number or proportion of the population in each subregion who do not have access to sufficient food to maintain a basal metabolic rate of 1.4, which is the Food and Agriculture Organization of the United Nations' (FAO) definition of undernourishment (FAO 1987).

The model generates outputs for continents made up principally of vulnerable developing countries (thus excluding China and the former Soviet Union, countries in North America, and in western and eastern Europe). As these model continents do not map directly on to the sub-regions, we generated estimates for each subregion by calculating the proportion of the population that lives within each continent in the food availability model. Where more than 90% of the subregion population live within a single model continent, we quoted the model estimate for that continent. Otherwise an average was calculated, weighted by the distribution of the subregion population among the various continents. While the subregions mapped reasonably well on to the climate change/food availability model continents, the aggregation meant that some of the geographical variation in vulnerability was lost. Most notably, despite severe problems in some countries, EUR-C was assumed

Figure 20.5 Key elements of the crop yields and world food trade study



ET Evapotranspiration.

Source: Rosenzweig et al. (1993).

not to suffer from malnutrition as it lies within the "developed" European continent of the food availability model.

Preliminary analysis correlating estimates for the 1990s at the level of the model regions (data not shown) indicated that the model output was positively related to more direct measures of malnutrition, such as the incidence of underweight, and stunting and wasting in children aged <5 years, as measured by the WHO Global Database on Child Growth and Malnutrition (WHO 2002). The aggregation of the food availability model means that this correlation was based on only a small number of independent data points. We therefore did not attempt to make any quantitative estimate of these relationships. Instead, the relative risk of the incidence of energy shortfall (Table 20.10) was interpreted as being directly proportional to the relative risk of suffering from the risk factor "underweight". The relative risk estimates were therefore applied to all diseases affected by the underweight risk factor (see chapter 2). These include diarrhoea and malaria. We therefore assumed that these diseases are affected in two distinct ways by climate change—through meteorological effects on the pathogens and vectors, and through increased susceptibility of the human population due to undernutrition.

In common with most climate change impact assessments, the published studies do not quote uncertainties around the various relationships in the model, either separately or aggregated. Small variations in the initial conditions of a single climate model (the HadCM2 ensemble) generated only slight variations in projections of crop production and incidence of food shortfall. Using a different climate model (HadCM3), however, generated markedly different projections (Parry et al. 1999). These comparisons relate to only part of the possible error, as they do not address uncertainties in the trade and social components of the model.

From the published descriptions of the model, there is no reason to assume that the estimates generated were systematically biased either upward or downward. In the absence of formal sensitivity analyses of the complete model, however, uncertainty estimates are arbitrary. We presented the relative risks generated above as mid-range estimates, with the upper and lower range covering a complete adaptation to any changes in agricultural output (i.e. no change in risk), to a doubling of the estimate of the relative risk calculated above. However, this uncertainty range should be treated with caution. Priorities for the future are investigations of:

- variation in output when using a wider range of food production models as inputs to the food trade/availability models;
- sensitivity of estimates to the various climate scenarios;
- sensitivity analyses to estimate uncertainty around the exposure–response relationships;

Table 20.10 Central, low and high estimates of the relative risk of malnutrition for alternative climate scenarios relative to baseline climate

		2000			2001			2005			2010			2020			2030		
Subregion	*Climate*[a]	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
AFR-D	2	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.03	1.00	1.06
	3	1.01	1.00	1.03	1.01	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.05	1.04	1.00	1.08	1.04	1.00	1.09
	4	1.01	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.04	1.02	1.00	1.04
AFR-E	2	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.03	1.00	1.06
	3	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.04	1.02	1.00	1.05	1.04	1.00	1.07	1.04	1.00	1.08
	4	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.04	1.02	1.00	1.05
AMR-A	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AMR-B	2	1.02	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.05	1.03	1.00	1.06	1.05	1.00	1.09	1.05	1.00	1.10
	3	1.03	1.00	1.07	1.04	1.00	1.07	1.05	1.00	1.10	1.07	1.00	1.13	1.10	1.00	1.20	1.11	1.00	1.22
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00
AMR-D	2	1.02	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.05	1.03	1.00	1.06	1.05	1.00	1.09	1.05	1.00	1.10
	3	1.03	1.00	1.07	1.04	1.00	1.07	1.05	1.00	1.10	1.07	1.00	1.13	1.10	1.00	1.20	1.11	1.00	1.22
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00
EMR-B	2	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.03	1.00	1.06
	3	1.02	1.00	1.04	1.02	1.00	1.04	1.03	1.00	1.06	1.04	1.00	1.08	1.06	1.00	1.12	1.06	1.00	1.13
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00
EMR-D	2	1.02	1.00	1.04	1.02	1.00	1.05	1.03	1.00	1.07	1.04	1.00	1.09	1.07	1.00	1.13	1.07	1.00	1.15
	3	1.03	1.00	1.07	1.04	1.00	1.07	1.05	1.00	1.10	1.07	1.00	1.13	1.10	1.00	1.20	1.11	1.00	1.22
	4	1.02	1.00	1.05	1.03	1.00	1.05	1.04	1.00	1.07	1.05	1.00	1.10	1.07	1.00	1.15	1.08	1.00	1.16

continued

Table 20.10 Central, low and high estimates of the relative risk of malnutrition for alternative climate scenarios relative to baseline climate (*continued*)

		2000			2001			2005			2010			2020			2030		
Subregion	*Climate*[a]	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
EUR-A	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
EUR-B	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
EUR-C	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SEAR-B	2	1.02	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.05	1.03	1.00	1.06	1.05	1.00	1.09	1.05	1.00	1.10
	3	1.03	1.00	1.06	1.03	1.00	1.07	1.04	1.00	1.09	1.06	1.00	1.12	1.09	1.00	1.18	1.10	1.00	1.19
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.00	1.00	1.01
SEAR-D	2	1.04	1.00	1.07	1.04	1.00	1.08	1.06	1.00	1.11	1.07	1.00	1.15	1.11	1.00	1.22	1.12	1.00	1.25
	3	1.05	1.00	1.10	1.06	1.00	1.12	1.08	1.00	1.16	1.10	1.00	1.21	1.16	1.00	1.31	1.17	1.00	1.35
	4	1.05	1.00	1.10	1.06	1.00	1.11	1.08	1.00	1.15	1.10	1.00	1.20	1.15	1.00	1.31	1.17	1.00	1.33
WPR-A	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
WPR-B	2	1.00	1.00	1.01	1.00	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.00	1.02
	3	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.03	1.03	1.00	1.05	1.03	1.00	1.05
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00

[a] 2 = s550, 3 = s750, 4 = unmitigated emissions.

- back-casting of climate/hunger relationships for verification of model accuracy;
- a finer geographical (e.g. country level) breakdown of the outputs of malnutrition models;
- correlations between model outputs and health outcomes at a high spatial resolution; and
- investigation of the synergistic effects of water availability and poverty on malnutrition.

3.9 Disasters caused by extreme weather events: coastal floods, inland floods and landslides

Natural disasters are ultimately a function of both the average and degree of variability of weather conditions, which are further modulated by multiple aspects of population vulnerability such as topography, housing quality and early warning systems (Alexander 1993; McMichael et al. 1996). They are therefore likely to be directly affected by the observed and predicted trend towards increasingly variable weather.

In addition to the axiomatic link between extreme weather events and weather-related deaths and injuries, there is strong statistical evidence that long-term weather cycles (e.g. the ENSO quasi-periodic cycle) correlate with incidence of deaths and injuries attributable to natural disasters (Bouma et al. 1997a; Kovats et al. 1999). The evidence for increased frequency of different categories of extreme events in the past, and the likelihood of changes in the future (Table 20.1), has been strengthened by recent demonstrations of increases in the frequency of large floods during the 20th century (Milly et al. 2002). This evidence is reinforced by projections of several-fold increases in the frequency of what are currently considered extreme wet seasons, for various regions over the world, using a range of climate models (Palmer and Ralsanen 2002).

Based on this, we presented estimates of consequences of increasing frequency of coastal flooding caused by sea-level rise, and inland flooding and landslides caused by increased frequency of extreme precipitation events, which are described by IPCC as very likely (90–99% probability) to increase in many areas under climate change. We did not attempt to estimate the effects of changing frequency and intensity of wind storms, owing to inconsistencies between models (IPCC 2001b) and a lack of quantitative projections of changes in exposure under different climate change scenarios. Our estimates excluded the direct effect of thermal extremes (e.g. heatstroke, increased risk of cardiovascular disease), which are dealt with elsewhere. We also excluded any potential longer-term health consequences arising through mechanisms such as population displacement, economic damage to public health infrastructure, increased risk of infectious disease epidemics and mental illness (Jovel 1989; Menne et al. 1999; WHO 1992).

Climate change is likely to have different effects on the frequency of coastal versus inland floods. Changes in the frequency of coastal floods were defined using published models (Hoozemans and Hulsburgen 1995; Nicholls et al. 1999) that estimate change in sea level for each climate scenario. These changes were applied to topographical and population distribution maps to estimate the change in incidence of exposure to flooding by subregion. The predictions did not account for changes in frequency of storm surges. The global model used here has been shown to be relatively accurate in validations against more detailed assessments at the national level (summarized in Nicholls et al. 1999).

Inland floods and landslides are not affected by sea-level rise, but will be influenced by any increase in the frequency of intense precipitation. Despite the clear causal link, this relationship is poorly researched (Pielke 1999), and has not previously been modelled as a health exposure. There are no published analyses of global relationships between intensity of precipitation, the likelihood of a declared disaster, or the magnitude of health consequences. Clearly, at the local level the frequency of health effects will be determined by the temporal distribution of rainfall (i.e. not only by the average amount of rain over an extended period, but by the peak amount falling in a week, day or hour), and modulated by topography and social aspects of vulnerability (Kundzewicz and Kaczmarek 2000). However, in the absence of detailed data on these variables and their effects, we made the *a priori* assumption that flood frequency is proportional to the frequency with which monthly rainfall exceeds the 1 in 10 year limit (i.e. upper 99.2% CI) of the baseline climate. We also assumed that determinants of vulnerability are distributed evenly throughout the population of a subregion—so that the change in relative risk of health consequences is proportional to the per capita change in risk of experiencing such an extreme event. For each $1° \times 1°$ population grid-cell, we estimated the 99.2% upper CI for the "baseline climate" using means and standard deviations derived from the 1961–1990 averages for each month of the year. Using equivalent data for future scenarios and time points, we estimated the change in frequency with which such a "1 in 10 year event" occurs.

The difference (in standard deviates of the new distribution) between the new mean and the previously defined "1 in 10 year limit" is given by:

$$((X_1 + 2.41 \times u_1) - X_2)/u_2$$

where X_1, u_1=mean and intra-annual deviation from 1961–1990 and X_2, u_2=mean and intra-annual deviation under new scenario.

The probability that this difference will be exceeded in any one month under the new distribution was taken from probability tables. This was divided by the frequency of occurrence under the baseline scenario (=0.008) to give the relative frequency of exceeding the 1 in 10 year limit

for each future scenario. Results were weighted by the population in each cell, and averaged across the countries in each subregion. The final measure of exposure was therefore the relative frequency with which each person in a subregion experiences 1 in 10 year rainfall events.

The process of estimating the disease burden of this change in frequency differs from that for other climate-sensitive health outcomes, as flood effects do not have a specific GBD code. Relative risks should therefore be applied to estimates of health consequences such as deaths and injuries *attributable to these climate events* under baseline climate (i.e. rather than change in total incidence of this outcome).

The EM-DAT database records numbers of deaths and injuries attributed to each natural disaster reported by the media or aid agencies in the last 100 years. Disasters are defined as events that resulted in at least one of the following conditions: (i) >10 people killed, (ii) >200 injured, (iii) a call for international assistance. Although this is the best comprehensive data source available on the current health consequences of natural disasters, all such sources may be subject to underreporting (Noji 1997). Estimates of attributable burden of disease derived from these figures are therefore conservative.

Individual events in the database were classified as inland or coastal floods on the basis of geography, or descriptions of events in the database. Total numbers of deaths for each class of event were summed for subregions. Although the EM-DAT database also records the number of injuries in flooding events, these are not included in this assessment, as they are considered particularly unreliable for floods (Guha-Sapir, personal communication, 2002). The effects of events that could not be identified as inland or coastal were assigned in proportion to the distribution of consequences of classified events in each subregion. The annual incidence of flood death under baseline climate conditions was estimated by dividing the annual average over the last 20 years by the subregional population in 1990.

These baseline incidence rates alter over time, depending on the balance between factors that decrease vulnerability (particularly improving flood defences as populations become richer), and those which increase vulnerability (particularly increasing population density in coastal zones and other flood-prone areas). Baseline estimates for future years were therefore adjusted as far as possible for these effects. For coastal flooding, the effects of projected changes in population distribution in relation to coastline and improving coastal defences in line with GDP were incorporated in the model of Nicholls et al. (1999). The baseline estimates of the incidence of deaths and injuries in years after 1990 were therefore scaled by the ratio of the model projections for numbers of people flooded in each year compared to that in 1990 (Table 20.11).

Such vulnerability effects have not been explicitly modelled for inland flooding. However, Yohe and Tol (2002) have carried out a cross-

Table 20.11 Annual incidence of deaths per 10 000 000 population caused by coastal floods, in the absence of climate change

Subregion	*1980–1999*	*2000*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	0	0	0	0	0	0	0
AFR-E	0	0	0	0	0	0	0
AMR-A	0	0	0	0	0	0	0
AMR-B	2.00	1.59	1.56	1.43	1.29	1.08	0.96
AMR-D	0.40	0.41	0.41	0.41	0.41	0.41	0.34
EMR-B	0	0	0	0	0	0	0
EMR-D	0	0	0	0	0	0	0
EUR-A	0	0	0	0	0	0	0
EUR-B	0	0	0	0	0	0	0
EUR-C	0.10	0.11	0.11	0.12	0.12	0.14	0.15
SEAR-B	0.10	0.11	0.11	0.11	0.12	0.12	0.11
SEAR-D	1.20	1.39	1.40	1.47	1.54	1.69	1.78
WPR-A	0.10	0.10	0.10	0.10	0.10	0.11	0.11
WPR-B	0.90	0.98	0.99	1.03	1.07	1.15	1.22

Source: EM-DAT (2002) for the period1980–1999. Estimates are based on changing GDP for other time points (see text).

sectional analysis of the effect of per capita income on the incidence of death due to all natural disasters (as reported in the EM-DAT database) for the period 1990–2000. They conclude that increasing wealth has a protective effect, best described by:

$$\text{Ln (proportion of population killed/decade)} = 4.7271 - 0.3858 \text{ (Ln GDP per capita)}$$

The effect of income is marginally non-significant at the 5% level ($P<0.07$), generic to all natural disasters and does not take account of the magnitude of the physical hazard. However, this is likely to introduce noise rather than bias in the relationship, and the relationship represents the only published basis for projection of the protective effects of economic development. The relationship is therefore applied to future projections of GDP (WHO/EIP/GPE, unpublished data, 2000). Our projected baseline incidence of deaths for years after 1990 were scaled by the ratio of the projections of deaths due to all natural disasters in that year divided by the estimate for 1990 (Table 20.12).

Some evidence from studies of a small number of earthquakes (Beinin 1981) and famines (Rivers 1982) suggests that women and young children are more vulnerable than men to the acute effects of natural disasters. However, there are insufficient data to derive subregional estimates

Table 20.12 Annual incidence of deaths per 10 000 000 population caused by inland floods and landslides, in the absence of climate change

Subregion	*1980–1999*	*2000*	*2001*	*2005*	*2010*	*2020*	*2030*
AFR-D	2.7	2.7	2.7	2.7	2.6	2.4	2.2
AFR-E	6.5	6.6	6.6	6.5	6.4	6.0	5.4
AMR-A	2.2	2.1	2.1	2.0	1.9	1.8	1.6
AMR-B	52.2	48.4	48.1	46.6	44.8	41.0	36.9
AMR-D	52.1	49.0	48.7	47.2	45.4	41.6	37.4
EMR-B	14.9	13.8	13.7	13.4	13.1	12.1	11.0
EMR-D	32.2	30.9	30.6	29.5	28.2	25.6	23.0
EUR-A	1.3	1.2	1.2	1.1	1.1	0.9	0.8
EUR-B	8.9	9.2	9.1	8.7	8.2	7.4	6.6
EUR-C	1.2	1.4	1.4	1.3	1.2	1.1	1.0
SEAR-B	9.9	8.2	8.1	7.5	6.8	5.8	5.1
SEAR-D	20.3	17.7	17.5	16.5	15.4	13.5	12.0
WPR-A	3.7	3.3	3.3	3.0	2.7	2.3	2.0
WPR-B	13.8	10.6	10.4	9.6	8.7	7.4	6.5

Source: EM-DAT (2002) for the period 1980–1999. Estimates are based on changing GDP for other time points (see text).

of the relative vulnerability of different age groups and sexes to the consequences of flooding: we therefore made the assumption, that all age groups are equally at risk.

The models presented here for coastal flooding assumed that protection evolves over time in proportion to projected increases in GDP. The mid-range estimates presented therefore incorporated an effect of increasing wealth, not only in the baseline estimates, but assumed the same proportional change in the relative risks (i.e. as described by Yohe and Tol 2002). This accounted for the effect of increasing wealth not only in reducing the likely health consequences of "baseline" (i.e. climate change independent floods), but also providing better adaptive capacity for increases driven by climate change. The mid-range estimates did not include any further adjustments for biological/behavioural adaptation to increased flood risk (Table 20.13).

As for other health outcomes of climate change, the only published sensitivity analyses relate to an unmitigated emissions scenario applied to the HadCM2 model with four slightly varying sets of initial conditions, and a comparison with the same emissions scenario run on the HadCM3 model. These resulted in almost no difference in model outputs over the time-scale considered in this assessment. Uncertainties in the model relate to the degree and manner to which individuals respond to

Table 20.13 Central, low and high estimates of the relative risk of death in coastal floods for alternative climate scenarios relative to baseline climate

		2000			*2001*			*2005*			*2010*			*2020*			*2030*		
Subregion	*Climate*[a]	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
AFR-D	2	1.07	1.04	1.14	1.07	1.04	1.15	1.09	1.05	1.18	1.11	1.05	1.21	1.13	1.06	1.26	1.44	1.22	1.89
	3	1.07	1.04	1.15	1.08	1.04	1.16	1.10	1.05	1.19	1.11	1.06	1.23	1.14	1.07	1.27	1.48	1.24	1.96
	4	1.10	1.05	1.20	1.11	1.05	1.21	1.13	1.06	1.25	1.15	1.07	1.30	1.18	1.09	1.36	1.64	1.32	2.29
AFR-E	2	1.06	1.03	1.12	1.06	1.03	1.13	1.07	1.04	1.15	1.08	1.04	1.17	1.09	1.05	1.19	1.12	1.06	1.25
	3	1.06	1.03	1.13	1.07	1.03	1.14	1.08	1.04	1.16	1.09	1.04	1.18	1.10	1.05	1.20	1.13	1.07	1.27
	4	1.09	1.04	1.17	1.09	1.04	1.18	1.10	1.05	1.21	1.12	1.06	1.23	1.13	1.07	1.27	1.18	1.09	1.35
AMR-A	2	1.03	1.02	1.06	1.04	1.02	1.07	1.06	1.03	1.11	1.08	1.04	1.16	1.12	1.06	1.25	1.13	1.06	1.25
	3	1.03	1.02	1.07	1.04	1.02	1.08	1.06	1.03	1.12	1.09	1.04	1.17	1.13	1.07	1.27	1.14	1.07	1.27
	4	1.05	1.03	1.10	1.06	1.03	1.12	1.09	1.04	1.17	1.12	1.06	1.24	1.18	1.09	1.37	1.19	1.09	1.38
AMR-B	2	1.24	1.12	1.48	1.27	1.13	1.54	1.38	1.19	1.75	1.52	1.26	2.04	1.84	1.42	2.69	1.90	1.45	2.81
	3	1.26	1.13	1.51	1.29	1.14	1.57	1.40	1.20	1.80	1.55	1.28	2.11	1.90	1.45	2.80	1.96	1.48	2.93
	4	1.34	1.17	1.68	1.38	1.19	1.75	1.53	1.26	2.05	1.73	1.36	2.46	2.18	1.59	3.37	2.27	1.64	3.54
AMR-D	2	2.07	1.53	3.14	2.15	1.58	3.30	2.45	1.73	3.91	2.77	1.89	4.55	3.27	2.14	5.55	3.58	2.29	6.17
	3	2.14	1.57	3.28	2.23	1.61	3.45	2.55	1.77	4.10	2.89	1.94	4.78	3.42	2.21	5.84	3.76	2.38	6.52
	4	2.50	1.75	4.00	2.62	1.81	4.23	3.04	2.02	5.08	3.49	2.24	5.97	4.19	2.59	7.38	4.64	2.82	8.28
EMR-B	2	1.14	1.07	1.29	1.16	1.08	1.32	1.22	1.11	1.45	1.31	1.16	1.63	1.52	1.26	2.03	1.53	1.27	2.06
	3	1.15	1.08	1.31	1.17	1.08	1.34	1.24	1.12	1.48	1.33	1.17	1.67	1.55	1.28	2.10	1.57	1.28	2.13
	4	1.20	1.10	1.40	1.22	1.11	1.45	1.31	1.16	1.63	1.44	1.22	1.88	1.72	1.36	2.45	1.75	1.37	2.50
EMR-D	2	1.53	1.27	2.07	1.57	1.28	2.14	1.68	1.34	2.36	1.79	1.39	2.58	1.94	1.47	2.88	3.01	2.01	5.02
	3	1.57	1.28	2.14	1.61	1.30	2.21	1.73	1.36	2.45	1.84	1.42	2.68	2.00	1.50	3.00	3.18	2.09	5.36
	4	1.75	1.38	2.50	1.80	1.40	2.59	1.96	1.48	2.91	2.11	1.55	3.21	2.32	1.66	3.63	3.91	2.46	6.82

EUR-A	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.09	1.04	1.18
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.10	1.05	1.20
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.14	1.07	1.29
EUR-B	2	1.57	1.28	2.13	1.64	1.32	2.27	1.95	1.48	2.91	2.45	1.73	3.91	4.05	2.52	7.09	4.78	2.89	8.55
	3	1.60	1.30	2.21	1.68	1.34	2.36	2.02	1.51	3.03	2.55	1.77	4.10	4.24	2.62	7.49	5.02	3.01	9.05
	4	1.79	1.40	2.59	1.89	1.45	2.79	2.34	1.67	3.68	3.04	2.02	5.08	5.27	3.14	9.54	6.31	3.65	11.61
EUR-C	2	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.01	1.02	1.02	1.01	1.03	1.02	1.01	1.04	1.03	1.01	1.06
	3	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.01	1.03	1.02	1.01	1.03	1.02	1.01	1.05	1.03	1.02	1.06
	4	1.01	1.01	1.02	1.01	1.01	1.03	1.02	1.01	1.03	1.02	1.01	1.04	1.03	1.02	1.06	1.04	1.02	1.08
SEAR-B	2	1.11	1.06	1.22	1.12	1.06	1.24	1.15	1.08	1.31	1.19	1.09	1.38	1.24	1.12	1.49	1.28	1.14	1.56
	3	1.12	1.06	1.24	1.13	1.06	1.26	1.16	1.08	1.33	1.20	1.10	1.40	1.26	1.13	1.52	1.30	1.15	1.59
	4	1.16	1.08	1.32	1.17	1.09	1.34	1.22	1.11	1.43	1.26	1.13	1.53	1.34	1.17	1.68	1.39	1.20	1.78
SEAR-D	2	1.01	1.00	1.02	1.01	1.01	1.02	1.01	1.01	1.03	1.02	1.01	1.03	1.02	1.01	1.04	1.03	1.01	1.05
	3	1.01	1.01	1.02	1.01	1.01	1.02	1.01	1.01	1.03	1.02	1.01	1.03	1.02	1.01	1.04	1.03	1.01	1.05
	4	1.01	1.01	1.03	1.01	1.01	1.03	1.02	1.01	1.04	1.02	1.01	1.04	1.03	1.01	1.05	1.04	1.02	1.07
WPR-A	2	1.01	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.01	1.03	1.02	1.01	1.04	1.03	1.01	1.06
	3	1.01	1.00	1.01	1.01	1.00	1.01	1.01	1.01	1.02	1.02	1.01	1.03	1.02	1.01	1.05	1.03	1.02	1.06
	4	1.01	1.00	1.02	1.01	1.01	1.02	1.02	1.01	1.03	1.02	1.01	1.04	1.03	1.02	1.06	1.04	1.02	1.09
WPR-B	2	1.01	1.01	1.02	1.01	1.01	1.03	1.02	1.01	1.04	1.02	1.01	1.05	1.03	1.02	1.06	1.04	1.02	1.07
	3	1.01	1.01	1.03	1.01	1.01	1.03	1.02	1.01	1.04	1.02	1.01	1.05	1.03	1.02	1.07	1.04	1.02	1.08
	4	1.02	1.01	1.04	1.02	1.01	1.04	1.03	1.01	1.05	1.03	1.02	1.07	1.04	1.02	1.09	1.05	1.03	1.10

[a] 2 = s550, 3 = s750, 4 = unmitigated emissions.

the increased risk (Hoozemans et al. 1993). The lower estimates therefore assumed that 90% of the risk could be avoided either by highly efficient coastal defences or individual adaptations. The higher estimates assumed no adaptation either with increasing GDP or individual level measures.

The model for inland flooding was subject to the same uncertainty over adaptive responses. Both the baseline and relative risks were assumed to change with GDP, as described for coastal flooding. As outlined above, however, the uncertainty is also greater for a hazard driven by the magnitude and temporal variation of precipitation (which varies considerably between climate models), rather than the more predictable process of temperature-driven sea level rise. This consideration is particularly important as only one climate model was used in this assessment. Although the relative risks estimated by our method (e.g. 5.53 for our estimates for EUR-A by the year 2030 under an unmitigated emissions scenario) were broadly comparable to estimates of changes in frequency of extreme wet seasons generated using multi-climate model analyses (fivefold for northern Europe for the period 2060–2080; Palmer and Ralsanen 2002), more formal analyses would be necessary to give more accurate estimates. We therefore gave a larger uncertainty range around these predictions than for coastal flooding, by assuming a 50% greater exposure and no adaptation with GDP for the high estimate, and no increase in risk under any scenario for the lower estimate (Table 20.14). As for the other outcomes, this uncertainty range should be interpreted with caution.

The potential health consequences of changing frequency and intensity of extreme weather events are surprisingly poorly researched. Substantial improvements in assessment could be made through better estimates of the current health impacts of natural disasters, which suffer from poor baseline data and probably severe underreporting, particularly in developing countries (Noji 1997). Analyses could also be greatly improved by geo-referencing and more detailed descriptions of disasters to allow differentiation of inland and coastal events, detailed analysis of the relationships between intensity of precipitation and health effects, and projections of future precipitation at higher temporal and spatial resolution, using output from a range of climate models. In order to generate better uncertainty estimates, formal sensitivity analyses of the contributions of each model parameter to the final uncertainty estimates are required.

Finally, it should be stressed that the estimates given here represent the immediate acute consequences of natural disasters, which are likely to be only one component of the total attributable disease burden. Other outcomes of these natural disasters need to be considered, such as the probability of outbreaks of water, vector- and rodent-borne diseases, the effects of sequential disasters on both public health defences and stability of natural ecosystems limiting disease outbreaks (Epstein 1999), and

Table 20.14 Central, low and high estimates of the relative risk of death in inland floods/landslides for alternative climate scenarios relative to baseline climate

		2000			*2001*			*2005*			*2010*			*2020*			*2030*		
Subregion	*Climate*[a]	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
AFR-D	2	1.40	1.00	1.60	1.44	1.00	1.66	1.60	1.00	1.89	1.77	1.00	2.19	2.09	1.00	2.79	2.30	1.00	3.13
	3	1.30	1.00	1.45	1.32	1.00	1.50	1.44	1.00	1.68	1.58	1.00	1.89	1.83	1.00	2.34	1.99	1.00	2.64
	4	1.18	1.00	1.27	1.20	1.00	1.30	1.26	1.00	1.41	1.35	1.00	1.54	1.50	1.00	1.81	1.66	1.00	2.08
AFR-E	2	1.37	1.00	1.54	1.40	1.00	1.60	1.53	1.00	1.83	1.71	1.00	2.10	2.03	1.00	2.64	2.30	1.00	3.18
	3	1.26	1.00	1.41	1.30	1.00	1.45	1.39	1.00	1.60	1.53	1.00	1.81	1.75	1.00	2.22	1.99	1.00	2.65
	4	1.22	1.00	1.33	1.24	1.00	1.36	1.33	1.00	1.50	1.43	1.00	1.66	1.62	1.00	1.99	1.86	1.00	2.44
AMR-A	2	4.19	1.00	6.00	4.66	1.00	6.51	5.76	1.00	8.50	7.33	1.00	11.0	10.5	1.00	16.00	11.5	1.00	18.69
	3	3.62	1.00	5.10	4.00	1.00	5.50	4.90	1.00	7.15	6.19	1.00	9.19	8.77	1.00	13.29	9.66	1.00	15.61
	4	3.02	1.00	4.18	3.34	1.00	4.50	4.03	1.00	5.77	5.03	1.00	7.36	7.03	1.00	10.54	7.99	1.00	12.79
AMR-B	2	1.43	1.00	1.69	1.50	1.00	1.75	1.66	1.00	2.04	1.88	1.00	2.37	2.26	1.00	3.06	2.60	1.00	3.67
	3	1.59	1.00	1.98	1.72	1.00	2.07	1.93	1.00	2.46	2.25	1.00	2.94	2.78	1.00	3.91	3.18	1.00	4.65
	4	1.56	1.00	1.89	1.66	1.00	1.98	1.87	1.00	2.34	2.13	1.00	2.79	2.63	1.00	3.67	3.03	1.00	4.39
AMR-D	2	1.53	1.00	1.83	1.60	1.00	1.92	1.79	1.00	2.25	2.06	1.00	2.65	2.52	1.00	3.48	2.92	1.00	4.20
	3	1.32	1.00	1.50	1.36	1.00	1.56	1.48	1.00	1.75	1.63	1.00	2.01	1.92	1.00	2.50	2.26	1.00	3.10
	4	1.36	1.00	1.57	1.42	1.00	1.62	1.54	1.00	1.84	1.73	1.00	2.13	2.03	1.00	2.70	2.40	1.00	3.33
EMR-B	2	1.61	1.00	1.99	1.71	1.00	2.10	1.98	1.00	2.49	2.29	1.00	2.98	2.83	1.00	3.97	3.20	1.00	4.63
	3	1.85	1.00	2.40	2.01	1.00	2.53	2.37	1.00	3.09	2.82	1.00	3.79	3.57	1.00	5.19	4.04	1.00	6.03
	4	1.89	1.00	2.44	2.05	1.00	2.58	2.41	1.00	3.15	2.88	1.00	3.87	3.64	1.00	5.31	4.04	1.00	6.01
EMR-D	2	2.32	1.00	3.09	2.51	1.00	3.28	3.01	1.00	4.12	3.66	1.00	5.16	4.78	1.00	7.24	5.29	1.00	8.17
	3	2.07	1.00	2.68	2.23	1.00	2.85	2.62	1.00	3.52	3.14	1.00	4.36	4.05	1.00	6.04	4.56	1.00	6.94
	4	2.13	1.00	2.77	2.29	1.00	2.94	2.70	1.00	3.64	3.26	1.00	4.53	4.20	1.00	6.30	4.68	1.00	7.15

continued

Table 20.14 Central, low and high estimates of the relative risk of death in inland floods/landslides for alternative climate scenarios relative to baseline climate (*continued*)

		2000			*2001*			*2005*			*2010*			*2020*			*2030*		
Subregion	*Climate*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
EUR-A	2	2.13	1.00	2.83	2.34	1.00	3.01	2.67	1.00	3.75	3.44	1.00	4.66	3.99	1.00	6.48	5.30	1.00	8.28
	3	2.13	1.00	2.83	2.34	1.00	3.03	2.69	1.00	3.76	3.44	1.00	4.68	4.01	1.00	6.51	5.27	1.00	8.20
	4	2.22	1.00	2.98	2.44	1.00	3.18	2.82	1.00	3.97	3.64	1.00	4.95	4.24	1.00	6.93	5.53	1.00	8.65
EUR-B	2	1.31	1.00	1.44	1.32	1.00	1.48	1.42	1.00	1.66	1.55	1.00	1.89	1.79	1.00	2.32	2.32	1.00	3.22
	3	1.64	1.00	1.92	1.67	1.00	2.01	1.88	1.00	2.38	2.15	1.00	2.83	2.66	1.00	3.75	3.16	1.00	4.65
	4	1.35	1.00	1.53	1.38	1.00	1.57	1.50	1.00	1.78	1.66	1.00	2.04	1.94	1.00	2.56	2.46	1.00	3.45
EUR-C	2	1.33	1.00	1.42	1.30	1.00	1.45	1.39	1.00	1.62	1.52	1.00	1.83	1.75	1.00	2.25	2.45	1.00	3.42
	3	2.05	1.00	2.37	2.00	1.00	2.50	2.26	1.00	3.04	2.68	1.00	3.72	3.49	1.00	5.08	4.31	1.00	6.46
	4	1.70	1.00	1.90	1.66	1.00	1.99	1.84	1.00	2.37	2.11	1.00	2.82	2.67	1.00	3.72	3.45	1.00	5.04
SEAR-B	2	1.33	1.00	1.59	1.43	1.00	1.65	1.56	1.00	1.89	1.71	1.00	2.19	2.01	1.00	2.77	2.51	1.00	3.60
	3	1.61	1.00	2.11	1.81	1.00	2.23	2.04	1.00	2.67	2.34	1.00	3.22	2.89	1.00	4.33	3.57	1.00	5.37
	4	1.30	1.00	1.54	1.40	1.00	1.60	1.50	1.00	1.81	1.65	1.00	2.08	1.92	1.00	2.62	2.39	1.00	3.37
SEAR-D	2	1.21	1.00	1.36	1.26	1.00	1.41	1.34	1.00	1.56	1.45	1.00	1.74	1.65	1.00	2.10	1.73	1.00	2.22
	3	1.10	1.00	1.20	1.14	1.00	1.21	1.19	1.00	1.29	1.24	1.00	1.38	1.33	1.00	1.57	1.39	1.00	1.68
	4	1.05	1.00	1.11	1.08	1.00	1.11	1.09	1.00	1.15	1.13	1.00	1.20	1.18	1.00	1.30	1.21	1.00	1.36
WPR-A	2	1.46	1.00	1.80	1.58	1.00	1.87	1.73	1.00	2.19	1.95	1.00	2.59	2.35	1.00	3.37	2.91	1.00	4.29
	3	1.20	1.00	1.33	1.24	1.00	1.38	1.31	1.00	1.50	1.40	1.00	1.68	1.56	1.00	2.01	2.04	1.00	2.80
	4	1.30	1.00	1.50	1.36	1.00	1.54	1.45	1.00	1.75	1.59	1.00	1.99	1.85	1.00	2.49	2.32	1.00	3.28
WPR-B	2	1.17	1.00	1.35	1.26	1.00	1.38	1.31	1.00	1.51	1.42	1.00	1.69	1.58	1.00	2.04	1.88	1.00	2.50
	3	1.22	1.00	1.41	1.29	1.00	1.45	1.39	1.00	1.62	1.49	1.00	1.83	1.70	1.00	2.23	2.00	1.00	2.70
	4	1.40	1.00	1.60	1.44	1.00	1.66	1.60	1.00	1.89	1.77	1.00	2.19	2.09	1.00	2.79	2.30	1.00	3.13

[a] 2 = s550, 3 = s750, 4 = unmitigated emissions.

longer-term effects such as post-traumatic stress after floods (e.g. Phifer 1990) and of population displacement through coastal flooding. More importantly, it is necessary to expand the range of natural disasters beyond those considered in this chapter. There is an obvious causal chain between apparently increasing variability in precipitation frequency of droughts and their associated health consequences, particularly food shortages and possible famine (UNDMTP 1990; WHO/PTC 1995). While these impacts may be expected to increase under climate change, no models have yet been developed to link climate scenarios, frequency of drought, and associated health effects, so that quantitative estimation of climate change effects is not currently possible. Given the global importance of drought-related disease (WHO/EHA 1998) and the effectiveness of early preparation and rapid response for avoiding health impacts (Gupta 2000; WHO/EHA 2002), this is clearly a priority area for research.

3.10 Vector-borne diseases

Viruses, bacteria, protozoa and helminths transmitted by biting insects and other intermediate hosts are among the most important causes of ill-health in tropical regions (WHO 2000b). The climate sensitivity of such diseases has long been recognized (e.g. Celli 1933), and knowledge of the relationships has been used to help predict epidemics of vector-borne diseases since at least the early years of the last century (e.g. Christophers 1911; MacDonald 1957). More recently, there have been many quantitative studies of the effects of climate variables on the population biology of vectors and pathogens in the laboratory (e.g. reviews by Martens 1998b; Massad and Forattini 1998), and on spatial and temporal variations in vector abundance and disease incidence in the field (review by Kovats et al. 2000a). The results of a literature review are shown in Appendix B, Tables B.5 and B.6.

This climate sensitivity has prompted several studies correlating long-term changes in vector distribution (Lindgren et al. 2000) or disease incidence (e.g. Bouma et al. 1996; Loevinsohn 1994) with local climate trends, which apparently reflect global climate change. However, many of the inferences about resulting health consequences have been called into question, due to concurrent local changes in crucial non-climatic factors, such as human behaviour, disease reporting or control programmes (Hay et al. 2002; Mouchet 1998; Randolph 2001; Reiter 2001).

As for other health effects, it is unlikely that simply correlating long-term trends in disease against trends in climate will soon (or perhaps ever) give unequivocal evidence of the effects of gradual climate change. As climate varies naturally between years, long time-series (i.e. two to three decades) will be needed for statistical tests of association in long-term trends. It is almost inevitable that non-climatic determinants of risk will also change over such long periods, either obscuring or offering

alternative explanations for any effects of climate change. In addition, reliable multi-decadal time-series for vector-borne diseases in developing countries are rare, and it is problematic to interpret analyses of a few data sets as evidence of a general global pattern. Recent reviews therefore suggest that a general effect of climate change on vector-borne disease, while suspected and possible, is highly dependent on other non-climatic factors that act at a small scale (IPCC 2001a; Kovats et al. 2001, Reiter 2001, Reiter et al. 2003).

Despite the practical problems in making direct correlations with recent trends, the extreme climate sensitivity of vector-borne diseases means that it is almost inevitable that they will respond in some way to climate change, in some settings and to some degree. The most reliable basis for estimating such changes should come from information on the relationships between variations in climate and disease in either the past or present. Several studies have used such data to model the effect of predicted climate change on either the distribution of vector-borne diseases, and/or measures of risk within existing or predicted newly endemic areas. The data, techniques and assumptions used in the various analyses are reviewed in detail in later sections.

Whatever the quality of the data and modelling techniques used for predictions, they will remain contingent upon other determinants of disease. Socioeconomic conditions, control programmes, human immunity and the specific combinations of climate variables required by particular vector species or transmission cycles also affect disease incidence, and may be more important than global climate trends, particularly at small spatial scales (Mouchet and Manguin 1999; Randolph et al. 2000; Reiter 2001; Rogers and Randolph 2000; Sutherst 1998).

The IPCC has reviewed the observed and predicted effects of climate variability and change in the context of the other factors listed above (IPCC 2001a). On balance, the IPCC concludes that climate change is likely to expand the geographical distribution of several vector-borne diseases, including malaria, dengue and leishmaniasis to higher altitudes (high confidence) and higher latitudes with limited public health defences (medium/low confidence), and to extend the transmission seasons in some locations (medium/high confidence). For some vector-borne diseases in some locations, climate change may decrease transmission by reductions in rainfall or temperatures too high for transmission (medium/low confidence).

The associations between climate and infectious diseases have also been reviewed by the United States National Research Council (National Research Council 2001b). The review highlights the climate sensitivity of vector-borne diseases, and describes some studies modelling the potential consequences of future climate change on vector-borne diseases. The report re-emphasizes the need for caution when making future projections, and stresses the importance of public health provision in mitigating increases in incidence driven by climate change. No judgement is

made on the likely effects of future climate change on specific diseases, owing to the limited evidence.

Aside from analyses of the potential effect of gradual changes in average climate conditions, predicted increases in climate variability, including possible increases in the frequency and intensity of El Niño events (IPCC 2001b), may also affect vector-borne diseases. There is evidence that the El Niño cycle causes inter-annual variations in disease incidence in several areas (e.g. Bouma et al. 1997b). However, even in these sites, other determinants such as seasonal variations and control programmes exert a larger influence, and epidemic cycles also occur in areas where El Niño has little or no effect on climate, apparently driven by gradual post-epidemic waning of herd immunity (Hay et al. 2000). As the effect of climate variability differs so greatly between sites, global projections of the associations described so far are unlikely to be informative.

Modelling of specific diseases

Most modelling of the effects of climate change has focused on malaria, and to a lesser extent dengue. These are therefore the only vector-borne diseases considered here. Some preliminary modelling work has been carried out on schistosomiasis (Martens et al. 1997), but this is based on a relatively small data set and has not been validated against current distributions. Randolph and Rogers (2000) have also modelled the potential effects of climate change on tick-borne encephalitis (TBE) in Europe, demonstrating that increased temperatures are likely to reduce the endemic range. Although TBE is a relatively small public health problem in global terms, and was not covered in this assessment, this study does demonstrate that climate change may potentially decrease, rather than increase, the transmission of some diseases. Effects on other major vector-borne diseases have been investigated either qualitatively (e.g. Carcavallo and Curto de Casas 1996 for American trypanosomiasis) or in terms of distribution of vectors rather than human disease (e.g. Rogers and Packer 1993 for African trypanosomiasis).

Falciparum malaria

Falciparum malaria is unusual in that several research groups have independently modelled the relationships between climate and disease distribution (Appendix B, Table B.5). The models used can be broadly classified as structural/biological, based on aggregating the effect of climate on the individual components of the disease transmission cycle, or "statistical", derived from direct correlations between geographic or temporal variations in climate, and associated variations in disease incidence or distribution, either in the present or recent past.

Published biological models for the global distribution of falciparum malaria use laboratory data to define the relationship between temperature and the extrinsic incubation period of the parasite, and therefore

the probability of completing development during the lifetime of the mosquito and completing the transmission cycle (Martin and Lefebvre 1995). Later models incorporate temperature effects on the survival probability and biting frequency of mosquitoes (Jetten et al. 1996; Martens et al. 1995a, 1995b, 1999). In these later models, the various temperature-dependent relationships are aggregated into the entomological version of the equation for R_o, the number of cases arising from each new case in a completely susceptible population (Anderson and May 1991; Dye 1992; Garrett-Jones 1964). Because of the lack of data on several key parameters, these are set as biologically plausible constants, allowing the calculation of the critical vector density required for sustainable disease transmission (i.e. $R_o > 1$). This threshold is lower under more suitable (generally warmer) climate conditions. The inverse of the critical density threshold, the "transmission potential", is used as a relative measure of transmission intensity under different climatic conditions. The models also assume a threshold level of transmission potential required to sustain transmission, which allows the identification of areas that are climatically suitable for transmission under both observed and projected climate scenarios.

These studies highlight the extreme climate sensitivity of several stages of the malaria transmission cycle. By aggregating these effects into a single measure related to R_o, they demonstrate that even small temperature increases could potentially cause large relative increases in risk, particularly at the edges of the distribution where temperature may be a limiting factor. They also suggest that those areas climatically suitable for *Plasmodium falciparum* transmission could expand substantially.

While valid for their original purpose as sensitivity analyses for relative changes in risk, these models are not ideal for defining the most probable changes in either geographical distribution or disease burden within endemic areas. Both outputs require the calculation of absolute rather than relative values of R_o, so as to identify areas where $R_o > 1$, allowing disease transmission to persist. In these incomplete biological models, such calculations are partly dependent on parameter values that are arbitrarily defined in the absence of empirical data (Rogers and Randolph 2000). As the models are based on temperature relationships derived from the laboratory, they also rely on the assumption that meteorological station data accurately represent the climatic conditions that mosquitoes and parasites experience in the field—which disregards the possibility that vectors might exploit microhabitats that are very different from those in meteorological stations. Since the outputs from these models have not been validated against current disease distributions, they were not used in this assessment.

In the absence of data to generate complete biological models of all stages in the transmission cycle, an alternative approach is to use statistical relationships to define only the distributional limits of disease. Although this approach does not allow disaggregation of the specific

mechanisms driving the climate-sensitivity of vector-borne diseases, it is generally considered more objective than the use of incomplete biological models, in that model outputs are not dependent on arbitrarily defined parameter values.

The international MARA (Mapping malaria risk in Africa) collaboration generated a model that used a combination of biological and statistical approaches to define the limits of climate suitability for falciparum malaria in Africa (Craig et al. 1999). Laboratory data on the rate of development of falciparum parasites (Detinova 1962) and laboratory and localized field observations of temperature effects on mosquito survival (Haddow 1943; Jepson et al. 1947; Le Sueur 1991; Maharaj 1995) were used to define upper and lower thresholds for mean monthly temperatures, and winter minima, which would allow both mosquito survival and the completion of the parasite extrinsic incubation period during the lifetime of mosquitoes, thereby permitting transmission. Rainfall thresholds were defined by comparing regions with and without stable malaria transmission, which have similar temperature conditions but different precipitation profiles. In order to take account of uncertainty about the precise values of the upper and lower bounds of temperature and rainfall necessary for transmission, climatic conditions near to the thresholds were not defined as either entirely suitable or unsuitable. Instead, they were assigned a probability of suitability between 0 and 1, defined by a "fuzzy membership curve", which is assumed to follow a pre-specified sigmoidal shape between the plausible values for the upper and lower thresholds for each climate variable (e.g. a decreasing probability of suitability between mean temperatures of 32 and 40 °C). These relationships were applied to high-resolution interpolated maps of climate throughout Africa (Hutchinson et al. 1996) to define areas that meet all suitability conditions (i.e. both temperature and rainfall) throughout the continent. For validation, model outputs were visually compared with independent high-resolution maps of the edges of the distribution, based either on field surveys or expert opinion. The model showed a good fit to the observed distributions in both southern Africa and East Africa.

The main advantages of this approach are that the model:

- describes only the cut-offs for any level of transmission, rather than quantitative estimates of transmission risk: it therefore does not rely on arbitrarily defined parameter values to complete the R_o equation;
- represents uncertainty around the edges of the distribution;
- allows description of seasonal patterns of transmission, based on the suitability of individual months; and
- most importantly, has been compared with current and historical distribution maps which are apparently independent of the model building process.

The main caveats are:

- the reliance on laboratory data and a small number of field studies to define climate cut-offs;
- apparent subjectivity in at least one parameter estimate (the proportion of mosquitoes that need to survive the sporogonic cycle in order to maintain transmission, which defines the precise value of the lower temperature cut-off);
- the need to make an assumption about the shape of the "fuzzy membership curve"; and
- lack of systematic empirical validation (validation by visual comparison, rather than calculating diagnostic statistics).

While each of the assumptions can legitimately be questioned, the visual validation suggests that the data and assumptions used are at least reasonably accurate. Again, there are further caveats in using the model to try to describe the true global distribution of falciparum malaria (rather than just climatically suitable areas), either now or in the future. To use such models, it is necessary to make the assumptions that distributions vary directly with climate, without any interactive effect of control programmes, or socioeconomic conditions. The comparisons of model outputs with current data suggest that this is a reasonable assumption for most of sub-Saharan Africa, although control programmes have altered distributions in South Africa. The assumption is much less secure for other endemic regions, which are invariably richer.

The relationship between climate variables and the global distribution of malaria can also be defined in statistical terms. Rogers and Randolph (2000) converted WHO maps (WHO 1997) of the limits of reported malaria distribution in the 1990s into $0.5° \times 0.5°$ resolution grids, coding each cell as either endemic or non-endemic. These grid maps were overlaid on $0.5°$ grid maps describing climate surfaces for the period 1961–1990. A statistical model was generated by randomly selecting a subsample of 50% of the observations of disease presence or absence. First, grid-cells were assigned by applying *k*-means clustering to six groups based on climatic similarity, thereby allowing for potentially different climate–disease relationships in different ecological zones. Stepwise discriminant analysis was applied to find the combination of temperature, rainfall and humidity parameters that gave the greatest statistical differentiation between positive vs negative grid cells within each cluster. The model was then applied to the remaining 50% of the observations to assess model accuracy. The fit of the predicted distributions to the observed WHO malaria maps was found to be significantly better than for previous falciparum malaria models: 77.71% of grid cells correctly predicted for this model vs 75.79% for Martin and Lefebvre (1995) and 67.26% for Martens et al. (1999).

The principal advantage of this approach is that it is entirely data driven. The relationships between global climate and malaria distributions are defined using transparent statistical techniques, applied to both disease and climate data from throughout the globe. They therefore do not require either unsupported assumptions of parameter values in order to complete a biological model, or global extrapolation of data from a limited number of observations. In addition, by mapping the observed distribution rather than climatically suitable areas, such analyses do not require the assumption that distributions are defined only by climate.

Despite these advantages, the quality of the data available at the global scale places several limitations on global statistical models. WHO maps of observed distributions are based on a combination of field observations and expert opinion to draw "inclusive" boundaries of the extremes of distributions. In many cases the maps define large areas as endemic (e.g. all of sub-Saharan Africa north of South Africa, Namibia and Botswana), although significant areas are actually disease free, often apparently due to unsuitable climate. WHO maps (and therefore the statistical relationships) also do not differentiate between malaria caused by *P. falciparum* and *Plasmodium vivax*, which have quite different sensitivities to temperature (Detinova 1962; MacDonald 1957). It is unclear what effect these two simplifications may have on the climate sensitivity of the model.

For this assessment, projected changes in temperature and rainfall under each of the alternative climate scenarios relative to the baseline (1961–1990) climate were mapped at the resolution of the HadCM2 climate model (3.75° longitude by 2.5° latitude). Maps of future climate were then generated by adding these values to maps of baseline climate for the 1961–1990 climate at 0.5° resolution (0.05° resolution for the MARA malaria model in Africa; Hutchinson et al. 1996).

An adapted version of the MARA climate model described in detail above (Tanser et al. 2003) was used to generate the mid-range estimates for this assessment. This decision was based on the independent (though continental rather than global) verification, plus the fact that the model was developed and tested using data from throughout Africa (where the overwhelming majority of the burden of malaria currently occurs, and where accuracy is therefore most important), and the potential for developing projections of increased force of infection and disease incidence in the near future.

The model was applied to the global climate maps for baseline climate, and for the unmitigated emissions scenario for the 2020s and 2050s. Relative risk estimates presented here were the ratios of the projected population at risk (i.e. living in areas climatically suitable for >1 month falciparum malaria transmission per year) in each subregion under climate change, relative to the population at risk under the 1961–1990 climate. As an approximation, estimates for the s750 and s550

scenarios were derived proportionately by multiplying the relative risks under unmitigated climate change by the ratio of global temperature change under each scenario/change under unmitigated emissions (Table 20.15). This model gave considerably larger estimates of changes in population at risk than the statistical model of Rogers and Randolph (2000), which predicted approximately no overall change under an unmitigated emissions scenario by the 2050s. In the absence of further comparisons and formal uncertainty assessments, our lower range estimate therefore included the possibility of no change in risk in any subregion. The upper range estimate is a doubling of the mid-range estimate from the MARA model. We emphasize that given the difficulties in validating any specific model and its suitability for extrapolation to other subregions, the choice of MARA climate model over the other possible models was somewhat arbitrary.

In addition, the calculation of disease burdens requires estimates of change in incidence within each subregion, rather than population at risk. In the absence of models for changes in malaria incidence within endemic regions, we therefore made the assumption that relative changes in incidence will vary in direct relation to predicted changes in population at risk—that is, a doubling of the population at risk within a region will lead to a doubling of the clinical disease incidence.

These measures are related in broad terms: countries or regions with higher populations at risk tend to have higher incidence and disease burdens. However, this relationship is a crude generalization, as it assumes that these relationships will remain constant as the population at risk expands or contracts. This may lead to underestimation of effects, if there is an increase in transmission within already at-risk populations, driving up infection incidence. Alternatively, this relationship may overestimate risk, depending on the extent to which increasing vectorial capacity promotes herd immunity (Rogers et al. 2002), and causes first infections to occur earlier in life, when patients suffer less severe clinical symptoms for some diseases, potentially conferring immunity on the more clinically vulnerable older age-groups (Coleman et al. 2001; Snow and Marsh 1995). In addition, socioeconomic conditions and control programmes clearly influence vector-borne diseases. Future changes in these factors are likely to affect (and hopefully reduce) transmission, as they have done in some regions in the past (e.g. Jetten and Takken 1994; Reiter 2001). The role of adaptation was discussed in section 2.5.

By applying relative risks to baseline incidences of zero in some subregions, our assessment did not allow for the spread of disease from endemic subregions to non-endemic subregions. This is a reasonable, but conservative assumption: non-endemic subregions have better developed health systems and a less amenable socioeconomic environment, in addition to usually being cooler. These factors may protect against re-establishment of vector-borne disease transmission (IPCC 2001b; Kuhn et al. 2003), providing they are maintained.

Table 20.15 Central, low and high estimates of the relative risk of falciparum malaria for alternative climate scenarios relative to baseline climate

		2000			2001			2005			2010			2020			2030		
Subregion	*Climate*[a]	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
AFR-D	2	1.00	1.00	1.01	1.00	1.00	1.01	1.00	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.00	1.02
	3	1.00	1.00	1.01	1.00	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.00	1.02	1.01	1.00	1.03
	4	1.01	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.04	1.02	1.00	1.05
AFR-E	2	1.02	1.00	1.04	1.02	1.00	1.04	1.03	1.00	1.06	1.04	1.00	1.08	1.06	1.00	1.12	1.07	1.00	1.15
	3	1.02	1.00	1.05	1.03	1.00	1.05	1.04	1.00	1.07	1.05	1.00	1.10	1.07	1.00	1.15	1.09	1.00	1.18
	4	1.04	1.00	1.08	1.04	1.00	1.09	1.06	1.00	1.12	1.08	1.00	1.16	1.12	1.00	1.23	1.14	1.00	1.28
AMR-A	2	1.08	1.00	1.15	1.08	1.00	1.17	1.11	1.00	1.23	1.15	1.00	1.30	1.23	1.00	1.46	1.27	1.00	1.53
	3	1.09	1.00	1.19	1.10	1.00	1.20	1.14	1.00	1.28	1.19	1.00	1.37	1.28	1.00	1.56	1.33	1.00	1.65
	4	1.15	1.00	1.29	1.16	1.00	1.32	1.22	1.00	1.44	1.29	1.00	1.59	1.44	1.00	1.88	1.51	1.00	2.03
AMR-B	2	1.02	1.00	1.04	1.02	1.00	1.05	1.03	1.00	1.07	1.04	1.00	1.09	1.07	1.00	1.13	1.08	1.00	1.16
	3	1.03	1.00	1.05	1.03	1.00	1.06	1.04	1.00	1.08	1.05	1.00	1.11	1.08	1.00	1.16	1.10	1.00	1.19
	4	1.04	1.00	1.09	1.05	1.00	1.09	1.06	1.00	1.13	1.09	1.00	1.17	1.13	1.00	1.26	1.15	1.00	1.30
AMR-D	2	1.01	1.00	1.02	1.01	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.05	1.03	1.00	1.07	1.04	1.00	1.09
	3	1.01	1.00	1.03	1.02	1.00	1.03	1.02	1.00	1.04	1.03	1.00	1.06	1.04	1.00	1.08	1.05	1.00	1.10
	4	1.02	1.00	1.04	1.02	1.00	1.05	1.03	1.00	1.07	1.04	1.00	1.09	1.07	1.00	1.13	1.08	1.00	1.17
EMR-B	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
EMR-D	2	1.03	1.00	1.07	1.04	1.00	1.07	1.05	1.00	1.10	1.07	1.00	1.13	1.10	1.00	1.20	1.15	1.00	1.30
	3	1.04	1.00	1.08	1.04	1.00	1.09	1.06	1.00	1.12	1.08	1.00	1.16	1.12	1.00	1.24	1.19	1.00	1.37
	4	1.06	1.00	1.13	1.07	1.00	1.14	1.10	1.00	1.19	1.13	1.00	1.26	1.19	1.00	1.38	1.29	1.00	1.59

continued

Table 20.15 Central, low and high estimates of the relative risk of falciparum malaria for alternative climate scenarios relative to baseline climate (*continued*)

		2000			2001			2005			2010			2020			2030		
Subregion	*Climate*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*	*Mid*	*Low*	*High*
EUR-A	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
EUR-B	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
EUR-C	2	1.07	1.00	1.13	1.07	1.00	1.15	1.10	1.00	1.20	1.13	1.00	1.27	1.20	1.00	1.40	1.25	1.00	1.50
	3	1.08	1.00	1.16	1.09	1.00	1.18	1.12	1.00	1.25	1.16	1.00	1.33	1.25	1.00	1.49	1.31	1.00	1.61
	4	1.13	1.00	1.26	1.14	1.00	1.29	1.19	1.00	1.39	1.26	1.00	1.52	1.39	1.00	1.78	1.48	1.00	1.97
SEAR-B	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SEAR-D	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.01
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.00	1.01
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.01	1.00	1.01	1.01	1.00	1.02
WPR-A	2	1.07	1.00	1.14	1.08	1.00	1.15	1.11	1.00	1.21	1.14	1.00	1.28	1.21	1.00	1.42	1.25	1.00	1.49
	3	1.09	1.00	1.17	1.09	1.00	1.19	1.13	1.00	1.26	1.17	1.00	1.34	1.26	1.00	1.52	1.30	1.00	1.60
	4	1.14	1.00	1.27	1.15	1.00	1.30	1.20	1.00	1.41	1.27	1.00	1.54	1.41	1.00	1.81	1.48	1.00	1.95
WPR-B	2	1.06	1.00	1.12	1.07	1.00	1.14	1.09	1.00	1.18	1.12	1.00	1.25	1.18	1.00	1.37	1.22	1.00	1.43
	3	1.08	1.00	1.15	1.08	1.00	1.17	1.11	1.00	1.23	1.15	1.00	1.30	1.23	1.00	1.45	1.26	1.00	1.53
	4	1.12	1.00	1.24	1.13	1.00	1.26	1.18	1.00	1.36	1.24	1.00	1.48	1.36	1.00	1.71	1.42	1.00	1.83

[a] 2 = s550, 3 = s750, 4 = unmitigated emissions.

Dengue

Several studies have explored the relationship between climate and the distribution of intensity of dengue transmission (see Appendix B, Table B.6). Most are derived from a series of biological models that relate climate variables to determinants of the population biology of *Aedes* vectors (Focks et al. 1993a, 1993b) and dengue transmission (Focks et al. 1995). Adaptations of these models to map climate relationships at the global scale (Jetten and Focks 1997; Martens et al. 1997; Patz et al. 1998) have used field and laboratory data to define the relationships between temperature and the length of the gonotrophic cycle (and the associated feeding frequency), larval weight (and therefore the need to take multiple feeds within a single cycle), and the extrinsic incubation period of dengue virus within the vector. Mosquito survival, human biting habit and the duration of human infectiousness are set as temperature-independent constants, with parameter values defined using field data from a number of sites.

As for the biological models of malaria, these relationships are aggregated into a simplified version of the R_o equation, excluding measures of vector abundance. This equation is again used to define the critical density threshold (the number of mosquitoes which would be required to maintain $R_o>1$), and its inverse, "transmission potential" or "transmission intensity". The model has been applied to local climate data in a series of sites, and the rank order of monthly predicted values of transmission potential showed good correspondence with the observed seasonal distribution of dengue cases (e.g. Pearson's $R=0.837$ in San Juan, Puerto Rico).

These models have similar characteristics to the biological models for malaria. By aggregating temperature effects from various stages of the transmission cycle, they demonstrate that overall dengue risk is likely to be extremely sensitive to even small changes in temperature. Sensitivity analyses show that a 2 °C increase in global temperature would lead to large (commonly 1–5 times) increases in transmission intensity in many regions of the world. However, these models are also subject to the same caveats. As there are no data available on mosquito abundance throughout the globe, or the relationship between vector abundance and climate variables, the transmission potential remains a relative rather than absolute measure of R_o. It is consequently unsuitable for defining the conditions under which transmission can persist, and therefore mapping the distributional limits of dengue either now or in the future. In addition, although the validation exercise demonstrates that transmission potential is correlated to the incidence of infection and clinical disease, in that months with higher transmission potential have more cases, the quantitative relationships have not been explicitly defined. It is therefore difficult to interpret exactly what effect a doubling of transmission potential (for example) would have on the burden of disease caused by dengue.

Methods similar to those used by Rogers and Randolph (2000) for malaria have recently been applied to define the relationship between multiple climate parameters and the reported geographic distribution of dengue. Hales et al. (2002) used WHO data (WHO 2000a) to map reports of dengue transmission during the period 1975–1996 at the level of the country, or smaller administrative area when this was specified in the report. These were converted into grid maps at 0.5° resolution. Logistic regression analysis was then used to correlate presence or absence of reported transmission against the average values of various climate parameters (monthly average rainfall, vapour pressure, and maximum, minimum and mean temperature) for the corresponding grid-cells throughout 1961–1990. Vapour pressure (approximately equivalent to absolute humidity, and reflecting both temperature and precipitation) gave the best discrimination between areas with and without reported dengue transmission, with no other climate variables adding significant explanatory power. The accuracy of the model was assessed by cross-validation, repeatedly using 95% of the data to generate predictions for the remaining 5%. The model gave correct predictions of observed presence or absence for an average of 89–92% of grid-cells, with the precise accuracy depending on the radius over which vapour pressure was considered. The limitations of this model are similar to the statistical model for malaria. To be consistent with the choice and exclusion of models for malaria, the model of Hales et al. (2002), which is similar to the excluded model of Rogers and Randolph (2000), was not used in the estimates of disease burden.

Future research

A comprehensive analysis would require the generation of predictive models for other important vector-borne diseases. There is also a need to investigate the effect of the predicted increases in climate variability (e.g. more frequent or more extreme El Niño events), especially on the frequency and intensity of epidemics.

Most importantly, more reliable estimation would require models that cover all stages of the causal chain from climate change to clinical outcome. These should explicitly address the role of socioeconomics and control in determining absolute, rather than relative, measures of R_o. This would allow better definition both of the geographical limits of transmission, and of variations in the incidence within existing or potentially newly endemic regions. They should also address the role of host-immunity in protecting individuals and populations from clinical disease, as exposure to infection changes. Such models would represent a considerable advance over the assumption made here, that changes in the proportion of the populations at risk within a subregion will be reflected in proportional changes in the burden of disease.

For the purpose of health assessments, the nature of the model (i.e. whether each of the biological processes are modelled separately and

then aggregated, or whether climate is statistically related directly to empirical measures of disease burden) is probably less important than model accuracy. There is a clear need for greater model validation, by comparison of predictions with geographical and temporal patterns of infection and disease in the present and recent past. The most urgent requirement for all of these objectives is the availability of better quality surveillance data on infection and disease, from a wide variety of geographical and socioeconomic settings.

Finally, the potential effects of future climate change on vector-borne diseases (or other diseases) should not divert attention or resources from current control efforts. On the contrary, they provide the additional argument that disease control now should also reduce vulnerability to climate change in the future.

4. Results

Summary measures of the effects of climate change on health are presented in this chapter only for the estimated current effects, using the relative risks obtained by extrapolation of the future predictions, as described in section 2.6. The estimates presented here did not attribute DALYs to cardiovascular deaths due to thermal extremes, and excluded any increase due to dengue (see sections 3.6 and 3.10). They should be interpreted with caution as, in contrast to most other risk factors, they relied on modelled rather than directly observed outcomes. They did, however, indicate the estimated distribution of impacts both among geographical regions and among the various causes of disease (Tables 20.16 and 20.17). The models may also be useful for the secondary purpose of indicating the magnitude of health impacts that might already be caused by climate change, but which may not be detected by direct observation using current surveillance systems.

The various causes considered here differed markedly in their contribution to the estimates of the overall burden of disease. In our analysis, climate-change effects on malnutrition, diarrhoea and vector-borne diseases appeared considerably more important than effects on flooding, or on deaths attributable to thermal extremes. It should be noted that, with the exception of malaria, these outcomes are relatively poorly studied in comparison with the direct effects of thermal stress.

The health consequences of climate change are distributed very unevenly among regions. Estimated DALY burdens *per capita* are several hundred times greater in the poorer regions of Africa, parts of the Eastern Mediterranean region and South-East Asia than in western Europe, North America and the more developed regions of the Western Pacific. This is largely a reflection of the much higher baseline incidence of the most important climate-sensitive diseases (malaria, diarrhoea and malnutrition) in these poorer regions, but also of greater vulnerability to climate change effects. Because these major climate-sensitive diseases

Table 20.16 Estimated mortality (000s) attributable to climate change in the year 2000, by cause and subregion

Subregion	*Malnutrition*	*Diarrhoea*	*Malaria*	*Floods*	*CVD*	*All causes*	*Total deaths/million population*
AFR-D	8	5	5	0	1	19	66.83
AFR-E	9	8	18	0	1	36	109.40
AMR-A	0	0	0	0	0	0	0.15
AMR-B	0	0	0	1	1	2	3.74
AMR-D	0	1	0	0	0	1	10.28
EMR-B	0	0	0	0	0	1	5.65
EMR-D	9	8	3	1	1	21	61.30
EUR-A	0	0	0	0	0	0	0.07
EUR-B	0	0	0	0	0	0	1.04
EUR-C	0	0	0	0	0	0	0.29
SEAR-B	0	1	0	0	1	2	7.91
SEAR-D	52	22	0	0	7	80	65.79
WPR-A	0	0	0	0	0	0	0.09
WPR-B	0	2	1	0	0	3	2.16
World	77	47	27	2	12	166	27.82

CVD Cardiovascular disease. As described in section 3.6, the estimated cardiovascular deaths represent temperature-related mortality displacement. Therefore no disease burden is estimated for deaths from this cause in Table 20.17.

mainly affect younger age groups, the health burden associated with climate change appears to be borne mainly by children rather than adults.

5. Discussion

The collective scientific evidence indicates that anthropogenic climate change has already begun and will continue, with potential consequences for human health. Global warming over the past quarter-century was of the order of half a degree centigrade. Such a gradual change is partly obscured by natural climate variability and affects health through complex causal pathways. These characteristics, coupled with considerably larger effects of other factors in the most vulnerable populations, mean that it is inherently difficult to measure directly net health losses or gains attributable to the climate change that have occurred until now.

However, climate change differs from most other health determinants in that considerable effort has been devoted to generating and evaluating formal models to forecast future climate, in response to likely trajectories of atmospheric gas composition. These models are in general agreement that over the next 50–100 years global warming will be approximately five times greater than has been experienced in the last

Table 20.17 Estimated disease burden (000s of DALYs) attributable to climate change in the year 2000, by cause and subregion

Subregion	*Malnutrition*	*Diarrhoea*	*Malaria*	*Floods*	*All causes*	*Total DALYs/million population*
AFR-D	293	154	178	1	626	2185.78
AFR-E	323	260	682	3	1267	3839.58
AMR-A	0	0	0	4	4	11.85
AMR-B	0	0	3	67	71	166.62
AMR-D	0	17	0	5	23	324.15
EMR-B	0	14	0	6	20	147.57
EMR-D	313	277	112	46	748	2145.91
EUR-A	0	0	0	3	3	6.66
EUR-B	0	6	0	4	10	48.13
EUR-C	0	3	0	1	4	14.93
SEAR-B	0	28	0	6	34	117.19
SEAR-D	1918	612	0	8	2538	2080.84
WPR-A	0	0	0	1	1	8.69
WPR-B	0	89	43	37	169	111.36
World	2846	1459	1018	193	5517	925.35

25 years, with associated changes in other potentially hazardous climate characteristics, such as the frequency of extreme precipitation events.

Such modelling is at a relatively early stage. Few modelling studies have estimated health effects at the global scale, and not all of these directly estimate incidence or prevalence of GBD outcomes. However, they provide the best current basis for making indicative forecasts in order to inform policy decisions. These models nevertheless make only crude adjustments for the effects of other variables (such as decreasing poverty), which may both determine the vulnerability of populations to potential health effects of climate change, and exert much larger independent effects on health. Taking each disease in turn:

1. We estimated a small proportional decrease in cardiovascular and respiratory disease mortality attributable to climate extremes in tropical regions and a slightly larger benefit in temperate regions, caused by warmer winter temperatures. Although these proportional changes are modest, they apply to significant causes of death. Uncertainties around these estimates are largely due to lack of knowledge of the degree to which populations physiologically and behaviourally adapt to increasing temperatures.

2. The relative risk for diarrhoea in 2030 in developing regions was estimated to be between 1 and 1.1 under unmitigated emissions compared with baseline climate. Richer countries (GDP >US$ 6000/year),

either now or in the future, were assumed to suffer little or no additional risk of diarrhoea. Again, these small changes in relative risk relate to a major cause of ill-health. Uncertainties were mainly due to poor characterization of variations in the relationship between climate and diarrhoea in more or less developed regions, which have different balances between pathogens preferring higher or lower temperatures.

3. Estimated effects on malnutrition varied markedly across subregions. By 2030, the relative risks for unmitigated emissions relative to no climate change varied from a large increase (RR=1–1.33) in SEAR-D to a small decrease (RR=1–0.99) in WPR-B. Developed countries were assumed to be immune to climate change effects on malnutrition. There was no consistent pattern of reduction in relative risk with intermediate levels of climate change stabilization. Apparent inconsistency in the estimates may be due to the high sensitivity of the models to regional variations in precipitation, for which future projections are much more uncertain than for temperature. Although these estimates are somewhat unstable, they are relatively large, and again relate to a major disease burden.

4. We estimated much larger proportional changes in the numbers of people killed in coastal floods (RR of up to 6.3 in EUR-B for unmitigated emissions compared to baseline conditions in 2030), but applied to a very low burden of disease. Consequences of inland floods were predicted to increase by a similar order of magnitude (RR= 1–18.5 in AMR-A), and are generally more common. In contrast to most other outcomes, the increase in relative risk tended to be at least as high in developed as developing subregions. However, these apply to baseline rates that *are* much higher in developing than developed countries. Both estimates are subject to uncertainty around the likely effectiveness of adaptation measures. Inland floods are subject to additional large uncertainties around the quantitative relationships between changes in the intra-annual variation in precipitation (on which our model is based), the magnitude and geographical distribution of extreme precipitation events, and in turn the frequency of flooding and its health consequences. The suggestion of a trend towards decreasing incidence with increasing GHG emissions in some regions is probably due to the uncertainties in predicting precipitation trends. As projections for precipitation are less secure than for temperature, mid-range estimates and uncertainties around the effects of inland floods could be much better described using multiple climate models, rather than the single model used in this assessment.

5. We estimated relatively large changes in the relative risk of falciparum malaria in countries at the edge of the current distribution, for example, increases in relative risk of falciparum malaria of between

1 and 1.83 in WPR-B by 2030. Relative changes were much smaller in areas that are already highly endemic for these diseases. The principal uncertainties specific to these estimates related to the reliability of extrapolations made between subregions, the relationship between changes in the population at risk of these diseases, and incidence (and therefore disease burden), and over the degree to which changes in the non-climatic influences on vector-borne diseases could affect not only the baseline rates of disease, but also interact with climate change to affect the relative risks.

The estimates for the year 2000 (Tables 20.16 and 20.17) suggest that, according to models summarizing our current knowledge of the relationships between climate and health, past climate change may already be causing some health consequences.

These relative risk estimates are much greater for projections into the future, as climate change continues. Both current and future estimates show extreme variations in the estimated effects among geographical regions. Negative consequences are overwhelmingly concentrated in the developing regions of the world (particularly in Africa and the poorer regions of the Eastern Mediterranean and South-East Asia). This is partly a function of variation in baseline climate (hot regions suffer more from increases in temperature), but more importantly due to population vulnerability (e.g. developed countries are assumed to be completely immune from some diseases, such as malnutrition).

Global models have not yet addressed all of the likely effects of climate change on health. The potential omissions are many infectious diseases, the health consequences of drought and famine (beyond those included in current estimates of malnutrition), population displacement, destruction of health infrastructure in natural disasters, increased pollution and aeroallergen levels, effects of plant pests and diseases on agriculture, and risk of conflict over declining natural resources. It is likely that these health consequences will be larger than those estimated in this chapter.

Although incomplete and encompassing a wide uncertainty range, the results of these analyses suggested that the attributable burden of climate change is likely to be significant, even under the relatively short (in climatological terms) time-scale considered for the comparative risk assessment (CRA). The effect of plausible reductions in climate change was estimated to be relatively small over this assessment period. However, the health gains would clearly be much greater over longer time periods. Given the long time-lag and apparent irreversibility of climate change, early mitigation should therefore result in long-term health benefits. Our results therefore indicate the urgent need for: (i) consideration of optimal policies to reduce climate change; (ii) strengthening of current actions to control climate-sensitive diseases, both as ends in themselves, and as adaptation to future climate change; and (iii) con-

tinued research to revise and narrow the uncertainty range around the estimates of disease burdens.

This assessment has highlighted the most important gaps in data and understanding that should be addressed for the next global burden of disease assessment. Marked improvements in our assessment of this risk factor would come from:

- the use of multiple climate models;
- climate–health relationships derived from a greater range of climatic and socioeconomic environments;
- more explicit and routine validation of the accuracy of disease models in the present or recent past;
- formal analyses to aggregate uncertainty arising from multiple causes (i.e. GHG emissions scenarios, climate models, climate–health relationships, and effect modifiers);
- efforts to formally model climate change effects through to disease burden, rather than intermediate indicators such as population at risk;
- a greater emphasis on investigating the consequences of increased climate variability, rather than gradual changes in mean conditions; and
- the development of analytical tools to assess outcomes acting through more complex causal mechanisms.

Acknowledgements

To Matthew Livermore, Tim Mitchell, David Viner, Pim Martens for provision of data and advice; Richard Tol for provision of unpublished data; Debarati Guha-Sapir for advice on EM-DAT disaster data; Clive Davies, Paul Coleman, Sandy Cairncross, Bo Draser, Alessandro Loretti, Charles Delacollette, Monika Blössner, Mercedes de Onis, Rosalie Woodruff, Will Checkley; and nine referees, for helpful review comments.

Steering committee

Carlos Corvalan, Annette Prüss-Ustün, Roberto Bertollini, Bettina Menne, Andy Haines and Alistair Woodward.

Note

1 See preface for an explanation of this term.

References

Alexander D (1993) *Natural disasters*. University College London Press, London.

Anderson RM, May RM (1991) *Infectious diseases of humans: dynamics and control*. Oxford University Press, Oxford.

Baker RHA, Sansford CE, Jarvis CH, Cannon RJC, MacLeod A, Walters KFA (2000) The role of climatic mapping in predicting the potential geographical distribution of non-indigenous pests under current and future climates. *Agriculture Ecosystems and Environment*, **82**:57–71.

Barer D, Ebrahim S, Smith C (1984) Factors affecting day to day incidence of stroke in Nottingham. *British Medical Journal*, **289**:662.

Beinin C (1981) An examination of health data following two major earthquakes in Russia. *Disasters*, **5**:142–146.

Bentham G (1997) Health. In: *Economic impacts of the hot summer and unusually warm year of 1995*. Palutikof JP, Subak S, Agnew MD, eds. University of East Anglia, Norwich.

Black RE, Lanata CF (1995) Epidemiology of diarrhoeal diseases in developing countries. In: *Infections of the gastrointestinal tract*. Blaser MJ, Smith PD, Ravdin JI, Greenberg HP, Guerrant RI, eds. Raven Press, New York.

BOM (2001) *Climate zones for urban design*. Available at http://www.bom.gov.au/climate/environ/design/climzone.shtml.

Bos ET, Vu MT, Massiah E, Bulatao RA (1994) *World population projections 1994–1995: estimates and projections with related demographic statistics*. World Bank, Johns Hopkins University Press, Baltimore, MD.

Bouma MJ, Dye C, van der Kaay HJ (1996) Falciparum malaria and climate change in the Northwest Frontier Province of Pakistan. *American Journal of Tropical Medicine and Hygiene*, **55**:131–137.

Bouma MJ, Kovats RS, Goubet SA, Cox JSH, Haines A (1997a) Global assessment of El Niño's disaster burden. *The Lancet*, **350**:1435–1438.

Bouma MJ, Poveda G, Rojas W et al. (1997b) Predicting high-risk years for malaria in Colombia using parameters of El Niño southern oscillation. *Tropical Medicine and International Health*, **2**:1122–1127.

Braga A, Zanobetti A, Schwartz J (2001) A tridimensional estimate of the lag structure between temperature and total daily deaths in 14 US cities. *Epidemiology*, **12**:180.

Broecker WS (1997) Thermohaline circulation, the Achilles heel of our climate system: will man-made CO_2 upset the current balance? *Science*, **278**: 1582–1588.

Campbell-Lendrum DH, Wilkinson P, Kuhn K et al. (2002) Monitoring the health impacts of global climate change. In: *Environmental change, climate and health: issues and research methods*, Martens P, McMichael AJ, eds. Cambridge University Press, Cambridge.

Carcavallo RU, Curto de Casas SI (1996) Some health effects of global warming in South America: vector-borne diseases. *Journal of Epidemiology*, **6**:S153–S157.

Celli A (1933) *A history of malaria in the Roman Campagna from ancient times*. AMS Press Inc., New York.

Chaudhury A, Nath G, Shukla B, Panda S, Singh TB (1996) Diarrhoea associated with *Candida* spp.: incidence and seasonal variation. *Journal of Diarrhoeal Diseases Research*, **14**:110–112.

Checkley W, Epstein LD, Gilman RH et al. (2000) Effect of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *The Lancet*, **355**:442–450.

Christophers RS (1911) *Malaria in the Punjab: scientific memoirs by officers of the Medical and Sanitary Department of the Government of India*. (New Series, No. 46.) Medical and Sanitary Department of the Government of India. Government Press, Calcutta.

Coleman PG, Perry BD, Woolhouse MEJ (2001) Endemic stability—a veterinary idea applied to human public health. *The Lancet*, **357**:1284–1286.

Cook SM, Glass RI, LeBaron CW, Ho MS (1990) Global seasonality of rotavirus infections. *Bulletin of the World Health Organization*, **68**:171–177.

Craig MH, Snow RW, Le Sueur D (1999) A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today*, **15**:105–111.

Curriero FC, Patz JA, Rose JB, Lele S (2001) The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *American Journal of Public Health*, **91**:1194–1199.

Curwen M, Devis T (1988) Winter mortality, temperature and influenza: has the relationship changed in recent years? *Population Trends*, **54**:17–20.

Detinova TS (1962) *Age grouping methods in diptera of medical importance, with special reference to some vectors of malaria*. World Health Organization, Geneva.

Drasar BS, Tomkins AM, Feacham RG (1978) *Seasonal aspects of diarrhoeal disease. Seasonal dimensions to rural poverty*. (Report to UK Overseas Development Association.) London School of Hygiene and Tropical Medicine, London.

Drinkwater BL, Horvath SM (1979) Heat tolerance and aging. *Medicine and Science in Sports and Exercise* **11**:49–55.

Dye C (1992) The analysis of parasite transmission by bloodsucking insects. *Annual Review of Entomology*, **37**:1–19.

Dyson T (1999) Prospects for feeding the world. *British Medical Journal*, **319**:988–990.

Eberhard ML, Nace EK, Freeman AR, Streit TG, Da Silva AJ, Lammie PJ (1999) *Cyclospora cayetanensis* infections in Haiti: a common occurrence in the absence of watery diarrhea. *American Journal of Tropical Medicine and Hygiene*, **60**:584–586.

Ebi KL, Exuzides KA, Lau E, Kelsh M, Barsnton A (2001) Association of normal weather conditions and El Niño events with viral pneumonia hospitalizations in females, California, 1983–1998. *American Journal of Public Health*, **91**:1200–1208.

EM-DAT (2002) *The OFDA/CRED international disaster database*. Université Catholique de Louvain, Brussels.

Energy Modeling Forum (1995) *Second round study design for EMF14*. (EMF Working Paper No. 14.1.) Energy Modeling Forum, Stanford, CA.

Epstein PR (1999) Climate and health. *Science*, **285**:347–348.

FAO (1987) *Fifth world food survey*. Food and Agriculture Organization of the United Nations, Rome.

Fischer G, Frohberg K, Keyzer MA, Parikh KS (1988) *Linked national models. A tool for international food policy analysis*. Kluwer, Dordrecht.

Fitzpatrick FA, Knox JC (2000) Spatial and temporal sensitivity of hydrogeomorphic response and recovery to deforestation, agriculture, and floods. *Physical Geography*, **21**:89–108.

Fleming DM, Norbury CA, Crombie DL (1991) *Annual and seasonal variation in the incidence of common diseases*. The Royal College of General Practitioners, London.

Focks DA, Daniels E, Haile DG, Keesling JE (1995) A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *American Journal of Tropical Medicine and Hygiene*, **53**:489–506.

Focks DA, Haile DG, Daniels E, Mount GA (1993a) Dynamic life table model for *Aedes aegypti* (diptera: culicidae) analysis of the literature and model development. *Journal of Medical Entomology*, **30**:1003–1017.

Focks DA, Haile DG, Daniels E, Mount GA (1993b) Dynamic life table model for *Aedes aegypti* (diptera: culicidae) simulation results and validation. *Journal of Medical Entomology*, **30**:1018–1028.

Garrett-Jones C (1964) Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. *Nature*, **204**: 1173–1175.

Goodess C, Hulme M, Osborn T (2001) *The identification and evaluation of suitable scenario development methods of the estimation of future probabilities of extreme weather events*. School of Environmental Sciences, University of Norwich, Norwich.

Guerrant RL, Kosek M, Lima AAM, Lorntz B, Guyatt HL (2002) Updating the DALYs for diarrhoeal disease. *Trends in Parasitology*, **18**:191–193.

Guest CS, Willson K, Woodward AJ et al. (1999) Climate and mortality in Australia. Retrospective study, 1979–1990, and predicted impacts in five major cities in 2030. *Climate Research*, **13**:1–15.

Hadley Centre (1999) *Climate change and its impacts: stabilisation of CO_2 in the atmosphere*. Hadley Centre, Reading.

Haddow AJ (1943) Measurements of temperature and light in artificial pools with reference to the larval habitat of *Anopheles (Myzomyia) gambiae* Giles and *A. (M.) funestus* Giles. *Bulletin of Entomological Research*, **34**:89.

Hajat S, Haines A, Atkinson RW, Bremner SA, Anderson HR, Emberlin J (2001) Association between air pollution and daily consultations with general practitioners for allergic rhinitis in London, United Kingdom. *American Journal of Epidemiology*, **153**:704–714.

Hales S, de Wet N, Maindonald J, Woodward A (2002) Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *The Lancet*, **360**:830–834.

Halstead SB (1988) Pathogenesis of dengue: challenges to molecular biology. *Science*, **239**:476–481.

Hay SI, Cox J, Rogers DJ et al. (2002) Climate change and the resurgence of malaria in the East African highlands. *Nature*, **415**:905–909.

Hay SI, Myers MF, Burke DS et al. (2000) Etiology of interepidemic periods of mosquito-borne disease. *Proceedings of the National Academy of Sciences of the United States of America*, **97**:9335–9339.

Hoozemans FMJ, Hulsburgen CH (1995) Sea-level rise: a worldwide assessment of risk and protection costs. In: *Climate change: impact on coastal habitation*. Eisma D, ed. Lewis Publishers, London.

Hoozemans FMJ, Marchand M, Pennekamp HA (1993) *A global vulnerability analysis: vulnerability assessment for population, coastal wetlands and rice production on a global scale*. 2nd edn. Delft Hydraulics, Delft.

Hutchinson MF, Nix HA, McMahon JP, Ord KD (1996) The development of a topographic and climate database for Africa. In: *Proceedings of the third international conference/workshop on integrating GIS and environmental modeling*, NCGIA, Santa Barbara, CA.

IBSNAT (1989) *International benchmark sites network for agrotechnology transfer*. Decision Support System for Agrotechnology Transfer Version 2.1 (DSSAT V2.1), Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, HI.

IPCC (1996) *Climate change 1995: impacts, adaptations and mitigation of climate change: contribution of Working Group II to the second assessment report of the Intergovernmental Panel on Climate Change*. Watson RT, Zinyowera MC, Moss RH, eds. Cambridge University Press, Cambridge.

IPCC (2001a) *Climate change 2001: impacts, adaptation and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change*. McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, White KS, eds. Cambridge University Press, Cambridge.

IPCC (2001b) *Climate change 2001: the scientific basis: contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change*. Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Xiaosu D, eds. Cambridge University Press, Cambridge.

Jepson WF, Moutia A, Courtois C (1947) The malaria problem in Mauritius: the bionomics of Mauritian anophelines. *Bulletin of the Entomological Research*, **38**:177–208.

Jetten TH, Focks DA (1997) Potential changes in the distribution of dengue transmission under climate warming. *American Journal of Tropical Medicine and Hygiene*, **57**:285–287.

Jetten TH, Martens WJM, Takken W (1996) Model simulations to estimate malaria risk under climate change. *Journal of Medical Entomology*, **33**:361–371.

Jetten TH, Takken W (1994) *Anophelism without malaria in Europe: a review of the ecology and distribution of the genus Anopheles in Europe*. Wageningen Agricultural University Press, Wageningen.

Johns TC, Gregory JM, Stott PA, Mitchell JFB (2001) Correlations between patterns of 19th and 20th century surface temperature change and HadCM2 climate model ensembles. *Geophysical Research Letters*, **28**:1007–1010.

Jovel JR (1989) *Natural disasters and their economic and social impact*. Economic Commissions, Latin America and the Caribbean, Santiago de Chile.

Kalkstein LS, Greene JS (1997) An evaluation of climate/mortality relationships in large US cities and the possible impacts of a climate change. *Environmental Health Perspectives*, **105**:84–93.

Karl TR, Knight RW (1998) Secular trends of precipitation amount, frequency, and intensity in the United States. *Bulletin of the American Meteorological Society*, **79**:231–241.

Katz RW (2002) Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research*, **20**:167–185.

Katz RW, Brown BG (1992) Extreme events in a changing climate: variability is more important than averages. *Climatic Change*, **21**:289–302.

Keatinge WR, Coleshaw SRK, Easton JC, Cotter F, Mattock MB, Chelliah R (1986a) Increased platelet and red-cell counts, blood-viscosity, and plasma-cholesterol levels during heat-stress, and mortality from coronary and cerebral thrombosis. *American Journal of Medicine*, **81**:795–800.

Keatinge WR, Coleshaw SRK, Holmes J (1989) Changes in seasonal mortalities with improvement in home heating in England and Wales from 1964 to 1984. *International Journal of Biometeorology*, **33**:71–76.

Keatinge WR, Coleshaw SRK, Millard CE, Axelsson J (1986b) Exceptional case of survival in cold water. *British Medical Journal*, **292**:171–172.

Khaw KT (1995) Temperature and cardiovascular mortality. *The Lancet* **345**:337–338.

Kilbourne EM (1989) Heat waves. In: *The public health consequences of disasters*. Gregg MB, ed. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta, GA.

Kilbourne EM (1992) Illness due to thermal extremes. In: *Public health and preventative medicine*. Last JM, Wallace RB, eds. Appleton Lang, Norwalk, CT.

Knutti R, Stocker TF, Joos F, Plattner G-K (2002) Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416**:719–723.

Konno T, Suzuki H, Katsushima N et al. (1983) Influence of temperature and relative humidity on human rotavirus infection in Japan. *Journal of Infectious Diseases*, **147**:125–128.

Kovats RS, Bouma MJ, Haines A (1999) *El Niño and health*, World Health Organization, Geneva.

Kovats RS, Campbell-Lendrum DH, McMichael AJ, Woodward A, Cox JS (2001) Early effects of climate change: do they include changes in vector-borne disease? *Philosophical Transactions of the Royal Society, Series B*, **356**:1057–1068.

Kovats RS, Campbell-Lendrum DH, Reid C, Martens P (2000a) *Climate and vector-borne disease: an assessment of the role of climate in changing disease patterns*. ICIS/LSHTM/UNEP, Maastricht.

Kovats RS, Menne B, McMichael AJ, Corvalán C, Bertollini R (2000b) *Climate change and human health: impact and adaptation*. (WHO Technical Report WHO/SDE/OEH/00.4.) World Health Organization, Geneva.

Kuhn KG, Campbell-Lendrum DH, Armstrong B, Davies CR (2003) Malaria in Britain: past, present and future. *Proceedings of the National Academy of Sciences, USA*, **100**:9997–10001.

Kundzewicz ZW, Kaczmarek Z (2000) Coping with hydrological extremes. *Water International*, **25**:66–75.

Kunst A, Looman C, Mackenbach J (1993) Outdoor air temperature and mortality in the Netherlands—a time series analysis. *American Journal of Epidemiology*, **137**:331–341.

Kunzli N, Kaiser R, Medina S et al. (2000) Public-health impact of outdoor and traffic-related air pollution: a European assessment. *The Lancet*, **356**: 795–801.

Le Sueur D (1991) The ecology, over-wintering and population dynamics of the *Anopheles gambiae* (diptera: culicidae) complex in northern Natal, South Africa. [Dissertation.] University of Natal, Pietermaritzburg.

Lindgren E, Talleklint L, Polfeldt T (2000) Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick *Ixodes ricinus*. *Environmental Health Perspectives*, **108**: 119–123.

Loevinsohn ME (1994) Climatic warming and increased malaria incidence in Rwanda. *The Lancet*, **343**:714–718.

Longstreth J (1999) Public health consequences of global climate change in the United States—some regions may suffer disproportionately. *Environmental Health Perspectives*, **107**:169–179.

MacDonald G (1957) *The epidemiology and control of malaria*. Oxford University Press, London.

Mackenbach JP, Borst V, Schols JM (1997) Heat-related mortality among nursing home patients. *The Lancet*, **349**:1297–1298.

Maharaj R (1995) Effects of temperature on members of the *Anopheles gambiae* complex (diptera: culicidae) in South Africa—implications for malaria transmission and control. [Dissertation.] University of Natal, Pietermaritzburg.

Martens P, Kovats RS, Nijhof S et al. (1999) Climate change and future populations at risk of malaria. *Global Environmental Change—Human and Policy Dimensions*, **9**:S89–S107.

Martens P, McMichael AJ (2002) *Environmental change, climate and health*. Cambridge University Press, Cambridge.

Martens W (1998a) Climate change, thermal stress and mortality changes. *Social Science Medicine*, **46**:331–334.

Martens WJ, Jetten TH, Rotmans J, Niessen LW (1995a) Climate change and vector-borne diseases: a global modelling perspective. *Global Environmental Change*, **5**:195–209.

Martens WJ, Niessen LW, Rotmans J, Jetten TH, McMichael AJ (1995b) Potential impact of global climate change on malaria risk. *Environmental Health Perspectives*, **103**:458–464.

Martens WJM (1998b) Health impacts of climate change and ozone depletion: an ecoepidemiologic modeling approach. *Environmental Health Perspectives*, **106**:241–251.

Martens WJM, Jetten TH, Focks DA (1997) Sensitivity of malaria, schistosomiasis and dengue to global warming. *Climatic Change*, **35**:145–156.

Martin PH, Lefebvre MG (1995) Malaria and climate—sensitivity of malaria potential transmission to climate. *Ambio*, **24**:200–207.

Mason SJ, Waylen PR, Mimmack GM, Rajaratnam B, Harrison JM (1999) Changes in extreme rainfall events in South Africa. *Climatic Change*, **41**:249–257.

Massad E, Forattini OP (1998) Modelling the temperature sensitivity of some physiological parameters of epidemiologic significance. *Ecosystem Health*, **4**:119–129.

McGregor GR, Walters S, Wordley J (1999) Daily hospital respiratory admissions and winter air mass types, Birmingham, UK. *International Journal of Biometeorology*, **43**:21–30.

McMichael AJ, Githeko A (2001) Human health. In: *Climate change 2001: impacts, adaptation and vulnerability.* McCathy JJ, Canziani OF, Leary NA, Dokken DJ, White KS, eds. Cambridge University Press, Cambridge.

McMichael AJ, Haines A, Slooff R, Kovats RS (1996) *Climate change and human health: an assessment by a task group on behalf of the World Health Organization, the World Meteorological Organization and the United Nations Environment Programme*. (WHO/EHG/96.7.) World Health Organization, Geneva.

McMichael AJ, Kovats RS (2000) Strategies for assessing health impacts of global environmental change. In: *Implementing ecological integrity: restoring regional and global environmental and human health*. Westra L, ed. Kluwer Academic Publishers, Dordrecht.

McMichael AJ, Kovats RS, Cawthorne A et al. (2000a) *Fast-track: climate change and human health*. Department of Environment, Transport and the Regions Global Atmosphere Division, London.

McMichael AJ, Campbell-Lendrum DH, Corvalán CF et al. eds. (2003) *Climate change and human health: risks and responses*. World Health Organization, Geneva.

Menne B, Pond K, Noji EK, Bertollini R (1999) *Floods and public health consequences, prevention and control measures*. WHO European Centre for Environment and Health, Rome.

Milly PCD, Wetherald RT, Dunne KA, Delworth TL (2002) Increasing risk of great floods in a changing climate. *Nature*, **415**:514–517.

Mouchet J (1998) Malaria epidemics on the Highlands of Madagascar and of East and South Africa. *Bulletin de la Societe de Pathologie Exotique*, **91**:64–66.

Mouchet J, Manguin S (1999) Global warming and malaria expansion. *Annales de la Societe Entomologique de France*, **35**:549–555.

Murray CJL, Lopez AD, eds. (1996) *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of the WHO, Cambridge, MA.

National Research Council (2001a) *Abrupt climate change: inevitable surprises*. National Academy Press, Washington, DC.

National Research Council (2001b) *Under the weather: climate, ecosystems, and infectious disease*. National Academy Press, Washington, DC.

New M, Hulme M, Jones P (1999) Representing twentieth-century space-time climate variability. Part I. Development of a 1961–90 mean monthly terrestrial climatology. *Journal of Climate*, **12**:829–856.

Nicholls RJ, Hoozemans FMJ, Marchand M (1999) Increasing flood risk and wetland losses due to global sea—level rise: regional and global analyses. *Global Environmental Change-Human and Policy Dimensions*, **9**:S69–S87.

Noji EK (1997) *The public health consequences of disasters*. Oxford University Press. New York.

Otter-Nacke S, Godwin DC, Ritchie JT (1986) *Testing and validating the CERES-wheat model in diverse environments*. (AGGRISTARS YM-15-00407). Johnson Space Center, Houston, TX.

Palmer TN, Ralsanen J (2002) Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature*, **415**:512–514.

Parry M, Rosenzweig C, Iglesias A, Fischer G, Livermore M. (1999) Climate change and world food security: a new assessment. *Global Environmental Change-Human and Policy Dimensions*, **9**:S51–S67.

Patz JA, Martens WJM, Focks DA, Jetten TH (1998) Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environmental Health Perspectives*, **106**:147–153.

Patz JA, McGeehin MA, Bernard SM et al. (2000) The potential health impacts of climate variability and change for the United States: executive summary of the report of the health sector of the US National Assessment. *Environmental Health Perspectives*, **108**:367–376.

Phifer JF (1990) Psychological distress and somatic symptoms after natural disaster: differential vulnerability among older adults. *Psychology and Aging*, **5**:412–420.

Pielke RA (1999) Nine fallacies of floods. *Climatic Change*, **42**:413–438.

Purohit SG, Kelkar SD, Simha V (1998) Time series analysis of patients with rotavirus diarrhoea in Pune, India. *Journal of Diarrhoeal Diseases Research*, **16**:74–83.

Randolph S (2001) Tick-borne encephalitis in Europe. *The Lancet*, **358**: 1731–1732.

Randolph SE, Green RM, Peacey MF, Rogers DJ (2000) Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data. *Parasitology*, **121**:15–23.

Randolph SE, Rogers DJ (2000) Fragile transmission cycles of tick-borne encephalitis virus may be disrupted by predicted climate change. *Proceedings of the Royal Society of London Series B*, **267**:1741–1744.

Reiter P (2001) Climate change and mosquito-borne disease. *Environmental Health Perspectives*, **109**:141–161.

Reiter P et al. (2003) Texas lifestyle limits transmission of dengue virus. *Emerging Infectious Diseases*, **9**:86–89.

Rivers JPW (1982) Women and children last: an essay on sex discrimination in disasters. *Disasters*, **6**:256–267.

Robins-Browne RM (1984) Seasonal and racial incidence of infantile gastroenteritis in South Africa. *American Journal of Epidemiology*, **119**:350–355.

Rogers DJ, Packer MJ (1993) Vector-borne diseases, models and global change. *The Lancet*, **342**:1282–1284.

Rogers DJ, Randolph SE (2000) The global spread of malaria in a future, warmer world. *Science*, **289**:1763–1766.

Rogers DJ, Randolph SE, Snow RW, Hay SI (2002) Satellite imagery in the study and forecast of malaria. *Nature*, **415**:710–715.

Rosenzweig C, Iglesias A, Yang XB, Epstein PR, Chivian E (2001) Climate change and extreme weather events: implications for food production, plant diseases, and pests. *Global Change and Human Health*, **2**:90–104.

Rosenzweig C, Parry ML (1994) Potential impact of climate-change on world food-supply. *Nature*, **367**:133–138.

Rosenzweig C, Parry ML, Fischer G, Frohberg K (1993) Climate change and world food supply. Environmental Change Unit, University of Oxford, Oxford.

Rothwell PM, Wroe SJ, Slattery J, Warlow CP (1996) Is stroke incidence related to season or temperature? *The Lancet*, **347**:934–936.

Schlesinger ME, Williams LJ (1997) COSMIC—COuntry Specific Model for Intertemporal Climate, computer software. Electric Power Research Institute, Palo Alto, CA.

Schneider SH (2002) Can we estimate the likelihood of climatic changes at 2100? *Climatic Change*, **52**:441–451.

Schwartz J, Samet J, Patz J (2001) The effects of temperature and humidity on hospital admissions for heart disease. *Epidemiology*, **12**:61.

Schwartz J, Spix C, Touloumi G et al. (1996) Methodological issues in studies of air pollution and daily counts of death or hospital admissions. *Journal of Epidemiology and Community Health*, **50**:S3–11.

Shindell DT, Rind D, Lonergan P (1998) Increased polar stratospheric ozone losses and delayed eventual recovery owing to increasing greenhouse-gas concentrations. *Nature*, **392**:589–592.

Singh RB, Hales S, de Wet N, Raj R, Hearnden M, Weinstein P (2001) The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environmental Health Perspectives*, **109**:155–159.

Snow RW, Marsh K (1995) Will reducing *Plasmodium falciparum* transmission alter malaria mortality among African children? *Parasitology Today*, **11**: 188–190.

Stott PA, Kettleborough JA (2002) Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, **416**:723–726.

Sutherst RW (1998) Implications of global change and climate variability for vectorborne diseases: generic approaches to impact assessments. *International Journal for Parasitology*, **28**:935–945.

Sutherst RW, Yonow T, Chakraborty S, O'Donnell C, White N (1996) A generic approach to defining impacts of climate change on pests, weeds and diseases in Australia. In: *Greenhouse: coping with climatic change*. Bouma WJ, Pearman GJ, Manning MR, eds. CSIRO, Melbourne.

Tanser FC, Sharp B, Le Sueur D (2003) Potential effect of climate change on malaria transmission in Africa. *The Lancet*, **362**:1792–1798.

Tett SFB, Johns TC, Mitchell JFB (1997) Global and regional variability in a coupled AOGCM. *Climate Dynamics*, **13**:303–323.

Tol RSJ, Dowlatabadi H (2001) Vector-borne diseases, development and climate change. *Integrated Assessment*, **2**:173–181.

UNDMTP (1990) DMTP training modules, United Nations Disaster Management Training Programme. Available at http://www.undmtp.org/.

Vassallo M, Navarro K, Allen S (1995) Factors associated with high risk of marginal hyperthermia in elderly patients living in an institution. *Postgraduate Medical Journal*, **71**:213–216.

Walther GR, Post E, Convey P et al. (2002) Ecological responses to recent climate change. *Nature*, **416**:389–395.

Wang XD, Smith KR (1999) Secondary benefits of greenhouse gas control: health impacts in China. *Environmental Science and Technology*, **33**:3056–3061.

Waterlow J, Armstrong DG, Fowden L, Riley R (1998) *Feeding a world population of more than eight billion people*. Oxford University Press. Oxford.

Wayne P, Foster S, Connolly J et al. (2002) Production of allergenic pollen by ragweed (Ambrosia artemisiifolia L.) is increased in CO_2-enriched atmospheres. *Annals of Allergy, Asthma and Immunology*, **88**:279–282.

Wheeler JG, Sethi D, Cowden JM et al. (1999) Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *British Medical Journal,* **318**: 1046–1050.

Whitman S, Good G, Donoghue ER, Benbow N, Shou W, Mou S (1997) Mortality in Chicago attributed to the July 1995 heat wave. *American Journal of Public Health*, **87**:1515–1518.

WHO (1992) *Psychological consequences of disasters*. World Health Organization, Geneva.

WHO (2000a) *WHO report on global surveillance of epidemic-prone infectious diseases*. (WHO/CDS/CSR/ISR/2000.1.) Anker M, Schaaf D, eds. World Health Organization, Geneva.

WHO (2000b) *The world health report 2000*. World Health Organization, Geneva.

WHO (2002) *Global database on child growth and malnutrition*. Available at http://www.who.int/nutgrowthdb/.

WHO/EHA (1998) *Emergency health training programme for Africa*. Panafrican Emergency Training Centre, Addis Ababa.

WHO/EHA (2002) *Technical hazard sheet—drought*. World Health Organization, Department of Emergency and Humanitarian Action, Geneva.

WHO/PTC (1995) *Drought: preparedness and response in the health sector*. World Health Organization/Panafrican Emergency Training Centre, Addis Ababa.

Woodhouse PR, Khaw KT, Plummer M (1993) Seasonal variation of blood pressure and its relationship to ambient temperature in an elderly population. *Journal of Hypertension*, **11**:1267–1274.

Woodhouse PR, Khaw KT, Plummer M, Foley A, Meade TW (1994) Seasonal variations of plasma fibrinogen and factor VII activity in the elderly: winter infections and death from cardiovascular disease. *The Lancet*, **343**:435–439.

Woodward A, Hales S, Weinstein P (1998) Climate change and human health in the Asia Pacific region: who will be most vulnerable? *Climate Research*, **11**:31–38.

World Bank (1994) *World population projections 1994–1995*. Johns Hopkins University Press, Baltimore, MD.

World Climate (2002) Average temperature at Laucala Bay, Fiji. Available at http://www.worldclimate.com/cgi-bin/data.pl?ref=S18E178+1102+91690W.

Yohe G, Tol RSJ (2002) Indicators for social and economic coping capacity—moving toward a working definition of adaptive capacity. *Global Environmental Change*, **12**:25–40.

Ziska LH, Caulfield F (2000) The potential influence of rising atmospheric carbon dioxide (CO_2) on public health: pollen production of common ragweed as a test case. *World Resource Review*, **12**:449–457.

Appendix A: Uncertainty around climate predictions

Results presented here were generated using COSMIC (Country Specific Model for Intertemporal Climate) (Schlesinger and Williams 1997). This contains simplified versions of the 14 different climate models used by the IPCC, and gives output at the country level. Each model was run specifying three different plausible sensitivities (1.5, 2.5 and 4.5 °C) to the effects of a doubling of atmospheric CO_2, generating 42 alternative future climate scenarios for each GHG emissions scenario. Estimates for changes in annual average temperature and precipitation were generated for each subregion by taking population weighted averages of the country level estimates. Means, ranges and 95% confidence intervals assume independence, and equal probability for each future model and climate sensitivity. All values in figures and tables are for changes relative to 1990.

Figure A.1 Mean and 95% CI of COSMIC model outputs for temperature change in the 2020s (relative to 1990) for each subregion under each GHG emissions scenario

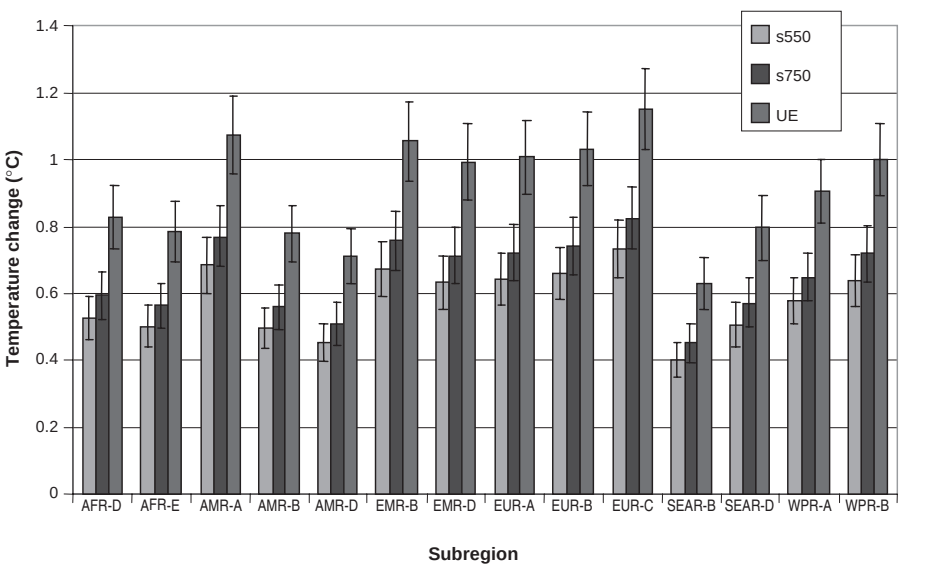


Table A.1 Mean and range of COSMIC model outputs, for temperature change in the 2020s, relative to 1990[a]

(a) S550

Subregion	*Mean dT (°C) 95% CI*	*Min–max range*	*COSMIC version HadCM2 (2.5 °C sensitivity)*
AFR-D	0.53 (0.46–0.59)	0.24–0.97	0.53
AFR-E	0.50 (0.44–0.56)	0.23–0.96	0.50
AMR-A	0.68 (0.60–0.77)	0.32–1.21	0.70
AMR-B	0.50 (0.44–0.56)	0.25–0.87	0.47
AMR-D	0.45 (0.39–0.51)	0.21–0.86	0.42
EMR-B	0.67 (0.59–0.76)	0.28–1.25	0.67
EMR-D	0.63 (0.55–0.71)	0.28–1.19	0.62
EUR-A	0.64 (0.56–0.72)	0.31–1.12	0.64
EUR-B	0.66 (0.58–0.74)	0.28–1.18	0.73
EUR-C	0.73 (0.65–0.82)	0.37–1.23	0.72
SEAR-B	0.40 (0.35–0.45)	0.17–0.77	0.35
SEAR-D	0.51 (0.44–0.57)	0.22–0.98	0.50
WPR-A	0.58 (0.51–0.64)	0.27–0.97	0.54
WPR-B	0.64 (0.56–0.71)	0.31–1.25	0.61

(b) S750

Subregion	*Mean dT (°C) 95% CI*	*Min–max range*	*COSMIC version HadCM2 (2.5 °C sensitivity)*
AFR-D	0.59 (0.52–0.66)	0.28–1.08	0.61
AFR-E	0.56 (0.49–0.63)	0.27–1.06	0.57
AMR-A	0.77 (0.68–0.86)	0.37–1.35	0.79
AMR-B	0.56 (0.49–0.62)	0.29–0.97	0.53
AMR-D	0.51 (0.45–0.57)	0.24–0.95	0.48
EMR-B	0.76 (0.67–0.85)	0.33–1.39	0.76
EMR-D	0.71 (0.63–0.80)	0.33–1.32	0.71
EUR-A	0.72 (0.64–0.81)	0.36–1.25	0.72
EUR-B	0.74 (0.65–0.83)	0.32–1.31	0.83
EUR-C	0.82 (0.73–0.92)	0.43–1.37	0.81
SEAR-B	0.45 (0.39–0.51)	0.20–0.86	0.40
SEAR-D	0.57 (0.50–0.64)	0.25–1.09	0.57
WPR-A	0.65 (0.58–0.72)	0.31–1.08	0.61
WPR-B	0.72 (0.63–0.80)	0.36–1.39	0.69

Table A.1 Mean and range of COSMIC model outputs, for temperature change in the 2020s, relative to 1990[a] (*continued*)

(c) Unmitigated emissions

Subregion	*Mean dT (°C) 95% CI*	*Min–max range*	*COSMIC version HadCM2 (2.5 °C sensitivity)*
AFR-D	0.83 (0.73–0.92)	0.41–1.47	0.85
AFR-E	0.78 (0.70–0.87)	0.39–1.45	0.80
AMR-A	1.07 (0.96–1.19)	0.54–1.83	1.11
AMR-B	0.78 (0.69–0.86)	0.42–1.32	0.75
AMR-D	0.71 (0.63–0.79)	0.35–1.29	0.67
EMR-B	1.06 (0.94–1.17)	0.48–1.88	1.08
EMR-D	0.99 (0.88–1.11)	0.48–1.79	1.00
EUR-A	1.01 (0.90–1.12)	0.52–1.69	1.02
EUR-B	1.03 (0.92–1.15)	0.46–1.78	1.17
EUR-C	1.15 (1.03–1.27)	0.63–1.86	1.15
SEAR-B	0.63 (0.55–0.70)	0.29–1.17	0.56
SEAR-D	0.80 (0.70–0.89)	0.36–1.49	0.80
WPR-A	0.91 (0.81–1.00)	0.45–1.46	0.86
WPR-B	1.00 (0.89–1.11)	0.53–1.89	0.97

[a] Results of the simplified version of the HadCM2 model at medium (2.5 °C) sensitivity are shown for comparison.

Figure A.2 Mean and 95% CI of COSMIC model outputs for precipitation change in the 2020s (relative to 1990), for each subregion under each GHG emissions scenario

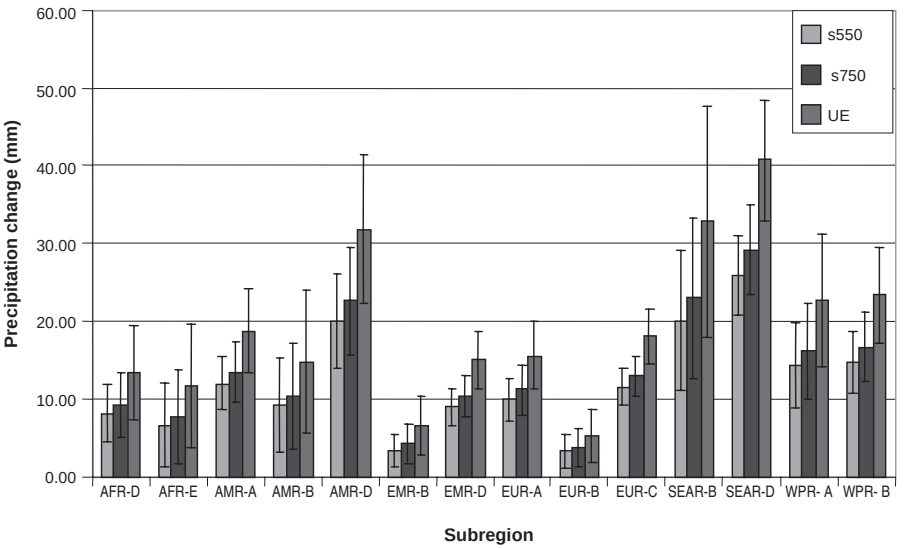


Table A.2 Mean and range of COSMIC model outputs, for change in annual precipitation in the 2020s, relative to 1990[a]

(a) S550

Subregion	*Mean dPrecipn (mm) 95% CI*	*Min–max range*	*COSMIC version HadCM2 (2.5 °C sensitivity)*
AFR-D	8.18 (4.36–11.99)	–15.8–43.88	4.92
AFR-E	6.71 (1.29–12.14)	–31.83–57.55	14.88
AMR-A	12.01 (8.53–15.48)	–10.38–54.00	13.56
AMR-B	9.17 (3.12–15.22)	–41.47–60.79	5.28
AMR-D	20.03 (13.87–26.18)	–31.17–71.40	33.48
EMR-B	3.45 (1.38–5.52)	–7.10–25.61	2.52
EMR-D	9.05 (6.71–11.38)	–1.83–32.97	5.76
EUR-A	9.94 (7.10–12.77)	–9.58–43.69	14.88
EUR-B	3.33 (1.18–5.47)	–9.55–15.18	7.56
EUR-C	11.50 (9.14–13.86)	0.72–37.68	12.48
SEAR-B	20.11 (11.05–29.16)	–35.40–135.16	7.20
SEAR-D	25.87 (20.74–31.01)	4.40–71.19	17.64
WPR-A	14.39 (8.93–19.84)	–33.56–68.61	1.32
WPR-B	14.82 (10.82–18.82)	–7.36–55.37	13.44

(b) S750

Subregion	*Mean dPrecipn (mm) 95% CI*	*Min–max range*	*COSMIC version HadCM2 (2.5 °C sensitivity)*
AFR-D	9.25 (4.97–13.52)	–17.66–48.79	5.60
AFR-E	7.67 (1.66–13.68)	–31.02–63.98	16.83
AMR-A	13.52 (9.63–17.40)	–11.79–60.07	15.43
AMR-B	10.32 (3.54–17.10)	–46.11–67.59	5.97
AMR-D	22.61 (15.73–29.49)	–34.64–79.40	37.88
EMR-B	4.24 (1.62–6.86)	–7.89–37.10	3.03
EMR-D	10.33 (7.71–12.94)	–2.05–36.66	6.47
EUR-A	11.19 (8.02–14.35)	–10.85–48.58	16.87
EUR-B	3.74 (1.34–6.15)	–10.82–16.88	8.60
EUR-C	12.95 (10.34–15.57)	0.83–41.90	14.18
SEAR-B	23.00 (12.69–33.32)	–39.37–150.27	8.23
SEAR-D	29.15 (23.45–34.85)	5.07–79.18	19.94
WPR-A	16.22 (10.11–22.33)	–37.30–76.29	1.49
WPR-B	16.70 (12.24–21.17)	–8.19–62.12	15.22

Table A.2 Mean and range of COSMIC model outputs, for change in annual precipitation in the 2020s, relative to 1990[a] (*continued*)

(c) Unmitigated emissions

Subregion	*Mean dPrecipn (mm) 95% CI*	*Min–max range*	*COSMIC version HadCM2 (2.5 °C sensitivity)*
AFR-D	13.39 (7.44–19.35)	–23.85–66.26	7.96
AFR-E	11.67 (3.69–19.64)	–37.40–86.89	23.74
AMR-A	18.81 (13.49–24.13)	–16.60–81.56	21.78
AMR-B	14.76 (5.64–23.88)	–52.87–92.27	8.42
AMR-D	31.84 (22.38–41.31)	–47.05–107.81	53.44
EMR-B	6.54 (2.75–10.34)	–10.49–54.18	4.69
EMR-D	15.10 (11.37–18.82)	–2.55–49.78	9.14
EUR-A	15.58 (11.24–19.92)	–15.30–65.98	23.79
EUR-B	5.22 (1.90–8.53)	–15.27–22.92	12.13
EUR-C	18.04 (14.50–21.58)	1.22–56.88	19.99
SEAR-B	32.75 (17.95–47.54)	–53.44–220.93	11.62
SEAR-D	40.72 (33.00–48.44)	7.38–107.89	28.12
WPR-A	22.67 (14.23–31.11)	–50.64–104.49	2.09
WPR-B	23.33 (17.18–29.47)	–11.13–86.26	21.45

[a] Results of the simplified version of the HadCM2 model at medium (2.5 °C) sensitivity are shown for comparison.

Appendix B: Literature review

For all impacts, we attempted to find all models that directly related climate change to the selected health effects, either at the global or large-regional level. We describe all such models, and outline reasons for using or not using them in the assessment. Note that the comments given relate to their relative suitability for generating the estimates required for this assessment, rather than a general judgement on their merits. For the effects of thermal extremes, we also include a summary of the study populations that were used to derive the heat and cold estimates used in this assessment.

Where existing global or large scale regional models are not considered appropriate (thermal extremes), or do not exist at all (diarrhoea and floods), the literature search describes the procedure for identifying relevant studies from which to derive new quantitative climate–health relationships and generate models.

THERMAL EXTREMES

Information sources:

- all relevant papers cited in IPCC (2001a)
- Medline search for all references (1966–2002) containing the terms "temperature"; "mortality"; "cardiovascular disease" (or CVD); "respiratory"; "weather"
- references cited in these papers.

Table B.1 Results of literature search for effects of climate change on deaths due to thermal extremes[a]

Reference	*Method*	*Inputs*	*Region*	*Key findings*	*Suitability for generating estimates for this assessment*
Langford and Bentham (1995)	Regression model: monthly mortality and temperature series (flu included)	Seasonal average dT from United Kingdom scenarios (Warrick and Barrow 1991)	England and Wales	Winter deaths avoided: 2010: all cause 3301, IHD 1308, CVD 429 2030: all cause 6353, IHD 2550, CVD 836 2050: all cause 8922, IHD 3631, CVD 1187	No population growth, ageing. No seasonal adjustment. Local rather than global
Martens (1998a)	MIASMA v1.0—empirical-statistical model, meta-analysis. A single temperature-mortality relationship was applied to all cities	ECHAM1-A, UKTR, GFDL89 (IPCC scenarios)	Cause+age specific. Global—20 cities	Changes in mortality rates for CVD (<65), CVD (>65), respiratory, and total mortality Net reductions in mortality in all cities except respiratory mortality in a few cities	Lack of control for seasonal variations, perhaps over-estimating heat effects
McMichael et al. (2000b)	MIASMA v.1.0—as for Martens (1998a)	HadCM2, ensemble mean+HadCM3	Global—20 cities	Net reductions in mortality in all cities except Athens, due to decreases in winter mortality	As for Martens (1998a)

continued

Table B.1 Results of literature search for effects of climate change on deaths due to thermal extremes[a] (*continued*)

Reference	*Method*	*Inputs*	*Region*	*Key findings*	*Suitability for generating estimates for this assessment*
Kalkstein and Smoyer (1993)	Model derived from observed relationship between synoptic air masses and mortality	GCM scenarios, no downscaling	Global—27 cities	Significant increases in heat-related mortality under various climate change scenarios, with or without acclimatization, projected for all cities	Methods used rely on synoptic classification of local air masses; therefore cannot be directly related to changes in temperature described by climate scenarios
Kalkstein and Greene (1997)	Model derived from observed relationship between synoptic air masses and mortality	GCM scenarios, no downscaling	44 cities in the USA	Increases in heat-related mortality are much larger overall than decreases in cold-related mortality	Methods used rely on synoptic classification of local air masses; therefore cannot be directly related to changes in temperature described by climate scenarios
Duncan et al. (1997)	Model derived from observed relationship between synoptic air masses and mortality	Climate scenarios for Canada	10 cities in Canada	240–1140 additional heat-related deaths/year in Montreal by 2050; 230–1220 additional deaths in Toronto, assuming no acclimatization. No climate/mortality relationship in some cities	Country specific
Guest et al. (1999)	Regression model: daily mortality and synoptic indices	CSIRO Mark 2 model, CO_2 doubling; high and low scenarios estimated	5 cities in Australia	Net decrease in heat-related mortality, particularly in the age group ≥65 years; range of 47–62 fewer deaths/year in all 5 cities. Significant increase in summer deaths in Sydney (76–239)	Country specific
Dessai (2003)	Empirical statistical model, observed-expected	2 regional climate models—PROMES and HadRM2	Lisbon, Portugal	Heat-related death rates increase by 57–113% by 2020s, by 97–255% by 2050s. Acclimatization assumptions, reduce estimates	Country specific

Key: CVD, cardiovascular disease; IHD, ischaemic hearth disease; GCM, global climate model.

[a] A total of 76 other publications were not used, either because they were analyses of specific episodes (35 studies), or because they did not match the criteria outlined above (41 studies).

Estimates of temperature–mortality relationship

Table B.2 below lists all studies (n=4) that meet the following criteria for inclusion in the meta-analysis. The study:

- uses daily time-series methods to analyse the relationship between daily mean temperature and mortality;
- Reports a coefficient from log linear regression that estimates the percentage changes in mortality per degree centigrade change in temperature, above a reported threshold temperature;
- has controls for the following confounders: season, air pollution and influenza;
- is published in English language only; and
- reports confidence intervals around the coefficient and is within the range of other reported estimates.

The studies are classified according to climate zone. The Netherlands population is used to approximate a population in the "cold" zone as there are no other appropriate time-series studies.

Note that some studies are included that are not yet published. However,

- they are the only studies available that provide estimates for developing country populations; and
- the results will be submitted to peer reviewed journals and the methods are at least as rigorous as those applied in previous published studies.

All estimates have been rounded to one decimal place. Where more than one estimate was available, the average estimate was calculated.

Table B.2 Results of literature search for temperature–mortality relationship

Population/ climate zone	*Age*	*Cause of death*	*% change[a] per 1 °C decrease*	*Cut-point (°C)*	*Lag*	*% change[a] per 1 °C increase*	*Cut-point (°C)*	*Lag*	*Reference*
			COLD			HEAT			
Netherlands	All	All cause	0.41	16.5	7–15	1.23	16.5	1–2	Kunst et al. (1993)
Netherlands	All	CVD	0.46	16.5	7–15	1.13	16.5	1–2	Kunst et al. (1993)
Netherlands	All	Respiratory	1.43	16.5	7–15	3.11	16.5	1–2	Kunst et al. (1993)
COLD		CVD	0.5			1.1			
Bulgaria									
Sofia	All	All	2.69 (0.88–4.54)	–2	0–13	1.93 (1.41–2.45)	17	0–2	ISOTHURM (forthcoming)
Chile									
Santiago	All	All	5.21 (3.55–6.89)	11	0–13	0.92 (0.44–1.31)	16	0–2	ISOTHURM (forthcoming)
Slovenia									
Ljubljana	All	All	0.77 (–0.16–1.70)	7	0–13	2.25 (1.09–3.42)	17	0–2	ISOTHURM (forthcoming)
South Africa									
Cape Town	All	All	3.32 (2.89–3.75)	19	0–13	1.02 (–0.32–2.38)	21	0–2	ISOTHURM (forthcoming)
Spain									
Madrid	All	All	1.7	20	7	0.97	20	0–1	Alberdi (1998)
Valencia	All	All	3.2 (1.8–4.6)	15	7–14	3.6 (1.2–6.0)	24	1–2	Ballester et al. (1997)
Valencia	>70	All	3.7 (2.1–5.4)	15	7–14	5.0 (2.1–8.0)	24	1–2	Ballester et al. (1997)
Valencia	All	CVD	4.3 (2.1–6.4)	15	7–14	2.3 (–1.5–4.5)	24	1–2	Ballester et al. (1997)
Valencia	All	CVD	1.5 (–0.3–3.3)	15	3–6	2.9 (–0.4–7.4)	24	3–6	Ballester et al. (1997)
Valencia	All	Respiratory	1.7 (–0.4–6.0)	15	7–14	5.7 (–2.9–8.2)	24	1–2	Ballester et al. (1997)

TEMPERATE		CVD	2.9			2.6			
Brazil									
São Paulo	All	All cause	3.92 (3.43–4.40)	19	0–13	2.28 (2.11–3.66)	23	0–2	ISOTHURM (forthcoming)
São Paulo	≥65	All cause	5.5	20	0–21	2.5 (2.1–2.8)	20	0–1	Gouveia et al. (2003)
El Salvador									
San Salvador	All	All cause	−13.5 (−32.35–10.9)	23	0–13	1.59 (0.86–2.32)	23	0–2	ISOTHURM (forthcoming)
Mexico									
Mexico City	All	All cause	8.60 (7.86–9.34)	15	0–13	0.6 (0.21–1.00)	18	0–2	ISOTHURM (forthcoming)
Mexico									
Monterrey	All	All cause	5.54 (4.52–6.58)	17	0–13	19.85 (14.69–25.25)	31	0–2	ISOTHURM (forthcoming)
Thailand									
Bangkok	All	All cause	4.13 (1.71–6.61)	28	0–13	7.66 (5.87–9.47)	30	0–2	ISOTHURM (forthcoming)
WARM HUMID	All	All cause	5.5			5.7			
India									
Delhi	All	All cause	1.36 (0.56–2.16)	19	0–13	3.03 (2.48–3.58)	28	0–2	ISOTHURM (forthcoming)
HOT DRY	All	All cause	1.4			3.0			

CVD Cardiovascular disease.

[a] % change = (RR − 1) × 100.

DIARRHOEA

Information sources:

- all relevant papers cited in IPCC (2001a)
- Medline search for all references (1966–2002) containing the terms "Diarrhoea (or diarrhea) and climate (or climatic) change"; "diarrhoea (or diarrhea) and climate"; "diarrhoea (or diarrhea) and weather"
- references cited in these papers.

Table B.3 Results of literature search for effects of climate change on diarrhoea[a]

Global models relating climate/climate change to diarrhoea	*Local studies quantitatively relating all-cause diarrhoea to temperature*	*Local studies showing seasonal patterns in all- or multiple-cause diarrhoea (non-quantitative)*	*Local studies showing pathogen-specific associations with climate*
None	Checkley et al. (2000); Singh et al. (2001)	Anjaneyulu et al. (1975); Becker (1981); Brewster and Greenwood (1993); Dean and Jones (1972); Hoge et al. (1996); Jin et al. (1996); Ling and Cheng (1993); Merlin et al. (1986); Parashar et al. (1999); Pinfold et al. (1991); Robins-Browne (1984); Rousham and Mascietaylor (1995); Saidi et al. (1997); Sawchuk (1993); Shaikh et al. (1990); Sutra et al. (1990); Tsukamoto et al. (1978); Van den Broeck et al. (1993); Williams et al. (1986); Yang et al. (1990)	Showing association: Adegbola et al. (1994); Adkins et al. (1987); Aggarwal et al. (1983); Ansari et al. (1991); Armah et al. (1994); Beards and Graham (1995); Bockemuhl (1976); Callejas et al. (1999); Chakravarti et al. (1992); Chan et al. (1998); Chaudhury et al. (1996); Cook et al. (1990); Cunliffe et al. (1998); da Rosa e Silva et al. (2001); Douglas and Kurien (1997); Eberhard et al. (1999); Fujita (1990); Henry and Bartholomew (1990); Hirschl et al. (1987); Ijaz et al. (1985); Konno et al. (1983); Laursen et al. (1994); LeBaron et al. (1990); Malakooti et al. (1997); Muhuri (1996); Musa et al. (1999); Nchito et al. (1998); Parashar et al. (1998a, 1998b); Pazzaglia et al. (1993); Purohit et al. (1998); Qiao et al. (1999); Ram et al. (1990a, 1990b); Reyes et al. (1996); Rytlewska et al. (2000); Sallon et al. (1991); Sethi et al. (1984); Shahid et al. (1987); Shkarin et al. (1983); Sorvillo et al. (1998); Stewien et al. (1991); Stintzing et al. (1981); Tswana et al. (1990); Utsalo et al. (1991); Vlasov et al. (1983); Wilcox et al. (1992); Wuhib et al. (1994) Failing to find association: Bishop et al. (2001); Conteas et al. (1998); Varoli et al. (1989)

[a] A total of 597 publications were rejected on the basis of lack of information in abstract, or irrelevance to quantitative assessment (mainly reviews, clinical and laboratory studies).

MALNUTRITION

Information sources:

- all relevant papers cited in IPCC (2001a)
- Medline search for all references (1966–2002) containing the terms "Climate (or climatic) change and malnutrition (or food)"
- references cited in these papers.

Table B.4 Results of literature search for effects of climate change on malnutrition[a]

Study	*Model type*	*Model output*	*Area covered*	*Conclusions*	*Suitability for generating estimates for this assessment*
Fischer et al. (1994)	Coupled crop and food trade model	Socioeconomic impacts, divided by region	Global	Significant impacts under climate change, concentrated in specific regions	More recent models available
Rosenzweig and Parry (1994)	Coupled crop and food trade model	Prevalence of insufficient energy intake, divided by region	Global	Significant impacts under climate change, concentrated in developing countries	More recent models available
Parry et al. (1999)	Coupled crop and food trade model	Prevalence of insufficient energy intake, divided by region	Global	Significant impacts under climate change, concentrated in developing countries	Not run on stabilization scenarios
Parry et al. (2001)	Coupled crop and food trade model	Global prevalence of insufficient energy intake	Global	Significant increase in food shortage, concentrated in developing countries	Not run on stabilization scenarios
Arnell et al. (2002)	Coupled crop and food trade model	Global prevalence of insufficient energy intake	Global	Significant impacts under climate change, CO_2 stabilization would generally reduce prevalence	Regional breakdowns of these results used as a proxy measure of prevalence of malnutrition

[a] A total of 281 other publications were rejected as they are either specific to particular crops or regions (10 studies), or irrelevant to quantitative assessment (mainly reviews, impacts on other ecological systems).

COASTAL AND INLAND FLOODING

Information sources:

- all relevant papers cited in IPCC (2001a)
- ISI-Web of Science search for "Floods (or flooding) and climate (or climatic) change"; "Floods (or flooding) and health (or death or injury or mortality)"; "Extreme precipitation and health (or death or injury or mortality)"
- Medline search for "Floods (or flooding) and climate (or climatic) change"
- References cited in these papers
- Inspection of descriptions of all flood events listed in EM-DAT enhanced database 1980–1999 (EM-DAT 2002).

Table B.5 Results of literature search for effects of climate change on deaths and injuries due to floods and landslides[a]

Large area studies quantitatively relating precipitation or sea-level rise to deaths/injuries	*Global or large regional models predicting climate change effects on frequency of extreme precipitation events or inland flooding*	*Global models predicting climate change effects on frequency of coastal flooding*	*Causes of flood events which caused deaths and/or injuries reported in the EM-DAT enhanced database (1980–1999)*	*Causes of landslides which caused deaths and/or injuries reported in the EM-DAT enhanced database (1980–1999)*
None	Booij (2002), western Europe	Baarse (1995)	Associated with precipitation	63 associated with precipitation
	Fowler and Hennessy (1995), specific locations worldwide	Hoozemans and Hulsburgen (1995)	2 with high tides and storm surges	10 associated with other causes (e.g. landslides at mines, cliffs collapsing)
	Jones and Reid (2001) United Kingdom	Nicholls and Mimura (1998)	19 other causes (e.g. cyclones, ice melt, dams bursting)	115 without specific information on cause
	Milly et al. (2002) specific locations worldwide	Nicholls et al. (1999)	559 without specific information on cause	
	Palmer and Ralsanen (2002) northern Europe, South-East Asia			

[a] A total of 342 other publications were rejected as they are either specific to particular regions, or irrelevant to quantitative assessment (mainly reviews, descriptions of specific events).

MALARIA

Information sources:

- all relevant papers cited in IPCC (2001a)
- Medline search for all references (1966–2002) containing the terms "Malaria and climate (or climatic) change"
- references cited in these papers.

Table B.6 Results of literature search for effects of climate change on malaria

Study	*Model type*	*Model output*	*Area covered*	*Conclusions*	*Suitability for generating estimates for this assessment*
Matsuoka and Kai (1994)	Biological model based on ecoclimatic index (EI) described in Sutherst et al. (1996)	Maps of changes in endemicity and the distribution of malarious areas	Global, China	Change in distribution patterns under CC	Lack of detail in methods, and independent verification
Martin and Lefebvre (1995)	Biological model based on EIP	Maps of TP. Areas at risk of epidemics and endemic malaria	Global	Expansion of distribution, increasing TP under CC	Consideration only of effects on parasite, rather than vector
Martens et al. (1995b)	Biological model—MIASMA v1.0	Changes in malaria risk relative to 1990. Maps of TP and change in TP	Global	Expansion of distribution, increasing TP under CC	Prediction of climate suitability for transmission, rather than actual transmission. Uncertainty over cut-off values to define endemic/non-endemic areas. Vector distributions not considered
Martens et al. (1995a)	Biological model—MIASMA v1.0	Changes in malaria risk relative to 1990. Maps of TP and change in TP	Global	Expansion of distribution, increasing TP under CC	Prediction of climate suitability for transmission, rather than actual transmission. Uncertainty over cut-off values to define endemic/non-endemic areas. Vector distributions not considered

continued

Table B.6 Results of literature search for effects of climate change on malaria (*continued*)

Study	*Model type*	*Model output*	*Area covered*	*Conclusions*	*Suitability for generating estimates for this assessment*
Bryan et al. (1996)	CLIMEX model	Vector distributions only	Northern Australia	Change in distribution under CC	Single vector, local rather than global predictions
Jetten et al. (1996)	Biological model—MIASMA v1.0	Maps of TP and change in TP	Global	Expansion of distribution, increasing TP under CC	Prediction of climate suitability for transmission, rather than actual transmission. Uncertainty over cut-off values to define endemic/non-endemic areas. Vector distributions not considered
Martens (1997)	MIASMA v1.0	Maps of TP and change in TP, *P. vivax* and *P. falciparum*	Global	Expansion of distribution, increasing TP under CC	Prediction of climate suitability for transmission, rather than actual transmission. Uncertainty over cut-off values to define endemic/non-endemic areas. Vector distributions not considered
Lindsay and Martens (1998)	MIASMA v1.0—biological model based on vectorial capacity	Maps—epidemic potential	African Highlands	Increase in latitude under CC	Regional projections only
Rogers (1996)	Empirical statistical mapping	Maps of distribution of *Anopheles gambiae*	Southern Africa	Change in distribution under CC	Regional projections only
Martens et al. (1999)	Biological model, overlaying a population grid. MIASMA v2.0	Maps of TP, maps of changes in seasonal transmission, additional population at risk	Global	Expansion of distribution, increasing TP under CC	Prediction of climate suitability for transmission, rather than actual transmission

Arnell et al. (2002)	Biological model, overlaying a population grid. MIASMA v2.0	Maps of changes in *P. falciparum* TP, population at risk, seasons suitable for transmission relative to 1961–1990	Global	Expansion of distribution, increasing TP under CC	Prediction of climate suitability for transmission, rather than actual transmission
Rogers and Randolph (2000)	Empirical-statistical model	Changes in populations at risk: *P. falciparum*	Global	Approximately no change in distribution under climate change	Additional population at risk ±25 million. Validation of model derived from a subset of the data against the remaining observations. (Used to inform uncertainty range around the projections)
Tanser et al. (2003)	MARA biological/ statistical model (Craig et al. 1999)	Changes in populations at risk—months suitable for transmission	Africa	Expansion in population at risk under CC	Biological model based on localized field studies, used to predict population at risk and numbers of months at risk throughout Africa, but not elsewhere. Potential for developing into predictions of incidence. Validated against detailed independent data set. (Used to determine mid-range estimates of population at risk, as a relative measure of change in incidence)

Key: CC, climate change; EIP, extrinsic incubation period; TP, transmission potential (also called EP, epidemic potential).

[a] A further 41 publications were rejected due to lack of relevance for making global projections (mainly reviews, local studies).

DENGUE

Information sources:

- all relevant papers cited in IPCC (2001a)
- Medline search for all references (1966–2002) containing the terms "Dengue and climate (or climatic) change"
- Secondary references cited in these papers.

Table B.7 Results of literature search for projected effects of climate change on dengue[a]

Study	*Model type*	*Model output*	*Area covered*	*Conclusions*	*Suitability for generating estimates for this assessment*
Martens et al. (1997)	Biological model summarizing climate effects on vector and parasite population dynamics MIASMA v1.0, coupled with DENSIM model of Focks et al. (1995)	Maps of change in epidemic potential under climate change	Global	Large increase in transmission potential with small temperature increase	Relationship between TP and disease incidence not characterized. Uncertainty over definitions for threshold TP to define limits of distribution

Patz et al. (1998)	Biological model summarizing climate effects on vector- and parasite population dynamics	Maps of change in epidemic potential under climate change	City specific (Athens, Bangkok, Mexico City, Philadelphia, Puerto Rico, San Juan)	Large increase in transmission potential with small temperature increase	Model is most accurate for non-endemic areas bordering endemic areas and may underestimate changes in transmission in temperate zones. Relationship between TP and disease incidence not characterized
Jetten and Focks (1997)	Biological model summarizing climate effects on vector- and parasite population dynamics	Maps of changes in Critical Density Threshold	Global	Large increase in areas suitable for transmission and length of transmission season under climate change	Relationship between TP and disease incidence not characterized. Uncertainty over setting of cut-offs for transmission
Hales et al. (2002)	Empirical statistical model	Maps of population at risk for dengue transmission	Global	Absolute humidity accurately explains current distribution of dengue. Large increases in population at risk predicted under climate change	Validation of model derived from a subset of the data against the remaining observations. (Used to determine population at risk, as a relative measure of change in incidence)

[a] A further 12 publications were rejected due to lack of relevance for making global projections (mainly reviews, local studies).

Appendix references

Adegbola RA, Demba E, de Veer G, Todd J (1994) Cryptosporidium infection in Gambian children less than 5 years of age. *Journal of Tropical Medicine and Hygiene*, **97**:103–107.

Adkins HJ, Escamilla J, Santiago LT, Ranoa C, Echeverria P, Cross JH (1987) Two-year survey of etiologic agents of diarrheal disease at San Lazaro Hospital, Manila, Republic of the Philippines. *Journal of Clinical Microbiology*, **25**:1143–1147.

Aggarwal P, Sarkar R, Gupta JP, Ahuja S, Ray K, Rai Chowdhuri AN (1983) Current epidemiological aspects of cholera in Delhi. *Journal of Communicable Diseases*, **15**:26–32.

Alberdi JC (1998) Daily mortality in Madrid community, 1986–1992: relationship with meteorological variables. *European Journal of Epidemiology*, **14**:571–578.

Anjaneyulu G, Banerji SC, Indrayan A (1975) A study of fly density and meteorological factors in occurrence of diarrhoea in a rural area. *Indian Journal of Public Health*, **19**:115–121.

Ansari SA, Springthorpe VS, Sattar SA (1991) Survival and vehicular spread of human rotaviruses—possible relation to seasonality of outbreaks. *Reviews of Infectious Diseases*, **13**:448–461.

Armah GE, Mingle JAA, Dodoo AK et al. (1994) Seasonality of rotavirus infection in Ghana. *Annals of Tropical Paediatrics*, **14**:223–229.

Arnell N, Cannell MG, Hulme M et al. (2002) The consequences of CO_2 stabilization for the impacts of climate change. *Climatic Change*, **53**:413–446

Baarse G (1995) *The development of an operational tool for global vulnerability assessment (GVA) update of the number of people at risk due to sea level rise and increased flooding probabilities*. Ministry of Transport, Public Works and Water Management, The Hague.

Ballester DF, Corella D, Perez-Hoyos S, Saez M, Hervas A (1997) Mortality as a function of temperature. A study in Valencia, Spain, 1991–1993. *International Journal of Epidemiology*, **26**:551–561.

Beards G, Graham C (1995) Temporal distribution of rotavirus G-serotypes in the West Midlands region of the United Kingdom, 1983–1994. *Journal of Diarrhoeal Diseases Research*, **13**:235–237.

Becker S (1981) Seasonality of deaths in Matlab, Bangladesh. *International Journal of Epidemiology*, **10**:271–280.

Bishop RF, Masendycz PJ, Bugg HC, Carlin JB, Barnes GL (2001) Epidemiological patterns of rotaviruses causing severe gastroenteritis in young children throughout Australia from 1993 to 1996. *Journal of Clinical Microbiology*, **39**:1085–1091.

Bockemuhl J (1976) Salmonellosis and shigellosis in Togo (West Africa), 1971–1973. I. Carrier rates in the rural population. *Tropenmedizin und Parasitologie*, **27**:112–120.

Booij MJ, (2002) Extreme daily precipitation in Western Europe with climate change at appropriate spatial scales. *International Journal of Climatology*, **22**:69–85.

Brewster DR, Greenwood BM (1993) Seasonal-variation of pediatric diseases in the Gambia, West-Africa. *Annals of Tropical Paediatrics*, **13**:133–146.

Bryan JH, Foley DH, Sutherst RW (1996) Malaria transmission and climate change in Australia. *Medical Journal of Australia*, **164**:345–347.

Callejas D, Estevez J, Porto-Espinoza L et al. (1999) Effect of climatic factors on the epidemiology of rotavirus infection in children under 5 years of age in the city of Maracaibo, Venezuela [in Spanish]. *Investigacion Clinica*, **40**:81–94.

Chakravarti A, Broor S, Natarajan R, Setty VS, Mittal SK (1992) Epidemiological and clinical characteristics of acute diarrhoea in children due to human rotavirus. *Journal of Tropical Pediatrics*, **38**:192–193.

Chan PKS, Tam JS, Nelson EAS et al. (1998) Rotavirus infection in Hong Kong: epidemiology and estimates of disease burden. *Epidemiology and Infection*, **120**:321–325.

Chaudhury A, Nath G, Shukla B, Panda S, Singh TB (1996) Diarrhoea associated with *Candida* spp.: incidence and seasonal variation. *Journal of Diarrhoeal Diseases Research*, **14**:110–112.

Checkley W, Epstein LD, Gilman RH et al. (2000) Effect of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *The Lancet*, **355**:442–450.

Conteas CN, Berlin OG, Lariviere MJ et al. (1998) Examination of the prevalence and seasonal variation of intestinal microsporidiosis in the stools of persons with chronic diarrhea and human immunodeficiency virus infection. *American Journal of Tropical Medicine and Hygiene*, **58**:559–561.

Cook SM, Glass RI, LeBaron CW, Ho MS (1990) Global seasonality of rotavirus infections. *Bulletin of the World Health Organization*, **68**:171–177.

Cunliffe NA, Kilgore PE, Bresee JS et al. (1998) Epidemiology of rotavirus diarrhoea in Africa: a review to assess the need for rotavirus immunization. *Bulletin of the World Health Organization*, **76**:525–537.

da Rosa e Silva ML, Naveca FG, Pires de Carvalho I (2001) Epidemiological aspects of rotavirus infections in Minas Gerais, Brazil. *Brazilian Journal of Infectious Diseases*, **5**:215–222.

Dean AG, Jones TC (1972) Seasonal gastroenteritis and malabsorption at an American military base in the Philippines. I. Clinical and epidemiologic investigations of the acute illness. *American Journal of Epidemiology*, **95**:111–127.

Dessai SR (2003) Heat stress and mortality in Lisbon. Part II. An assessment of the potential impacts of climate change. *International Journal of Biometeorology*, **48**:37–44.

Douglas AS, Kurien A (1997) Seasonality and other epidemiological features of haemolytic uraemic syndrome and E-coli O157 isolates in Scotland. *Scottish Medical Journal*, **42**:166–171.

Duncan K, Guidotti TL, Cheng K et al. (1997) The Canadian Country study: impacts and adaptation, Health Sector. In: *The Canadian Country study: climate impacts and adaptation, National Sectoral Volume VII*. Environment Canada. Toronto.

Eberhard ML, Nace EK, Freeman AR, Streit TG, Da Silva AJ, Lammie PJ (1999) *Cyclospora cayetanensis* infections in Haiti: a common occurrence in the absence of watery diarrhea. *American Journal of Tropical Medicine and Hygiene*, **60**:584–586.

Fischer G, Frohberg K, Parry ML, Rosenzweig C (1994) Climate-change and world food-supply, demand and trade—who benefits, who loses. *Global Environmental Change-Human and Policy Dimensions*, **4**:7–23.

Fowler AM, Hennessy KJ (1995) Potential impacts of global warming on the frequency and magnitude of heavy precipitation. *Natural Hazards*, **11**:283–303.

Fujita Y (1990) Rotavirus infection—clinical symptoms and influence of climate [in Japanese]. *Kansenshogaku Zasshi*, **64**:1255–1263.

Gouveia N, Hajat S, Armstrong B (2003) Socio-economic differentials in the temperature-mortality relationship in Sao Paulo, Brazil. *International Journal of Epidemiology*, **32**:390–397.

Guest CS, Willson K, Woodward AJ et al. (1999) Climate and mortality in Australia. Retrospective study, 1979–1990, and predicted impacts in five major cities in 2030. *Climate Research*, **13**:1–15.

Hales S, de Wet N, Maindonald J, Woodward A (2002) Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *The Lancet*, **360**:830–834.

Henry FJ, Bartholomew RK (1990) Epidemiology and transmission of rotavirus infections and diarrhoea in St. Lucia, West Indies. *West Indian Medical Journal*, **39**:205–212.

Hirschl AM, Lior H, Wolf D et al. (1987) Occurrence, serotypes and biotypes of thermophilic Campylobacters isolated in Vienna. *Zentralblatt für Bakteriologie, Mikrobiologie, und Hygiene, Serie A*, **266**:94–103.

Hoge CW, Shlim DR, Echeverria P, Rajah R, Herrmann JE, Cross JH (1996) Epidemiology of diarrhea among expatriate residents living in a highly endemic environment. *Journal of the American Medical Association*, **275**:533–538.

Hoozemans FMJ, Marchand M, Pennekamp HA (1993) *A global vulnerability analysis: vulnerability assessment for population, coastal wetlands and rice production on a global scale*. 2nd edn. Delft Hydraulics, Delft.

Ijaz MK, Sattar SA, Johnson-Lussenburg CM, Springthorpe VS, Nair RC (1985) Effect of relative humidity, atmospheric temperature, and suspending medium on the airborne survival of human rotavirus. *Canadian Journal of Microbiology*, **31**:681–685.

ISOTHURM (forthcoming) ISOTHURM Study Group: international study of temperature and heat-waves on urban mortality in low and middle income countries. *The Lancet*,

Jetten TH, Focks DA (1997) Potential changes in the distribution of dengue transmission under climate warming. *American Journal of Tropical Medicine and Hygiene*, **57**:285–287.

Jetten TH, Martens WJM, Takken W (1996) Model simulations to estimate malaria risk under climate change. *Journal of Medical Entomology*, **33**:361–371.

Jin SX, Kilgore PE, Holman RC, Clarke MJ, Gangarosa EJ, Glass RI (1996) Trends in hospitalizations for diarrhea in United States children from 1974 through 1992: estimates of the morbidity associated with rotavirus. *Pediatric Infectious Disease Journal*, **15**:397–404.

Jones PD, Reid PA (2001) Assessing future changes in extreme precipitation over Britain using regional climate model integrations. *International Journal of Climatology*, **21**:1337–1356.

Kalkstein LS, Greene JS (1997) An evaluation of climate/mortality relationships in large US cities and the possible impacts of a climate change. *Environmental Health Perspectives*, **105**:84–93.

Kalkstein LS, Smoyer KE (1993) The impact of climate change on human health: some international implications. *Experientia*, **49**:969–979.

Konno T, Suzuki H, Katsushima N et al. (1983) Influence of temperature and relative humidity on human rotavirus infection in Japan. *Journal of Infectious Diseases*, **147**:125–128.

Kunst A, Looman C, Mackenbach J (1993) Outdoor air temperature and mortality in the Netherlands—a time series analysis. *American Journal of Epidemiology*, **137**:331–341.

Langford IH, Bentham G (1995) The potential effects of climate-change on winter mortality in England and Wales. *International Journal of Biometeorology*, **38**:141–147.

Laursen E, Mygind O, Rasmussen B, Ronne T (1994) Gastroenteritis: a waterborne outbreak affecting 1600 people in a small Danish town. *Journal of Epidemiology and Community Health*, **48**:453–458.

LeBaron CW, Lew J, Glass RI, Weber JM, Ruiz-Palacios GM (1990) Annual rotavirus epidemic patterns in North America. Results of a 5-year retrospective survey of 88 centers in Canada, Mexico, and the United States. Rotavirus Study Group. *Journal of the American Medical Association*, **264**:983–988.

Lindsay SW, Martens WJM (1998) Malaria in the African highlands: past, present and future. *Bulletin of the World Health Organization*, **76**:33–45.

Ling JM, Cheng AF (1993) Infectious diarrhea in Hong-Kong. *Journal of Tropical Medicine and Hygiene*, **96**:107–112.

Malakooti MA, Alaii J, Shanks GD, Phillips-Howard PA (1997) Epidemic dysentery in western Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **91**:541–543.

Martens P (1997) *Health impacts of climate change and ozone depletion: an eco-epidemiological modelling approach*. Maastricht University, Maastricht.

Martens W (1998a) Climate change, thermal stress and mortality changes. *Social Science Medicine*, **46**:331–334.

Martens WJ, Jetten TH, Rotmans J, Niessen LW (1995a) Climate change and vector-borne diseases: a global modelling perspective. *Global Environmental Change*, **5**:195–209.

Martens WJ, Niessen LW, Rotmans J, Jetten TH, McMichael AJ (1995b) Potential impact of global climate change on malaria risk. *Environmental Health Perspectives*, **103**:458–464.

Martens WJM, Jetten TH, Focks DA (1997) Sensitivity of malaria, schistosomiasis and dengue to global warming. *Climatic Change*, **35**:145–156.

Martens P, Kovats RS, Nijhof S et al. (1999) Climate change and future populations at risk of malaria. *Global Environmental Change—Human and Policy Dimensions*, **9**:S89–S107.

Martin PH, Lefebvre MG (1995) Malaria and climate—sensitivity of malaria potential transmission to climate. *Ambio*, **24**:200–207.

Matsuoka Y, Kai K (1994) An estimation of climatic change effects on malaria. *Journal of Global Environment Engineering*, **1**:1–15.

McMichael AJ, Kovats S, Martens P et al. (2000b) *Climate change and human health: final report to the Department of Environment, Transport and the Regions*. London School of Hygiene and Tropical Medicine/ICIS, London/Maastricht.

Merlin M, Roure C, Kollo B et al. (1986) Evaluation of morbidity, mortality and therapeutic procedures in diarrheal diseases of young children in Cameroon. *Médecine Tropicale (Marseilles)*, **46**:355–357.

Milly PCD, Wetherald RT, Dunne KA, Delworth TL (2002) Increasing risk of great floods in a changing climate. *Nature*, **415**:514–517.

Muhuri PK (1996) Estimating seasonality effects on child mortality in Matlab, Bangladesh. *Demography*, **33**:98–110.

Musa HA, Shears P, Kafi S, Elsabag SK (1999) Water quality and public health in northern Sudan: a study of rural and peri-urban communities. *Journal of Applied Microbiology*, **87**:676–682.

Nchito M, Kelly P, Sianongo S et al. (1998) Cryptosporidiosis in urban Zambian children: an analysis of risk factors. *American Journal of Tropical Medicine and Hygiene*, **59**:435–437.

Nicholls RJ, Mimura N (1998) Regional issues raised by sea level rise and their policy implications. *Climate Research*, **11**:5–18.

Nicholls RJ, Hoozemans FMJ, Marchand M (1999) Increasing flood risk and wetland losses due to global sea—level rise: regional and global analyses. *Global Environmental Change-Human and Policy Dimensions*, **9**:S69–S87.

Palmer TN, Ralsanen J (2002) Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature*, **415**:512–514.

Parashar UD, Chung MA, Holman RC, Ryder RW, Hadler JL, Glass RI (1999) Use of state hospital discharge data to assess the morbidity from rotavirus diarrhea and to monitor the impact of a rotavirus immunization program: a pilot study in Connecticut. *Pediatrics*, **104**:489–494.

Parashar UD, Holman RC, Bresee JS et al. (1998a) Epidemiology of diarrheal disease among children enrolled in four west coast health maintenance organizations. *Pediatric Infectious Disease Journal*, **17**:605–611.

Parashar UD, Holman RC, Clarke MJ, Bresee JS, Glass RI (1998b) Hospitalizations associated with rotavirus diarrhea in the United States, 1993 through 1995: surveillance based on the new ICD-9-CM rotavirus-specific diagnostic code. *Journal of Infectious Diseases*, **177**:13–17.

Parry M, Rosenzweig C, Iglesias A, Fischer G, Livermore M. (1999) Climate change and world food security: a new assessment. *Global Environmental Change-Human and Policy Dimensions*, **9**:S51–S67.

Parry ML, Arnell NW, McMichael A et al. (2001) Millions at risk: defining critical climate change threats and targets. *Global Environmental Change*, **11**:181–183.

Patz JA, Martens WJM, Focks DA, Jetten TH (1998) Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environmental Health Perspectives*, **106**:147–153.

Pazzaglia G, Bourgeois AL, Araby I et al. (1993) Campylobacter-associated diarrhea in Egyptian infants—epidemiology and clinical manifestations of disease and high-frequency of concomitant infections. *Journal of Diarrhoeal Diseases Research*, **11**:6–13.

Pinfold JV, Horan NJ, Mara DD (1991) Seasonal effects on the reported incidence of acute diarrhoeal disease in northeast Thailand. *International Journal of Epidemiology*, **20**:777–786.

Purohit SG, Kelkar SD, Simha V (1998) Time series analysis of patients with rotavirus diarrhoea in Pune, India. *Journal of Diarrhoeal Diseases Research*, **16**:74–83.

Qiao HP, Nilsson M, Abreu ER et al. (1999) Viral diarrhea in children in Beijing, China. *Journal of Medical Virology*, **57**:390–396.

Ram S, Khurana S, Khurana SB, Sharma S, Vadehra DV (1990a) Seasonal fluctuations in the occurrence of enteroinvasive Escherichia coli diarrhoea. *Indian Journal of Medical Research*, **91**:258–262.

Ram S, Khurana S, Khurana SB, Sharma S, Vadehra DV, Broor S (1990b) Bioecological factors and rotavirus diarrhoea. *Indian Journal of Medical Research*, **91**:167–170.

Reyes M, Hedlund KO, Lorenzana I, Ehrnst A (1996) Respiratory infection and iatrogenic diarrhea in Honduras and El Salvador during the 1991–1992 season. *American Journal of Tropical Medicine and Hygiene*, **54**:260–264.

Robins-Browne RM (1984) Seasonal and racial incidence of infantile gastroenteritis in South Africa. *American Journal of Epidemiology*, **119**:350–355.

Rogers DJ (1996) Regional impacts of climate change; changes in disease vector distributions. In: *Climate change and Southern Africa: an exploration of some potential impacts and implications in the SADC region*. Hulme M, ed. University of East Anglia, Norwich.

Rogers DJ, Randolph SE (2000) The global spread of malaria in a future, warmer world. *Science*, **289**:1763–1766.

Rosenzweig C, Parry ML (1994) Potential impact of climate-change on world food-supply. *Nature*, **367**:133–138.

Rousham EK, Mascietaylor CGN (1995) Seasonality and child morbidity in rural Bangladesh. *American Journal of Human Biology*, **7**:369–379.

Rytlewska M, Bako W, Ratajczak B et al. (2000) Epidemiological and clinical characteristics of rotaviral diarrhoea in children from Gdansk, Gdynia and Sopot. *Medical Science Monitor* **6**:117–122.

Saidi SM, Iijima Y, Sang WK et al. (1997) Epidemiological study on infectious diarrheal diseases in children in a coastal rural area of Kenya. *Microbiology and Immunology*, **41**:773–778.

Sallon S, el Showwa R, el Masri M, Khalil, M, Blundell N, Hart CA (1991) Cryptosporidiosis in children in Gaza. *Annals of Tropical Paediatrics*, **11**:277–281.

Sawchuk LA (1993) Societal and ecological determinants of urban health: a case study of pre-reproductive mortality in 19th century Gibraltar. *Social Science and Medicine*, **36**:875–892.

Sethi SK, Al-Nakib W, Khuffash FA, Majeed HA (1984) Acute diarrhoea and rotavirus infections in young children in Kuwait. *Annals of Tropical Paediatrics*, **4**:117–121.

Shahid NS, Rahman AS, Sanyal SC (1987) Cryptosporidium as a pathogen for diarrhoea in Bangladesh. *Tropical and Geographical Medicine*, **39**:265–270.

Shaikh K, Wojtyniak B, Mostafa G, Khan MU (1990) Pattern of diarrheal deaths during 1966–1987 in a demographic surveillance area in rural Bangladesh. *Journal of Diarrhoeal Diseases Research*, **8**:147–154.

Shkarin VV, Ouchfoun A, Minaev VI, Naceur D (1983) Epidemiology of bacillary dysentery in Algeria. II. The seasonality of dysentery. *Zhurnal Mikrobiologii, Epidemiologii i Immunobiologii*, 53–58.

Singh RB, Hales S, de Wet N, Raj R, Hearnden M, Weinstein P (2001) The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environmental Health Perspectives*, **109**:155–159.

Sorvillo F, Beall G, Turner PA et al. (1998) Seasonality and factors associated with cryptosporidiosis among individuals with HIV infection. *Epidemiology and Infection*, **121**:197–204.

Stewien KE, Dacunha LCF, Alvim AD et al. (1991) Rotavirus associated diarrhea during infancy in the city of S Luis (Ma), Brazil—a 2-year longitudinal-study. *Revista do Instituto de Medicina Tropical de São Paulo*, **33**:459–464.

Stintzing G, Back E, Tufvesson B, Johnsson T, Wadstrom T, Habte D (1981) Seasonal fluctuations in the occurrence of entero-toxigenic bacteria and rotavirus in pediatric diarrhea in Addis-Ababa. *Bulletin of the World Health Organization*, **59**:67–73.

Sutra S, Srisontrisuk S, Panpurk W et al. (1990) The pattern of diarrhea in children in Khon Kaen, northeastern Thailand: I. The incidence and seasonal variation of diarrhea. *Southeast Asian Journal of Tropical Medicine and Public Health*, **21**:586–593.

Tanser FC, Sharp B, Le Sueur D (2003) Potential effect of climate change on malaria transmission in Africa. *The Lancet*, **362**:1792–1798.

Tsukamoto T, Kinoshita Y, Shimada T, Sakazaki R (1978) Two epidemics of diarrhoeal disease possibly caused by *Plesiomonas shigelloides*. *Journal of Hygiene*, **80**:275–280.

Tswana SA, Jorgensen PH, Halliwell RW, Kapaata R, Moyo SR (1990) The incidence of rotavirus infection in children from 2 selected study areas in Zimbabwe. *Central African Journal of Medicine*, **36**:241–246.

Utsalo SJ, Eko FO, Antia-Obong OE (1991) Cholera and *Vibrio parahaemolyticus* diarrhoea endemicity in Calabar, Nigeria. *West African Journal of Medicine*, **10**:175–180.

Van den Broeck J, Eeckels R, Devlieger H (1993) Child morbidity patterns in two tropical seasons and associated mortality rates. *International Journal of Epidemiology*, **22**:1104–1110.

Varoli O, Gatti M, Piscolla F (1989) A one-year study on thermophilic campylobacters isolated from fecal specimens. *Microbiologica*, **12**:263–265.

Vlasov VI, Zamotin BA, Burykh VM (1983) Method for the short-term prognosis of annual morbidity indices in Sonne-type dysentery. *Zhurnal Mikrobiologii, Epidemiologii i Immunobiologii*, 53–56.

Warrick RA, Barrow EM (1991) Climate change scenarios for the UK. *Transactions of the Institute of British Geographers*, **16**:387–399.

Wilcox MH, Cook AM, Eley A, Spencer RC (1992) *Aeromonas* spp. as a potential cause of diarrhoea in children. *Journal of Clinical Pathology*, **45**:959–963.

Williams EH, Hayes RJ, Smith PG (1986) Admissions to a rural hospital in the West Nile District of Uganda over a 27 year period. *Journal of Tropical Medicine and Hygiene*, **89**:193–211.

Wuhib T, Silva TM, Newman RD et al. (1994) Cryptosporidial and micro-sporidial infections in human immunodeficiency virus-infected patients in northeastern Brazil. *Journal of Infectious Diseases*, **170**:494–497.

Yang CR, Meng ZD, Wang X, Li YL, Zhang YX, Zhao QP (1990) Diarrhea surveillance in children aged under 5 years in a rural area of Hebei Province, China. *Journal of Diarrhoeal Diseases Research*, **8**:155–159.

Chapter 21

Selected occupational risk factors

Marisol Concha-Barrientos,
Deborah Imel Nelson, Timothy Driscoll,
N. Kyle Steenland, Laura Punnett,
Marilyn A. Fingerhut, Annette Prüss-Üstün,
James Leigh, SangWoo Tak and Carlos Corvalan

Summary

Many of the 2.9 billion workers across the globe are exposed to hazardous risks at their workplaces. This chapter examines the disease and injury burden produced by selected occupational risk factors: occupational carcinogens, airborne particulates, noise, ergonomic stressors and risk factors for injuries. Owing primarily to lack of data in developing countries, we were unable to include important occupational risks for some cancers, reproductive disorders, dermatitis, infectious diseases, ischaemic heart disease, musculoskeletal disorders (MSDs) of the upper extremities, and other conditions such as workplace stress. Mesothelioma and asbestosis due to asbestos exposure, silicosis and coal workers' pneumoconiosis are almost exclusively due to workplace exposure, but limitations in global data precluded a full analysis of these outcomes.

The economically active population (EAP) aged ≥15 years, which includes people in paid employment, the self-employed, and those who work to produce goods and services for their own household consumption, were considered the group at risk of exposure to occupational hazards. Both formal and informal sectors of employment are included in the EAP, but child labour was excluded. Exposure was quantified based on the economic sector (where people do the work) and on occupation (what people do). Our sources of data to delineate categories of exposed workers included economic databases and publications of the International Labour Organization (ILO) and the World Bank and the published scientific literature. For most risk factors the workers were grouped into high- and low-exposure categories, and the exposed population was distributed by age, sex and subregion.[1] Risk estimates for the occupational hazards were obtained from the published epidemiological literature, particularly from studies of large populations, reviews and meta-analyses when available.

The occupational risk factors in our study accounted for an estimated 37% of back pain, 16% of hearing loss, 13% of chronic obstructive

pulmonary disease (COPD), 11% of asthma, 8% of injuries, 9% of lung cancer and 2% of leukaemia. These work-related risks caused 775 000 deaths worldwide in 2000. There were five times as many deaths in males as in females (647 000 vs 128 000). The leading occupational cause of death among the six risk factors was unintentional injuries (41%) followed by COPD (40%) and cancer of the trachea, bronchus or lung (13%). Workers who developed outcomes related to the occupational risk factors lost about 22 million years of healthy life. By far the main cause of years of healthy life lost (measured in disability-adjusted life years [DALYs]), within occupational diseases, was unintentional injuries with 48% of the burden. This was followed by hearing loss due to occupational noise (19%) and COPD due to occupational agents (17%). Males experienced almost five times greater loss of healthy years (DALYs) than females. Low back pain and hearing loss have in common the fact that they do not directly produce premature mortality, but they cause substantial disability and have multiple consequences for the individual and society, particularly for workers suffering the outcomes at an early age.

The major source of uncertainty in our analysis was characterizing exposure, which was based solely on economic subsectors and/or occupations and involved a large number of extrapolations and assumptions. High-quality exposure data are lacking, especially in developing countries, and European and American exposure estimates were thus applied in many instances in developing regions. This extrapolation could have substantial impact on the accuracy of analysis for the developing regions if exposures, as usually occur, vary from place to place and over time. Diseases with long latency (e.g. cancers) are more susceptible to the assumptions and extrapolations. In addition to problems produced by the length of the latency period, the magnitude of the excess risk may vary depending on the age of the person when exposure began, the duration and strength of exposure and other concomitant exposures. The turnover of workers is another issue that affects both exposure and risk assessment. Sources of uncertainty in hazard estimates (relative risk and mortality rates) include variations determined from the literature (once again caused by the use of different exposure proxies), extrapolations to regions with different working conditions, the application to females of risk measures from male cohorts, and the application of the same relative risk values to all age groups (e.g. carcinogens). Restricting the analysis to persons aged ≥15 years excludes the quantification of child labour. The exclusion of children in the estimation was due to the wide variation in the youngest age group for which countries reported economic activity rates (EARs). In addition to inconsistent data on EARs for children, there were virtually no data available on their exposure to occupational risk factors or the relative risks of such exposures. Specific, focused research on children is needed to quantify the global burden of disease due to child labour and the resulting implications.

1. Introduction

Throughout the world, most adults—and many children—spend much of their waking hours at work. Work provides a number of economic and other benefits. At the same time, people at work face a variety of hazards owing to chemicals, biological agents, physical factors, adverse ergonomic conditions, allergens, a complex network of safety risks, and many and varied psychosocial factors. In addition to injuries, more than 100 occupational diseases have been classified according to the tenth revision of the International Classification of Diseases and Related Health Problems (ICD-10). Broadly, these include respiratory, musculoskeletal, cardiovascular, reproductive, neurotoxic, skin and psychological disorders, hearing loss and cancers.

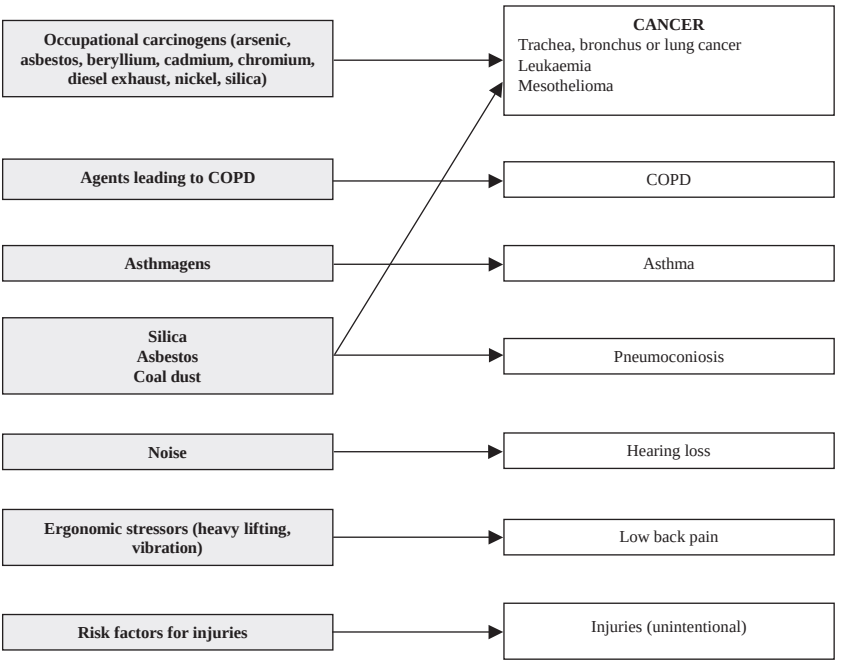
Of the wide variety of work-related exposures, only the most widespread are evaluated here. Other criteria for selection of risk factors include adequacy of exposure information and the applicability of health outcome data to all regions of the globe, and the inclusion of the relevant health outcomes in the global burden of disease (GBD) database of diseases and injuries.

Exposure to occupational hazards can adversely affect the human body. Adverse effects range from asymptomatic physiological and biochemical changes to symptoms of illness, to diagnosed diseases and, finally, to death. For some risk factors there is a very clear connection between the exposure and the disease. For example, the primary route of exposure to airborne particulates, gases and vapours is inhalation, whereby these agents gain access to the respiratory system and are either deposited (in the case of particulates) or enter the circulatory system (gases and vapours). Many risk factors cause more than one type of outcome of interest. For example, exposure to asbestos can result in malignant conditions of the lung and the pleura, malignant conditions of the peritoneum, and nonmalignant conditions of the lung (asbestosis). Some exposures, such as occupational noise, are well characterized. Others have not been well characterized or are multi-faceted, but the condition they cause is clear (such as occupational injuries).

Following a general description of methods and data sources, individual sections provide details of specific aspects of methodology and results for each of the selected occupational health risk factors that were analysed: occupational carcinogens, occupational airborne particulates, occupational noise, occupational ergonomic stressors and occupational risk factors for injuries.

In this study, the term "occupational risk factor" is defined as a chemical, physical, biological or other agent that may cause harm to an exposed person in the workplace and is potentially modifiable. Figure 21.1 shows the selected risk factors along with related health outcomes. Owing to complex etiology and lack of data, a different approach was developed for some conditions such as asthma and low back pain, using

Figure 21.1 Relationship between occupational risk factors and outcomes[a]



[a] Only selected relationships were quantified.

occupation as a proxy for exposure to the causative agents. The utility of this work as a risk-based framework has thus been limited.

1.1 Excluded exposures and outcomes

No effects specific to the hazards associated with child labour are addressed in this report owing to a lack of data. Other excluded risks or outcomes include respiratory diseases other than COPD and asthma; some infectious diseases; less widespread cancers and carcinogens (e.g. bladder cancer and cancer of the liver); MSDs such as carpal tunnel syndrome; intentional injuries in the workplace; organ and systemic diseases resulting from occupational exposure to solvents, pesticides and heavy metals such as lead or mercury; maternal and perinatal conditions resulting from occupational exposures; skin disorders, including dermatitis, dermatosis and melanoma; ischaemic heart disease and other outcomes associated with work-related stress.

Malignant mesothelioma of the pleura and peritoneum is virtually uniquely due to asbestos exposure. Occupational dusts can also result in

nonmalignant respiratory diseases other than asthma and COPD. The most important of these are silicosis, asbestosis and coal workers' pneumoconiosis, which are caused by exposure to silica, asbestos and coal dust, respectively. While evidence for a causal relationship is strong, lack of data on accumulated exposure, especially in developing countries, restricted the ability to provide a detailed assessment of attributable mortality and disease burden for these outcomes. Preliminary estimates are provided in the note under Table 21.62 in Section 7.

Because of lack of available data and difficulties in quantification, it was not possible to conduct a global quantitative analysis for the health consequences of stress at work. Overall, the evidence indicates that incidence of stress-related cardiovascular disease is likely to be higher in the blue-collar occupations when the following factors are present: restricted discretion, shiftwork (particularly nightshift), effort-reward imbalance, high demands, poor psychosocial work environment, social isolation, physical inactivity or occupational violence. These risk factors may be interactive. Nurminen and Karjalainen (2001) estimated for Finland an attributable fraction of 16.9% (18.9% for men and 9.1% for women) for ischaemic heart disease due to the combined occupational risk factors of shiftwork, noise, and exposure to engine exhaust and environmental tobacco smoke. For ischaemic heart disease, Steenland et al. (2003) used an attributable fraction of 6–18% for individuals in the United States of America aged 24–64 years, based on the combined effects of noise, job strain (stress), shiftwork and environmental tobacco smoke. Occupational dermatitis accounts for about 10% of all occupational disease in the United States (Emmett 2002) but exposure data are lacking at global level.

Although there were adequate global data to analyse the risks to health care workers from contaminated sharps (e.g. syringe needles and scalpels), the full analysis has been omitted from this chapter. Since health care workers make up only 0.6% of the global population, the contribution to hepatitis B, hepatitis C and HIV/AIDS infections on a global level was close to zero. However, health care workers are at high risk of preventable infection from bloodborne pathogens, owing to occupational exposure to infected blood and body fluids.

1.2 Choice of theoretical-minimum-risk exposure distributions

For some occupational hazards, a theoretical minimum exposure of zero is not possible, as there is some low-level environmental exposure. Two occupational risk factors (carcinogens and airborne particulates) involve workplace exposure at concentrations higher than the environmental or background levels of these substances. For noise, the theoretical minimum was defined as less than 80 dBA, a level found not to have an increased risk of causing hearing loss (NIOSH 1998). For the other risk factors (ergonomic stressors and work-related risk factors for injuries), a category of workers with the lowest risks was identified as the com-

parison group for occupational categories of workers with higher risks. Thus, the theoretical minimum risk corresponds to "no occupational exposure above levels found in the defined comparison group". Selection of a defined comparison group provides a realistic basis for a theoretical minimum, but it does not establish the lowest rate of adverse outcome that could ever be experienced. While it is not expected that occupational exposures will be eliminated in the foreseeable future, it is possible to control exposures through recognized industrial hygiene practices. Engineering controls (including prevention, substitution of materials, process automation, enclosure, process elimination, isolation of workers and process change) constitute effective methods of minimizing exposures (Burton 1997). Administrative controls (such as education and training, work practice controls, worker rotation, maintenance and housekeeping) provide another means of risk reduction.

1.3 Data sources

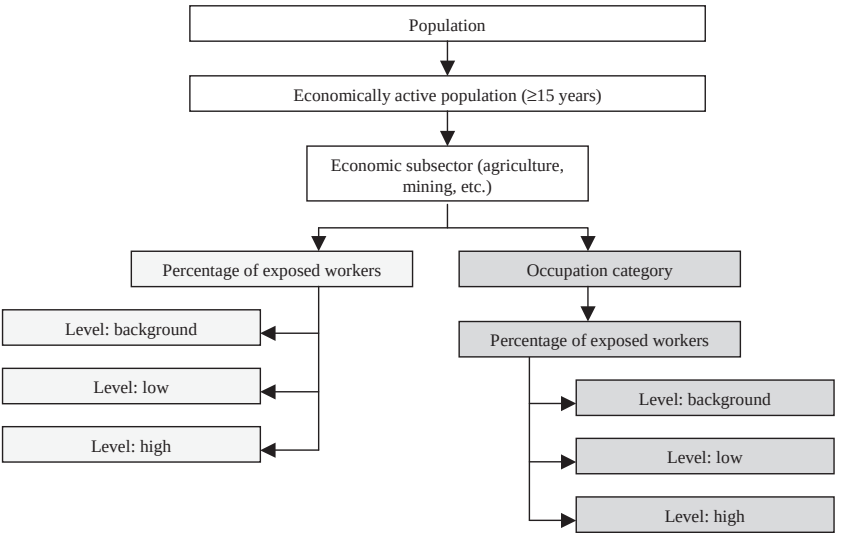
A systematic assessment of the literature was carried out to identify studies on occupational exposures and health outcomes. This included searching Medline, occupational health and safety databases such as OSHROM and NIOSHTIC and databases of various organizations; reviewing relevant references cited in publications identified through the initial literature search and of references cited in these secondary references; communicating with relevant experts; and seeking other information recommended by referees following the initial review of the draft manuscript. PubMed was searched using keywords for exposures and health outcomes, including (separately and in combination, with no limit on year of publication): exposure, occupational, cancer, carcinogen, silica, silicosis, benzene, asbestos, asbestosis, pneumoconiosis and developing country. Names of regions (e.g. Africa, Asia) and specific countries were also used as keywords. A systematic search was conducted using Ovid Healthstar and the former HealthSTAR databases, covering the period 1975–2001. Keywords included: asbestos; asthmagens; chronic obstructive lung disease; cancer and diesel exhaust; arsenic; benzene and leukaemia; ionizing radiation and leukaemia; back (for low back pain); injury; accidents; ergonomics; and hearing impairment and noise.

Studies of large populations, reviews and meta-analyses were specifically sought. Reports and publications were critically assessed to determine their methodology, validity and the characteristics of the population studied.

1.4 Estimating risk factor levels

In general, since the types of risk factor to which workers are exposed are primarily influenced by where the work is performed (economic sector) and the type of work they do (occupation), the assessment of proportion of population exposed in each subregion was based on (Figure 21.2):

Figure 21.2 Exposure assessment overview



- economic sector distribution (total nine sectors), used for carcinogens and agents leading to COPD) (Equation 1);
- occupational distribution (occupation within economic sector) (total seven occupations), used for asthmagens, noise and ergonomic stressors (Equation 2); and
- exposure could not be estimated for injury risk factors, and thus estimates of disease burden were made based on the reported rates of the outcome (injury mortality) rather than on exposure.

$$\mathrm{PEP}(r,g,a) = \mathrm{EAR}(r,g,a) \times \mathrm{OT}(r) \times \mathrm{EPF}(r) \sum_{i=1}^{9} (\mathrm{PW}(\mathrm{es}(r,g)i) \times \mathrm{PEW}(\mathrm{es}(r,g)i) \quad (1)$$

$$\mathrm{PEP}(r,g,a) = \mathrm{EAR}(r,g,a) \times \mathrm{OT}(r) \times \mathrm{EPF}(r) \sum_{i=1}^{7} (\mathrm{PW}(\mathrm{oc}(r,g)i) \times \mathrm{PEW}(\mathrm{oc}(r,g)i) \quad (2)$$

where

PEP(r,g,a) = proportion of the population occupationally exposed to a specific risk factor in that subregion, by sex and age, at low or high level

EAR(r,g,a) = economic activity rate, by subregion, sex and age

OT(r) = occupational turnover, if applicable, to account for workers exposed in the past, by subregion

EPF(r) = exposure partitioning factor, by subregion, to delineate proportion exposed at low or at high levels

PW(es(r,g)i) = proportion of the population working in economic subsector (i), by subregion and sex

PEW(es(r,g)i) = proportion of workers in economic subsector (i) with exposure to the specific risk factor, by subregion and sex

PW(oc(r,g)i) = proportion of the population working in occupational category (i), by subregion and sex

PEW(oc(r,g)i) = proportion of workers in occupational category (i) with exposure to the specific risk factor, by subregion and sex

The differences between the two equations are the term PW(es(r,g)i) in Equation 1, which is used when exposure data are available by economic sector, and the term PW(oc(r,g)i) in Equation 2, which is used when exposure data are available by occupational category. Occupational turnover (OT), defined as "the rate of replacement of workers due to departures from the workplace", was utilized for carcinogens because health effects due to these risk factors occur many years after exposure (latency) and it was therefore necessary to know how many persons had been exposed in the past to these risk factors. The effects of noise, ergonomic stressors and risk factors for injuries are relatively immediate; latent effects were therefore not a consideration for these risk factors. Additional detail on each term is provided below in the text, and is also summarized in Table 21.1.

The primary data sources used for the exposure assessment and the determination of some of the risk measures (see Table 21.2) included: the World Bank, ILO, the European Union carcinogen exposure (CAREX) database, published literature on prevalence and level of exposure to occupational risk factors, and published literature on epidemiology of health outcomes linked to occupational risk factors, as cited in the relevant sections for each risk factor.

Economic activity rate

EAR is defined as the proportion of the economically active population (EAP) among the overall population. EAR was calculated for each region and sex in persons aged ≥15 years, and used to estimate the proportion of the population potentially exposed to occupational risks. EAR provides the most comprehensive accounting of persons who may be exposed to occupational risks, as it includes people in paid employment, the self-employed, and people who work to produce goods and services

Table 21.1 Summary of determinants of population exposure to occupational risk factors

Term	*Comments*	*Application*	*Primary data sources*
EAP	Economically active population is calculated by application of the EAR to the national population	Used for injuries	ILO (2002a)
EAR	Economic activity rate, calculated as the EAP in each age group compared to the number of people in that age group, males and females ≥15 years	Used in exposure assessments of all risk factors	ILO (2002a)
PW(es)	Proportion working, i.e. fraction of EAP in economic sector. Data on distribution of EAP into three economic sectors (agriculture, industry, service) or nine economic subsectors. Country-level data were weighted by working-age population to develop subregional averages	Used for carcinogens, selected airborne particulates (agents leading to COPD)	World Bank (2001)
PW(oc)	Proportion working, i.e. fraction of EAP in occupational category. Country-level data for about 31 countries were weighted by working-age population to develop subregional averages. Owing to lack of country-level data, EMR-B was based on EMR-D data, EUR-C was based on EUR-B data and WPR-A was based on AMR-A data	Used for asthmagens, noise, ergonomic stressors	ILO (1995a); World Bank (2001)
PEW	Proportion exposed working, i.e. fraction of population working in economic sector (or in an occupational category) with exposure to risk factor. Owing to data limitations, data from developed countries were usually applied to developing countries, verified where possible by data on specific risk factors from specific countries	PEW(es): carcinogens PEW(es): selected airborne particulates (agents leading to COPD) PEW(oc): selected airborne particulates (asthmagens) PEW(oc): noise PEW(oc): ergonomic stressors	FIOH (1999); Kauppinen et al. (2000) FIOH (1999); Kauppinen et al. (2000); Korn et al. (1987); USEIA (2001) Karjalainen et al. (2002); Kogevinas et al. (1999) NIOSH (1998) Leigh and Sheetz (1989)
EPF	Exposure partitioning factor, i.e. proportion of PEW with low- or high-level exposure to risk factor	Carcinogens, selected airborne particulates, noise	NIOSH (1998, 1999, 2000a); Pearce et al. (1994); Yin et al. (1987)
OT	Occupational turnover factor. Used only for risk factors for which latent effects must be considered (carcinogens, selected airborne particulates). A factor of 4 was estimated on the basis of published data on labour turnover rates, published cohort data and modelling of cohorts with various mean lengths of exposure. Higher value used for specific regions for coal mining	Carcinogens	K. Steenland, personal communication, 2002

Table 21.2 Key sources, data supplied and special characteristics of the sources used to estimate exposure

Source	*Data supplied*	*Comments*
ILO (1995a, 2000, 2002b)	Employment in economic sectors and subsectors, and in occupations within economic sectors; EARs by age and sex for selected countries	Collected by national EAP surveys. Differences among and within countries (e.g. applicable ages, time period covered) limit international comparability
World Bank (2001)	Distribution of EAP (males and females) in agriculture, industry and services; participation of females in the EAP	Based on ILO data
FIOH (1999); Kauppinen et al. (2000)	Proportion of the working population with occupational exposure to carcinogens in the European Union, by economic sector and subsector, at the 3-digit classification level	Applicable to A subregions, extrapolated to B, C, D and E subregions
EIA (2001)	Country-level data on coal production	
ILO (1995b)	Country-level data on number of coal miners	
NIOSH (1991, 1998); USDHHS (1986)	Data on noise exposure of American workers	Applicable to A subregions, extrapolated to B, C, D and E subregions

for their own household consumption. According to ILO (2002b), the majority of those who work in the informal sector are included in the "employed" category, and the remainder are in the "unemployed" category; thus, the informal sector workers are included in this analysis. At the same time, persons in precarious or contingent employment often face an increased risk of occupational health and safety hazards, which are not quantified here (Quinlan 2002). The use of EAR for persons aged ≥15 years excludes children under 15 who work.

Estimates and projections of EAP were developed by ILO by applying estimates and projections of activity rates, by sex and age group, to the population estimates and projections assessed by the United Nations (ILO 1996). ILO estimates and projections of economic activity are taken primarily from population censuses and/or sample surveys carried out between 1975 and 1994. ILO also takes data from specific publications by national, interregional and/or international institutions.

Country-level data from the ILO electronic database were used to develop subregion-specific EARs for ages 15 years and above, for males and females (Table 21.3). EARs were estimated for 60–69-year olds by using data for 60–64-year olds. Data for people aged ≥65 years were

Table 21.3 Economic activity rates by subregion, sex and age group

		Age group (years)						
Subregion[a]	*Sex*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*	*Total ≥15*
AFR-D	Male	0.77	0.97	0.95	0.85	0.65	0.33	0.85
	Female	0.50	0.61	0.62	0.48	0.28	0.14	0.53
AFR-E	Male	0.78	0.97	0.95	0.86	0.66	0.33	0.86
	Female	0.64	0.72	0.69	0.54	0.36	0.18	0.65
AMR-A (95%)	Male	0.70	0.93	0.87	0.50	0.13	0.07	0.73
	Female	0.64	0.81	0.71	0.32	0.07	0.04	0.59
AMR-B	Male	0.78	0.97	0.89	0.66	0.33	0.17	0.82
	Female	0.46	0.53	0.39	0.20	0.07	0.04	0.42
AMR-D	Male	0.71	0.98	0.96	0.86	0.61	0.31	0.82
	Female	0.38	0.48	0.39	0.29	0.17	0.09	0.39
EMR-B (90%)	Male	0.66	0.97	0.92	0.74	0.45	0.23	0.79
	Female	0.33	0.37	0.26	0.18	0.09	0.05	0.31
EMR-D (40%)	Male	0.73	0.97	0.94	0.76	0.44	0.22	0.82
	Female	0.37	0.43	0.37	0.25	0.12	0.06	0.37
EUR-A	Male	0.66	0.96	0.84	0.35	0.05	0.03	0.68
	Female	0.59	0.74	0.56	0.14	0.02	0.01	0.47
EUR-B	Male	0.72	0.96	0.80	0.41	0.22	0.11	0.74
	Female	0.56	0.77	0.59	0.23	0.12	0.06	0.54
EUR-C	Male	0.72	0.97	0.89	0.30	0.11	0.06	0.74
	Female	0.61	0.94	0.74	0.17	0.05	0.03	0.58
SEAR-B	Male	0.74	0.98	0.94	0.73	0.44	0.22	0.83
	Female	0.55	0.70	0.65	0.44	0.21	0.11	0.58
SEAR-D (95%)	Male	0.77	0.98	0.95	0.72	0.53	0.27	0.85
	Female	0.45	0.57	0.50	0.32	0.16	0.08	0.47
WPR-A	Male	0.67	0.97	0.95	0.69	0.30	0.15	0.76
	Female	0.57	0.70	0.67	0.36	0.13	0.07	0.52
WPR-B (90%)	Male	0.81	0.98	0.92	0.61	0.29	0.15	0.84
	Female	0.77	0.89	0.67	0.29	0.09	0.05	0.71

[a] When data were not available for all countries, the percentage of the regional working age population (≥15 years) represented by data is indicated. Some very small countries, e.g. Grenada, were not included in these calculations.

Source: ILO (2002a).

applied to the 70–79 age group. The ≥80 age group was estimated at one half of the rate for the ≥65 age group (by comparison with country-level data, which is reported by some countries for elderly workers) (ILO 2001).

PROPORTION OF THE POPULATION WORKING IN AN ECONOMIC SECTOR OR OCCUPATIONAL CATEGORY

The distinction between "where people work", i.e. economic sector and "what they do", i.e. occupation, is important in exposure characterization. For example, within the economic subsector of manufacturing there

Table 21.4 Illustration of the ISIC classification system used in exposure assessment

Economic sector	*Economic subsectors*	*Occupational categories*						
		Professional	*Administration*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*
Agriculture								
Industry	Mining							
	Manufacturing							
	Electrical							
	Construction							
Services	Trade							
	Transport							
	Finance							
	Services							

Source: ILO 1987.

are people who work as production workers, but also people who work as clerical or sales people (Table 21.4). EAP was used for injuries. EAP by economic sector and subsector was used for carcinogens and agents leading to COPD, because available data do not distinguish exposures by occupational category within economic sectors. For asthmagens, noise and ergonomic stressors, the analyses were conducted on the basis of exposure by occupational category within economic sectors.

The approach used here is based on the International Standard Industrial Classification of All Economic Activities (ISIC), an economic classification system of the United Nations, which organizes all economic activities by economic sectors and relevant subgroupings (ILO 1987; UN 2000). The ISIC system is used almost universally by national and international statistical services to categorize economic activity, and therefore allowed us to make global comparisons. Table 21.4 illustrates the ISIC classification scheme of economic sectors, economic subsectors and occupational categories that were used to estimate exposures to workers in this project. We did not subdivide agriculture into economic subsectors.

Economic sector

For each subregion, a weighted proportion of working men and women (EAP) in each of the three economic sectors was constructed (Table 21.5) (World Bank 2001, data from 1990 and 1996–1998). Economic sector employment data were used to subdivide the number of workers in industry into the economic subsectors of mining, manufacturing, electricity (and other utilities) and construction. In a similar manner, the data for the service sector were subdivided into the economic subsectors of trade, transport, finance and services. The agriculture sector was not subdivided.

Table 21.5 EAP distribution in economic sectors and subsectors, by subregion and sex

			Industry				*Services*				
Subregion[a]	*Sex*	*Agriculture*	*Mining*	*Manufacturing*	*Electricity*	*Construction*	*Trade*	*Transport*	*Finance*	*Services*	*Sum*
AFR-D (20%)	Male	0.55	0.01	0.09	0.01	0.04	0.06	0.04	0.03	0.16	1.00
	Female	0.68	0.00	0.05	0.00	0.01	0.06	0.03	0.02	0.16	1.00
AFR-E (20%)	Male	0.55	0.01	0.09	0.01	0.04	0.05	0.03	0.03	0.18	1.00
	Female	0.65	0.00	0.03	0.00	0.00	0.05	0.02	0.02	0.22	1.00
AMR-A (95%)	Male	0.05	0.01	0.21	0.01	0.09	0.21	0.06	0.10	0.26	1.00
	Female	0.02	0.00	0.09	0.00	0.01	0.23	0.03	0.14	0.47	1.00
AMR-B	Male	0.20	0.01	0.15	0.04	0.08	0.10	0.05	0.25	0.12	1.00
	Female	0.12	0.00	0.12	0.00	0.01	0.14	0.01	0.05	0.55	1.00
AMR-D (70%)	Male	0.07	0.01	0.16	0.01	0.11	0.18	0.07	0.09	0.30	1.00
	Female	0.03	0.00	0.12	0.00	0.00	0.31	0.01	0.06	0.46	1.00
EMR-B (5%)	Male	0.15	0.01	0.16	0.03	0.11	0.15	0.08	0.05	0.27	1.00
	Female	0.09	0.00	0.09	0.00	0.01	0.20	0.05	0.09	0.47	1.00
EMR-D (20%)	Male	0.45	0.00	0.11	0.01	0.07	0.30	0.00	0.03	0.02	1.00
	Female	0.68	0.00	0.10	0.01	0.01	0.20	0.00	0.00	0.01	1.00
EUR-A	Male	0.06	0.01	0.27	0.01	0.11	0.01	0.00	0.05	0.48	1.00
	Female	0.05	0.00	0.15	0.00	0.02	0.01	0.01	0.12	0.64	1.00

continued

Table 21.5 EAP distribution in economic sectors and subsectors, by subregion and sex (*continued*)

			Industry				*Services*				
Subregion[a]	*Sex*	*Agriculture*	*Mining*	*Manufacturing*	*Electricity*	*Construction*	*Trade*	*Transport*	*Finance*	*Services*	*Sum*
EUR-B (70%)	Male	0.29	0.02	0.20	0.02	0.08	0.07	0.04	0.07	0.20	1.00
	Female	0.44	0.00	0.16	0.01	0.01	0.03	0.01	0.04	0.30	1.00
EUR-C (35%)	Male	0.21	0.04	0.14	0.03	0.15	0.05	0.20	0.12	0.06	1.00
	Female	0.16	0.03	0.12	0.02	0.04	0.12	0.24	0.12	0.15	1.00
SEAR-B (30%)	Male	0.46	0.01	0.12	0.01	0.07	0.14	0.06	0.01	0.13	1.00
	Female	0.45	0.00	0.15	0.00	0.01	0.22	0.01	0.01	0.15	1.00
SEAR-D (80%)	Male	0.53	0.02	0.13	0.02	0.03	0.01	0.06	0.04	0.17	1.00
	Female	0.80	0.01	0.10	0.00	0.01	0.00	0.00	0.02	0.06	1.00
WPR-A	Male	0.05	0.00	0.24	0.01	0.13	0.17	0.10	0.09	0.21	1.00
	Female	0.06	0.00	0.17	0.00	0.04	0.27	0.04	0.11	0.31	1.00
WPR-B (95%)	Male	0.44	0.03	0.14	0.01	0.05	0.09	0.06	0.02	0.16	1.00
	Female	0.40	0.01	0.12	0.01	0.01	0.17	0.06	0.06	0.17	1.00

[a] When data were not available for all countries, the percentage of the regional working age population (≥15 years) represented by data is indicated. Some very small countries, e.g. Grenada, were not included in these calculations.

Note: AFR-D and AFR-E (combined), EMR-B, EMR-D and SEAR-D data are based on 1990 employment data from the World Bank world development indicators, as EAP data were very limited. All others are taken from 1996–1998 World Bank EAP data. Subregional averages were calculated using country values weighted by the working-age population.

Occupational category

Regional tables of occupation within economic sector distributions were constructed using the number of employed people by occupation and economic sector. For comparison purposes, data were obtained from one source (ILO 1995a). For a subregion with only one country represented, the distribution of occupation within economic sector was assumed to represent the regional employment patterns. Where more than one country was represented, a weighted average was constructed. Where there were no data for the subregion, patterns for the most similar subregion were applied (EMR-B based on EMR-D, EUR-C based on EUR-B and WPR-A based on AMR-A). Because of limited data on occupational distribution by sex within an economic sector, the same distribution (i.e. proportional division) was applied within a subregion to ages 15 and above, and to males and females. The A subregions had higher proportions of EAP in the professional, managerial and administrative categories, while the B, C, D and E subregions had proportionally more workers in the production categories.

Proportion of workers in an economic sector or occupational category with exposure

Worldwide data on worker exposure are limited. Therefore, several assumptions were made, validated where possible, to establish the proportion of workers exposed to a specific risk factor within an economic sector (PEW). More detail is presented in the sections on specific risk factors.

Exposure partitioning factor (EPF)

In order to partition into high and low exposure groups those workers exposed to carcinogens, we chose the United States Occupational Health and Safety Administration (OSHA) Permissible Exposure Levels (PELs). For most carcinogens we were then able to estimate the risks for the low and high exposure groups from the literature.

The OSHA PELs state a level of the agent that can never be exceeded in the workplace (usually based on eight-hour time-weighted average exposures), and these have had the force of law in the United States as maximum limits of exposure since the creation of OSHA in 1971. Similar occupational exposure limits (OELs) have been promulgated as law by many countries, particularly in the A subregions, and as recommendations by professional expert groups. It is generally considered that a long-term mean exposure in a "minimally controlled" work environment will be in the range 0.3–0.5 times the PEL (Hewett 1996). For example, the American Industrial Hygiene Association suggests that a typical long-term average exposure may be one third of an eight-hour PEL (Roach 1992).

A different approach was used for asthmagens and agents leading to COPD. The actual disease-causing exposures themselves, within these occupations, are either generic (e.g. dust) or too numerous to be useful (e.g. there are over 200 known asthmagens). In both instances there were no international data on the number of workers exposed, which dictated the approach of using occupations or economic subsectors. For asthma, different relative risks were available for eight large occupational groups, while for COPD we partitioned the overall relative risk for the exposed population into high and low relative risks, and assigned these to different economic subsectors according to Korn et al. (1987).

Occupational turnover (OT)

Cancers and lung diseases have long latency periods and once the disease process has begun the worker continues to be at risk, even after exposure ceases. This means that persons who were exposed in the past must be considered as ever-exposed, even if they are currently working in non-exposed jobs or have retired. Furthermore, OT increases the number of persons ever exposed to an occupational risk. This approach was consistent with cohorts represented in the epidemiological studies from which relative risks were taken. The OT factor was not utilized in estimating the numbers of workers exposed to noise, ergonomic stressors or risk factors leading to occupational injuries, as these risk factors do not have latent effects. No turnover was estimated for asthma and COPD owing to a lack of sufficient information on the applicability to studies in which relative risk was measured. Table 21.6 presents data from the literature on annual OT rates in various countries and industries throughout the world, organized by subregion. These reports did not indicate if employees were new to the job or to the industry, although several studies were at the company level, indicating that the worker was new to the company. Therefore, to account for previously unexposed workers entering jobs with carcinogen or dust exposures, an annual turnover rate (ATR) of 10% was selected for all subregions.

An adjustment factor (noted as OT) to account for annual turnover in jobs with exposure to occupational carcinogens was determined as follows:

Computation of adjustment factor to correct for occupational turnover (OT)

$$\begin{aligned}\text{Adjustment factor, } OT &= P_t/P_0 \\ &= [\text{original workers} + \text{new workers} - \text{deaths}]/\text{original workers} \\ &= \{P_0 + [P_0 \times ATR \times t] - [(\text{mortality rate})(P_0 + (P_0 \times ATR \times t)]\}/P_0\end{aligned} \quad (3)$$

where

P_t = the proportion who have ever been occupationally exposed, during a period of 40 years, still living

Table 21.6 Turnover rates in various industries and countries

Country or area	*Basis of measurement*	*Annual turnover rate (ATR)*	*Comments*	*Source*
A subregions				
Italy	Metal—mechanical engineering industry	13.4%	Industry level, based on 2729 observations	Lucifora (1998)
Italy		26%	Total worker turnover rate, including accession and separation	Lucifora (1998)
Spain				
Basque	Industrial production cooperative (manufacturing)	3%	65 firms—cited as low rate	Johnson and Whyte (1977)
United States	Restaurant industry	500%	8 southern restaurants	Butler and Twaddle (1979)
United States	Garment manufacturing	140%	153 female workers at a single plant in the south-west	Koch and Rhodes (1981)
United States	One interstate trucking firm	40%	Expected rate for 1997—truck drivers	EIU (1997a)
United States				
California	Silicon Valley, one financial firm	25%	Software services group	EIU (1997f)
United States				
New Mexico	State-wide survey by New Mexico Department of Labor, January–March 2001	25% per quarter, ranging from 29% in agriculture to 15% in public administration	Agriculture rates show greatest seasonal variation	Moffett (2002)
B subregions				
Brazil	Brazilian labour market	47%	Cited as higher than most markets for which data are available	EIU (1997b)

continued

Table 21.6 Turnover rates in various industries and countries (*continued*)

Country or area	*Basis of measurement*	*Annual turnover rate (ATR)*	*Comments*	*Source*
China	Foreign Investment Enterprise, key managers and technicians	Support staff median rate, 7%, up to >50%; middle management median rate 4%, up to >35%		EIU (1996a)
China	One manufacturer of spun polyester	40% (1992–1993) 21% (1994) 17% (1995) 7% (expected for 1996)	Turnover reductions due to establishment of housing programme	EIU (1996b)
China	Foreign Investment Enterprise	11.8% (1999) 13.4% (2001)	Staff turnover varies widely by subregion, industry and type of enterprise	EIU (2001a, 2002)
China Hong Kong Special Administrative Region	Expatriate teachers, 2000	25%		EIU (2001c)
Eastern Caribbean	Informatics firms	2%	Cited as low rate	EIU (1997d)
Eastern Europe	Security personnel	160%		EIU (1998)
Lao People's Democratic Republic	Vientiane University College	Half of staff have 2-year tenure, average elsewhere is 3 months	Improved owing to staff training	EIU (1997c)
Republic of Korea	National level	3%	Cited as low rate	EIU (1997e)
D subregions				
India	Managers	Minimum 10% estimated	Mostly foreign-invested ventures	EIU (1995)
India	One computer manufacturer	9% (1994)	200 employees	EIU (1995)
E subregions				
Uganda	Uganda Railway Corporation (mid-1990s)	15%	Annual employee turnover rate	EIU (2001b)

P_0 = the proportion who are occupationally exposed at time t=0

ATR = turnover/year, taken as 0.10

t = time, taken as 40 years, a typical working lifetime

mortality = 20% of total cohort, based on published death rates of about 5 deaths per thousand over a period of 40 years (Minino and Smith 2001).

Equation 3 results in an adjustment factor of OT = 4 to correct for occupational turnover over a 40-year period with a median exposure duration of 10 years.

In addition to knowing the numbers of workers exposed to agents with latent health effects, in some cases it was also useful to know the duration of exposure to agents with latent effects for outcomes for which the risks were based on cumulative exposure. Cohort modelling was conducted to determine the typical duration of exposure (K. Steenland, personal communication, 2002). This modelling assumed that people worked for a maximum of 40 years, that 10% of the workers were replaced each year, and that 20% died over the 40-year period. Exposure durations were randomly selected from a log-normal distribution. Persons were also randomly assigned a starting age at entry between 20 and 45 years, and were assumed to retire at age 65 years if they had not already left the cohort by that age. A steady-state working population was produced by using a log-normal distribution for exposure with a geometric mean of 9 years. Using this, the mean length of exposure (in years) at the end of 40 years could be estimated (by age) for all persons ever exposed in the cohort. The average estimated length of exposure, as shown in Table 21.7, was 9.8 years, which is consistent with data on a wide range of cohorts presented in the published literature (Steenland et al. 1991a, 1991b, 2001b).

1.5 Risk factor–disease relationship

Risk measures (relative risks or mortality rates) for the health outcomes resulting from exposure to the risk factors considered in this study were determined primarily from peer-reviewed, published studies. Adjustments were made, as appropriate, to account for differences in levels of exposure, exposure duration and/or age, sex and subregion.

- For carcinogens leading to cancer of the lung, trachea or bronchus, and for leukaemogens, composite values were taken from the literature and adapted to exposure patterns in the various subregions.
- For asthma, the relative risks for different occupations were taken from Karjalainen et al. (2002), with the exception of work in agriculture, for which the relative risk was taken from Kogevinas et al. (1999).

Table 21.7 Exposure duration after 40 years in model cohort

Age group (years)	*Number*	*Total exposure (years)*	*Average exposure (years)*
15–29	12	50	4.2
30–44	86	575	6.7
45–59	117	1182	10.1
60–69	105	1195	11.4
70–79	53	618	11.7
≥80	19	234	12.3
All ages	392	3854	9.8

- For COPD, the relative risks for different economic subsectors were taken from Korn et al. (1987).
- For noise, relative risks of noise-induced hearing loss were calculated from data on hearing loss in workers with different levels of noise exposure in the United States (NIOSH 1998).
- The relative risks of low back pain, given employment in different occupational categories, were taken from Leigh and Sheetz (1989).
- Owing to heterogeneity of factors leading to occupational injuries, relative risks could not be extrapolated from one setting to another. As a result, the mortality rates for workers exposed to risk factors leading to injuries were estimated for different subregions from various sources, including Laborsta (ILO 2001).

2. Occupational carcinogens

The International Agency for Research on Cancer (IARC 2002) has classified 150 chemical or biological agents or exposure situations as known or probable human carcinogens. IARC has classified 87 agents, mixtures or exposure circumstances as Group 1 (carcinogenic to humans), including various chemical compounds, pharmaceuticals and bacterial and viral infections. Many are encountered in occupational settings, e.g. asbestos and cadmium. An additional 63 agents, mixtures or exposure circumstances have been classified as Group 2A (probably carcinogenic to humans). Those with occupational significance include diesel fumes and benzidine-based dyes (IARC 2001). Although IARC classifies agents according to their overall carcinogenicity, specific sites are also considered.

Work-related malignant conditions can arise from a large variety of occupational exposures. However, the main groups of conditions are relatively few—lung cancer and leukaemia. The exposures selected for assessment in this study were based on how common they may be, the

risk arising from exposure, the strength of evidence and the availability of data. Table 21.8 shows the definition of each of the chemical and physical agents, along with the related cancer.

The analysis included relevant Group 1 and 2A carcinogens, with the following exceptions.

- Tetrachloroethylene and trichloroethylene, both classified in Group 2A, were not included as carcinogens because the evidence for cancer is weak.
- The aromatic amines and dyes, including 2-naphthylamine, benzidine-based dyes, benzidine and 4,4'-methylenebis(2-chloroaniline) (also known as MOCA) were excluded owing to lack of data for developing countries.
- Occupational carcinogens with extremely limited exposures (e.g. bis-chloromethyl ether, also known as BCME) were not included.
- Compounds for which exposure estimates were not available from the CAREX database (e.g. soot, xenylamine, 4-nitrobiphenyl and polycyclic aromatic hydrocarbons) were not included.
- Although radon is an IARC Group I carcinogen with large estimated exposures, it was excluded from consideration owing (i) to worldwide differences in naturally occurring radon emissions, (ii) to wide variations in climate and construction methods, which substantially affect the concentration of radon retained in buildings, and (iii) to difficulties in separating occupational and nonoccupational radon exposures.

Other conditions have insufficient relevant exposure data, insufficient risk data or insufficient number of cases worldwide to allow them to be usefully included. These conditions include:

- bladder cancer (aromatic amines, benzidine dyes, MOCA);
- liver (vinyl chloride);
- nasal cavity and middle ear (hardwood dust, chromium VI compounds, nickel compounds);
- bone and articular cartilage (ionizing radiation);

Table 21.8 Occupational carcinogens and health outcomes

Occupational carcinogen	*Outcome*
Arsenic, asbestos, beryllium, cadmium, chromium, diesel exhaust, nickel, silica	Cancer of the trachea, bronchus or lung
Benzene, ethylene oxide, ionizing radiation	Leukaemia

- skin (arsenic, by-products of distillation, ionizing radiation); and
- lung cancer due to passive smoking in the workplace.

2.1 Exposure variable and theoretical-minimum-risk exposure

Exposure was divided into three categories: background, low and high. The occupational risk factors for cancer involve workplace exposure, at concentrations higher than background level, to various chemical and physical agents that are known to cause malignant neoplasms. Thus, the theoretical minimum risk corresponds to "no occupational exposure to physical, chemical or biological agents or other factors above background levels".

2.2 Estimating risk factor levels

The general exposure assessment methodology was described earlier. This assessment was based on the distribution of the EAP by economic subsector, because the primary exposure data sources used in this analysis organized carcinogen exposure data by economic subsector (Equation 1). The regional distributions of workers into economic subsectors were adjusted by data on the carcinogens to which people in the various economic subsectors were exposed. As described earlier, an adjustment factor of 4 was used to account for turnover in jobs with exposure to occupational carcinogens.

The primary data source on work-related exposure to carcinogens for each economic subsector (PEW(es(r,g)i) in Equation 1) is the CAREX database (FIOH 1999), which presents data on the number of workers in the European Union exposed to 139 carcinogens (IARC Group 1, 2A and selected 2B agents) at levels above background in 1990–1993. Table 21.9 lists the CAREX data for the carcinogens in our study. These estimates were based on national workforce data and exposure prevalence estimates from Finland and the United States, adjusted for the economic structure of each country, then refined by national experts.

It was assumed that the proportion of workers exposed to a particular carcinogen in a specific economic subsector was constant throughout the world. To check the validity of this assumption, the literature was searched for estimates of the number of workers exposed to silica. Silica was chosen as an indicator because there are more data on silica available for developing countries than on other carcinogens. This search yielded a range of study types, from rough estimates (Zou Changqi et al. 1997) to studies in which air concentrations were measured in workplaces (Yin et al. 1987). Estimates of the number of workers exposed to silica in China, Thailand and Viet Nam, and to benzene in China, were compared to the number of persons employed in that country, either in a specific economic sector or overall. The results obtained were compared with CAREX data. With a few exceptions, the estimated fraction

Table 21.9 Mean proportions of workers exposed to selected carcinogens, by economic sector and subsector, in the European Union

Carcinogen	*Agriculture*	*Mining*	*Manufacturing*	*Electrical*	*Construction*	*Trade*	*Transport*	*Finance*	*Services*
Silica	0.00372	0.23049	0.02327	0.01415	0.18860	0.00017	0.00476	0.00002	0.00061
Cadmium	0.00000	0.00000	0.00487	0.00287	0.00291	0.00002	0.00065	0.00000	0.00047
Nickel	0.00000	0.02025	0.01680	0.00352	0.00047	0.00007	0.00003	0.00000	0.00043
Arsenic	0.00054	0.00072	0.00400	0.00148	0.00134	0.00006	0.00000	0.00002	0.00011
Chromium	0.00000	0.00346	0.02079	0.00409	0.00237	0.00017	0.00370	0.00000	0.00225
Diesel fumes	0.00646	0.21970	0.01110	0.03358	0.05816	0.00485	0.13438	0.00000	0.00914
Beryllium	0.00000	0.00055	0.00207	0.00070	0.00004	0.00002	0.00011	0.00000	0.00003
Asbestos	0.01248	0.10248	0.00590	0.01702	0.05203	0.00292	0.00684	0.00016	0.00284
Benzene	0.00100	0.00200	0.00300	0.00100	0.00100	0.01000	0.00500	0.00000	0.02000
Ionizing radiation	0.00000	0.01100	0.00000	0.03400	0.00000	0.00000	0.00400	0.00000	0.00000
Ethylene oxide	0.00012	0.00137	0.00060	0.00006	0.00027	0.00000	0.00002	0.00000	0.00057

Source: Calculated from CAREX (FIOH 1999).

of workers exposed to silica was equal to or higher in these countries than indicated by CAREX (Juengprasert 1997; T. Nguyen, personal communication, 2001; NIEHS 1999; Phan Hong Son et al. 1999; Yin et al. 1987; Zou Changqi et al. 1997). For example, the proportion of workers exposed to silica in manufacturing in Viet Nam is 3.7%, compared to the CAREX estimate of 2.3%.

It was assumed that, within a given economic subsector, both male and female workers and younger and older workers had the same probability of exposure. For example, if 2.3% of people working in manufacturing were exposed to silica, it was assumed that 2.3% of males and 2.3% of females working in manufacturing were exposed to silica, young and old alike. There were, however, fewer females working in manufacturing, so that at the population level the proportion of females with exposure to silica was lower than that of males.

There are few data on the distribution of exposure monitoring values, which are needed to accurately estimate the proportion of workers exposed to above or below a specific value (EPF(r) in Equation 1). Therefore, the demarcation between low and high exposure was established as the PELs enforced by OSHA. Some reasons for selecting the PELs as partitioning values include the following.

- Exposure data for the United States are often reported based on "compliance with" or "exceeding" the PELs.
- The risks corresponding to low or high exposure have been linked to the PELs.
- As cancers have long latency periods, the exposures of concern have occurred several decades in the past. The OSHA PELs for many carcinogens have not changed since their adoption in 1971, allowing a stable benchmark for comparison (Table 21.10).

The peer-reviewed literature was searched for studies that included proportions of workers exposed above and below particular levels. There are many reports of exposures to contaminants in the literature, and even on the distribution of exposures at low and high levels in developed countries. However, there are few data on distribution of exposure values for developing countries. A summary of the major sources used to decide how to partition exposure values for carcinogens for the B, C, D and E subregions is presented in Tables 21.11 and 21.12 for benzene and metals, respectively.

The following data were used to partition exposure for A subregions:

- Finnish data (Partanen et al. 1995), indicating 11–94% exposed above 0.2 mg/m^3 respirable silica in a range of industries;
- NIOSH (1999) estimates of proportions of workers exposed above the PELs of 4% (asbestos) and 13.6% (silica); and

Table 21.10 OSHA permissible exposure levels (PELs) for carcinogens

Chemical/ physical agent	*PEL*	*Source*	*Comment*
Arsenic	Inorganic: 10 μg/m^3 Organic: 0.5 mg/m^3	OSHA, 29 CFR 1910.1018 OSHA, 29 CFR 1910.1000, Table Z-1	Effective 1978 Effective 1971
Asbestos	Varied[a]		
Benzene	10 ppm	OSHA, 29 CFR 1910.1000, Table Z-2	Effective 1971
	1 ppm	OSHA, 29 CFR 1910.1028	Effective 1987
Beryllium	2 μg/m^3	OSHA, 29 CFR 1910.1000, Table Z-2	Effective 1971
Cadmium	Fume: 0.1 mg/m^3	OSHA, 29 CFR 1910.1000, Table Z-2	Effective 1971
	Dust: 0.2 mg/m^3	OSHA, 29 CFR 1910.1000, Table Z-2	Effective 1971
	5 μg/m^3	29 CFR 1910.1027	Effective 1992
Chromium	Chromic acid and chromates: 0.1 mg/m^3	OSHA, 29 CFR 1910.1000, Table Z-2 (ceiling)	Effective 1971
	Chromium metal: 1 mg/m^3	OSHA, 29 CFR 1910.1000, Table Z-1	Effective 1971
Diesel exhaust	NA		
Ethylene oxide	1 ppm	OSHA, 29 CFR 1910.1047	Effective 1984
Ionizing radiation	Rems/calendar quarter: whole body, 1.25; hands, forearms, feet, ankles, 18.75; skin, 7.5	OSHA, 29 CFR 1910.1096, Table G-18	Effective 1974
Nickel	Metal, insoluble and soluble compounds: 1 mg/m^3	OSHA, 29 CFR 1910.1000, Table Z-1	Effective 1971
Silica	Respirable quartz: (10 mg/m^3)/ (per cent SiO_2 + 2)	OSHA, 29 CFR 1910.1000, Table Z-3	Effective 1971. For 100% silica dust, this is equivalent to 0.1 mg/m^3. Halve this value for cristobalite and tridymite

NA Not applicable.

[a] As shown in this table, most of the PELs have not changed since they were put in place. However, there were considerable changes in the United States PEL for asbestos during the years of interest to the current analysis, with a level before 1972 of 12 fibres/ml before the first OSHA-issued PEL in 1972 decreasing, through several steps, to 0.1 fibres/ml in 1994 (Martonik et al. 2001; Nelson 1997).

Source: USDOL OSHA (2002a).

Table 21.11 Occupational exposure to benzene in developing countries

Country	*Industry*	*Concentration*	*Year (or year reported)*
Egypt	Rubber coating	0–74 mg/m^3	(1986)
Turkey	Shoemaking	48–96 mg/m^3 672 mg/m^3 (maximum level)	1970
India	Petrol pump	4.5 mg/m^3 (mean)	1991
China	Various: paint, chemical, varnish works, shoemaking	0.06–850 mg/m^3	(1987)
Brazil	Steel workers Petrochemical	960–3200 mg/m^3 per day 140 mg/m^3, maximum of personal samples	(1993) (1993)

Source: Pearce et al. (1994).

Table 21.12 Occupational exposures to metals in developing countries

Country	*Industry*	*Concentration*	*Year (or year reported)*
China	Tin mine	Arsenic: 0.42 mg/m^3, mean Arsenic: 0.01 mg/m^3	1952 1980s
China	Cadmium refining	Cadmium: 0.04–0.074 mg/m^3	1970s
Singapore	Storage battery factory	Cadmium: 0.13–58.3 mg/m^3, geometric means of three sets of samples	1980
China	Chromate production	0.02–21.3 mg/m^3 0.55 mg/m^3, mean	1960s–1980 (1989)

Source: Pearce et al. (1994).

- NIOSH (2000b) data on miners, indicating silica exposures above the PEL for 8% of coal mine samples, 16% of metal mine samples, 9% of stone mine samples and 8% of sand and gravel facility samples.

For the B, C, D and E subregions, important evidence includes:

- Chinese data (Dosemeci et al. 1995), indicating roughly three quarters of samples above 0.1 mg/m^3 respirable silica;
- a study of a South African brickworks (Myers et al. 1989), in which 45% of presented sample values were above 0.1 mg/m^3 and roughly two thirds and four fifths of samples in medium and dusty areas, respectively, were above 0.1 mg/m^3 respirable silica;
- a study of a South African pottery (Rees et al. 1992), where roughly three quarters of samples that included silica analysis were above the Threshold Limit Value (TLV); and

Table 21.13 Exposure partition factors for carcinogens for the A and for the B, C, D and E subregions

Subregion	*Proportion of exposed workers with low exposures (at or below the PEL)*	*Proportion of exposed workers with high exposures (above the PEL)*
A	0.90	0.10
B, C, D and E	0.50	0.50

- the Chinese benzene study (Yin et al. 1987), in which 35% of over 50 000 workplaces had concentrations at or above 40 mg/m^3, in comparison to the current OSHA PEL of 3.2 mg/m^3 for benzene, and in which the benzene concentration in 86% of 141 shoe factories was above 25 mg/m^3.

Based on these data, partition factors for carcinogen exposures were determined for the A and for the B, C, D and E subregions, as shown in Table 21.13.

Lung carcinogens

The proportions of the population exposed to the occupational lung carcinogens included in the study (Table 21.8) are shown in Tables 21.14 and 21.16 by subregion, age, sex and level of exposure.

Leukaemogens

The proportions of the population exposed to occupational leukaemogens (Table 21.8) are presented in Table 21.15 by subregion, age, sex and level of exposure.

2.3 Risk factor–disease relationships

Relative risk estimates were used for lung carcinogens and leukaemogens. Table 21.17 summarizes the chemical or physical agent, the specific cancer and the key data sources that provided evidence of the link between the two. These review studies assessed risk measures for the main sites of occupational cancer, including the lung (which, for the purposes of this study, includes the trachea, bronchus and lung), the haematopoietic system (represented in this study by leukaemia) and malignant mesothelioma.

Relative risks for lung cancer and leukaemia were taken from studies of cohorts of workers with variable exposure durations and intensities, variable periods from the last exposure and variable lengths of follow-up. They therefore compare exposed with unexposed groups. In preparing relative risk estimates for exposure outcomes of interest, several assumptions were made:

Table 21.14 Proportions of the population exposed to lung carcinogens by subregion, age, sex and level of exposure

			Age group (years)					
Subregion	*Sex*	*Exposure level*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	Background	0.837	0.837	0.837	0.837	0.837	0.837
		Low	0.082	0.082	0.082	0.082	0.082	0.082
		High	0.082	0.082	0.082	0.082	0.082	0.082
	Female	Background	0.934	0.934	0.934	0.934	0.934	0.934
		Low	0.033	0.033	0.033	0.033	0.033	0.033
		High	0.033	0.033	0.033	0.033	0.033	0.033
AFR-E	Male	Background	0.839	0.839	0.839	0.839	0.839	0.839
		Low	0.080	0.080	0.080	0.080	0.080	0.080
		High	0.080	0.080	0.080	0.080	0.080	0.080
	Female	Background	0.929	0.929	0.929	0.929	0.929	0.929
		Low	0.035	0.035	0.035	0.035	0.035	0.035
		High	0.035	0.035	0.035	0.035	0.035	0.035
AMR-A	Male	Background	0.802	0.802	0.802	0.802	0.802	0.802
		Low	0.178	0.178	0.178	0.178	0.178	0.178
		High	0.020	0.020	0.020	0.020	0.020	0.020
	Female	Background	0.936	0.936	0.936	0.936	0.936	0.936
		Low	0.058	0.058	0.058	0.058	0.058	0.058
		High	0.006	0.006	0.006	0.006	0.006	0.006
AMR-B	Male	Background	0.793	0.793	0.793	0.793	0.793	0.793
		Low	0.103	0.103	0.103	0.103	0.103	0.103
		High	0.103	0.103	0.103	0.103	0.103	0.103
	Female	Background	0.951	0.951	0.951	0.951	0.951	0.951
		Low	0.024	0.024	0.024	0.024	0.024	0.024
		High	0.024	0.024	0.024	0.024	0.024	0.024
AMR-D	Male	Background	0.761	0.761	0.761	0.761	0.761	0.761
		Low	0.119	0.119	0.119	0.119	0.119	0.119
		High	0.119	0.119	0.119	0.119	0.119	0.119
	Female	Background	0.961	0.961	0.961	0.961	0.961	0.961
		Low	0.019	0.019	0.019	0.019	0.019	0.019
		High	0.019	0.019	0.019	0.019	0.019	0.019
EMR-B	Male	Background	0.760	0.760	0.760	0.760	0.760	0.760
		Low	0.120	0.120	0.120	0.120	0.120	0.120
		High	0.120	0.120	0.120	0.120	0.120	0.120
	Female	Background	0.963	0.963	0.963	0.963	0.963	0.963
		Low	0.019	0.019	0.019	0.019	0.019	0.019
		High	0.019	0.019	0.019	0.019	0.019	0.019
EMR-D	Male	Background	0.840	0.840	0.840	0.840	0.840	0.840
		Low	0.080	0.080	0.080	0.080	0.080	0.080
		High	0.080	0.080	0.080	0.080	0.080	0.080
	Female	Background	0.955	0.955	0.955	0.955	0.955	0.955
		Low	0.023	0.023	0.023	0.023	0.023	0.023
		High	0.023	0.023	0.023	0.023	0.023	0.023

Table 21.14 Proportions of the population exposed to lung carcinogens by subregion, age, sex and level of exposure (*continued*)

			Age group (years)					
Subregion	*Sex*	*Exposure level*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
EUR-A	Male	Background	0.802	0.802	0.802	0.802	0.802	0.802
		Low	0.179	0.179	0.179	0.179	0.179	0.179
		High	0.020	0.020	0.020	0.020	0.020	0.020
	Female	Background	0.937	0.937	0.937	0.937	0.937	0.937
		Low	0.057	0.057	0.057	0.057	0.057	0.057
		High	0.006	0.006	0.006	0.006	0.006	0.006
EUR-B	Male	Background	0.779	0.779	0.779	0.779	0.779	0.779
		Low	0.111	0.111	0.111	0.111	0.111	0.111
		High	0.111	0.111	0.111	0.111	0.111	0.111
	Female	Background	0.920	0.920	0.920	0.920	0.920	0.920
		Low	0.040	0.040	0.040	0.040	0.040	0.040
		High	0.040	0.040	0.040	0.040	0.040	0.040
EUR-C	Male	Background	0.654	0.654	0.654	0.654	0.654	0.654
		Low	0.173	0.173	0.173	0.173	0.173	0.173
		High	0.173	0.173	0.173	0.173	0.173	0.173
	Female	Background	0.801	0.801	0.801	0.801	0.801	0.801
		Low	0.099	0.099	0.099	0.099	0.099	0.099
		High	0.099	0.099	0.099	0.099	0.099	0.099
SEAR-B	Male	Background	0.798	0.798	0.798	0.798	0.798	0.798
		Low	0.101	0.101	0.101	0.101	0.101	0.101
		High	0.101	0.101	0.101	0.101	0.101	0.101
	Female	Background	0.922	0.922	0.922	0.922	0.922	0.922
		Low	0.039	0.039	0.039	0.039	0.039	0.039
		High	0.039	0.039	0.039	0.039	0.039	0.039
SEAR-D	Male	Background	0.805	0.805	0.805	0.805	0.805	0.805
		Low	0.098	0.098	0.098	0.098	0.098	0.098
		High	0.098	0.098	0.098	0.098	0.098	0.098
	Female	Background	0.934	0.934	0.934	0.934	0.934	0.934
		Low	0.033	0.033	0.033	0.033	0.033	0.033
		High	0.033	0.033	0.033	0.033	0.033	0.033
WPR-A	Male	Background	0.745	0.745	0.745	0.745	0.745	0.745
		Low	0.230	0.230	0.230	0.230	0.230	0.230
		High	0.026	0.026	0.026	0.026	0.026	0.026
	Female	Background	0.914	0.914	0.914	0.914	0.914	0.914
		Low	0.078	0.078	0.078	0.078	0.078	0.078
		High	0.009	0.009	0.009	0.009	0.009	0.009
WPR-B	Male	Background	0.769	0.769	0.769	0.769	0.769	0.769
		Low	0.115	0.115	0.115	0.115	0.115	0.115
		High	0.115	0.115	0.115	0.115	0.115	0.115
	Female	Background	0.875	0.875	0.875	0.875	0.875	0.875
		Low	0.063	0.063	0.063	0.063	0.063	0.063
		High	0.063	0.063	0.063	0.063	0.063	0.063

Table 21.15 Proportions of the population exposed to leukaemogens by subregion, age, sex and level of exposure

			Age group (years)					
Subregion	*Sex*	*Exposure level*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	Background	0.980	0.980	0.980	0.980	0.980	0.980
		Low	0.010	0.010	0.010	0.010	0.010	0.010
		High	0.010	0.010	0.010	0.010	0.010	0.010
	Female	Background	0.989	0.989	0.989	0.989	0.989	0.989
		Low	0.005	0.005	0.005	0.005	0.005	0.005
		High	0.005	0.005	0.005	0.005	0.005	0.005
AFR-E	Male	Background	0.979	0.979	0.979	0.979	0.979	0.979
		Low	0.010	0.010	0.010	0.010	0.010	0.010
		High	0.010	0.010	0.010	0.010	0.010	0.010
	Female	Background	0.984	0.984	0.984	0.984	0.984	0.984
		Low	0.008	0.008	0.008	0.008	0.008	0.008
		High	0.008	0.008	0.008	0.008	0.008	0.008
AMR-A	Male	Background	0.973	0.973	0.973	0.973	0.973	0.973
		Low	0.025	0.025	0.025	0.025	0.025	0.025
		High	0.003	0.003	0.003	0.003	0.003	0.003
	Female	Background	0.970	0.970	0.970	0.970	0.970	0.970
		Low	0.027	0.027	0.027	0.027	0.027	0.027
		High	0.003	0.003	0.003	0.003	0.003	0.003
AMR-B	Male	Background	0.980	0.980	0.980	0.980	0.980	0.980
		Low	0.010	0.010	0.010	0.010	0.010	0.010
		High	0.010	0.010	0.010	0.010	0.010	0.010
	Female	Background	0.977	0.977	0.977	0.977	0.977	0.977
		Low	0.011	0.011	0.011	0.011	0.011	0.011
		High	0.011	0.011	0.011	0.011	0.011	0.011
AMR-D	Male	Background	0.967	0.967	0.967	0.967	0.967	0.967
		Low	0.016	0.016	0.016	0.016	0.016	0.016
		High	0.016	0.016	0.016	0.016	0.016	0.016
	Female	Background	0.979	0.979	0.979	0.979	0.979	0.979
		Low	0.010	0.010	0.010	0.010	0.010	0.010
		High	0.010	0.010	0.010	0.010	0.010	0.010
EMR-B	Male	Background	0.969	0.969	0.969	0.969	0.969	0.969
		Low	0.015	0.015	0.015	0.015	0.015	0.015
		High	0.015	0.015	0.015	0.015	0.015	0.015
	Female	Background	0.984	0.984	0.984	0.984	0.984	0.984
		Low	0.008	0.008	0.008	0.008	0.008	0.008
		High	0.008	0.008	0.008	0.008	0.008	0.008
EMR-D	Male	Background	0.984	0.984	0.984	0.984	0.984	0.984
		Low	0.008	0.008	0.008	0.008	0.008	0.008
		High	0.008	0.008	0.008	0.008	0.008	0.008
	Female	Background	0.995	0.995	0.995	0.995	0.995	0.995
		Low	0.003	0.003	0.003	0.003	0.003	0.003
		High	0.003	0.003	0.003	0.003	0.003	0.003

Table 21.15 Proportions of the population exposed to leukaemogens by subregion, age, sex and level of exposure (*continued*)

			Age group (years)					
Subregion	*Sex*	*Exposure level*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
EUR-A	Male	Background	0.968	0.968	0.968	0.968	0.968	0.968
		Low	0.029	0.029	0.029	0.029	0.029	0.029
		High	0.003	0.003	0.003	0.003	0.003	0.003
	Female	Background	0.973	0.973	0.973	0.973	0.973	0.973
		Low	0.024	0.024	0.024	0.024	0.024	0.024
		High	0.003	0.003	0.003	0.003	0.003	0.003
EUR-B	Male	Background	0.978	0.978	0.978	0.978	0.978	0.978
		Low	0.011	0.011	0.011	0.011	0.011	0.011
		High	0.011	0.011	0.011	0.011	0.011	0.011
	Female	Background	0.983	0.983	0.983	0.983	0.983	0.983
		Low	0.009	0.009	0.009	0.009	0.009	0.009
		High	0.009	0.009	0.009	0.009	0.009	0.009
EUR-C	Male	Background	0.982	0.982	0.982	0.982	0.982	0.982
		Low	0.009	0.009	0.009	0.009	0.009	0.009
		High	0.009	0.009	0.009	0.009	0.009	0.009
	Female	Background	0.981	0.981	0.981	0.981	0.981	0.981
		Low	0.009	0.009	0.009	0.009	0.009	0.009
		High	0.009	0.009	0.009	0.009	0.009	0.009
SEAR-B	Male	Background	0.980	0.980	0.980	0.980	0.980	0.980
		Low	0.010	0.010	0.010	0.010	0.010	0.010
		High	0.010	0.010	0.010	0.010	0.010	0.010
	Female	Background	0.985	0.985	0.985	0.985	0.985	0.985
		Low	0.008	0.008	0.008	0.008	0.008	0.008
		High	0.008	0.008	0.008	0.008	0.008	0.008
SEAR-D	Male	Background	0.979	0.979	0.979	0.979	0.979	0.979
		Low	0.011	0.011	0.011	0.011	0.011	0.011
		High	0.011	0.011	0.011	0.011	0.011	0.011
	Female	Background	0.995	0.995	0.995	0.995	0.995	0.995
		Low	0.003	0.003	0.003	0.003	0.003	0.003
		High	0.003	0.003	0.003	0.003	0.003	0.003
WPR-A	Male	Background	0.975	0.975	0.975	0.975	0.975	0.975
		Low	0.023	0.023	0.023	0.023	0.023	0.023
		High	0.003	0.003	0.003	0.003	0.003	0.003
	Female	Background	0.979	0.979	0.979	0.979	0.979	0.979
		Low	0.019	0.019	0.019	0.019	0.019	0.019
		High	0.002	0.002	0.002	0.002	0.002	0.002
WPR-B	Male	Background	0.978	0.978	0.978	0.978	0.978	0.978
		Low	0.011	0.011	0.011	0.011	0.011	0.011
		High	0.011	0.011	0.011	0.011	0.011	0.011
	Female	Background	0.980	0.980	0.980	0.980	0.980	0.980
		Low	0.010	0.010	0.010	0.010	0.010	0.010
		High	0.010	0.010	0.010	0.010	0.010	0.010

Table 21.16 Proportions of the population exposed to asbestos by subregion, age, sex and level of exposure

			Age group (years)					
Subregion	*Sex*	*Exposure level*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	Background	0.961	0.961	0.961	0.961	0.961	0.961
		Low	0.019	0.019	0.019	0.019	0.019	0.019
		High	0.019	0.019	0.019	0.019	0.019	0.019
	Female	Background	0.978	0.978	0.978	0.978	0.978	0.978
		Low	0.011	0.011	0.011	0.011	0.011	0.011
		High	0.011	0.011	0.011	0.011	0.011	0.011
AFR-E	Male	Background	0.961	0.961	0.961	0.961	0.961	0.961
		Low	0.020	0.020	0.020	0.020	0.020	0.020
		High	0.020	0.020	0.020	0.020	0.020	0.020
	Female	Background	0.975	0.975	0.975	0.975	0.975	0.975
		Low	0.012	0.012	0.012	0.012	0.012	0.012
		High	0.012	0.012	0.012	0.012	0.012	0.012
AMR-A	Male	Background	0.973	0.973	0.973	0.973	0.973	0.973
		Low	0.025	0.025	0.025	0.025	0.025	0.025
		High	0.003	0.003	0.003	0.003	0.003	0.003
	Female	Background	0.991	0.991	0.991	0.991	0.991	0.991
		Low	0.008	0.008	0.008	0.008	0.008	0.008
		High	0.001	0.001	0.001	0.001	0.001	0.001
AMR-B	Male	Background	0.966	0.966	0.966	0.966	0.966	0.966
		Low	0.017	0.017	0.017	0.017	0.017	0.017
		High	0.017	0.017	0.017	0.017	0.017	0.017
	Female	Background	0.992	0.992	0.992	0.992	0.992	0.992
		Low	0.004	0.004	0.004	0.004	0.004	0.004
		High	0.004	0.004	0.004	0.004	0.004	0.004
AMR-D	Male	Background	0.965	0.965	0.965	0.965	0.965	0.965
		Low	0.017	0.017	0.017	0.017	0.017	0.017
		High	0.017	0.017	0.017	0.017	0.017	0.017
	Female	Background	0.994	0.994	0.994	0.994	0.994	0.994
		Low	0.003	0.003	0.003	0.003	0.003	0.003
		High	0.003	0.003	0.003	0.003	0.003	0.003
EMR-B	Male	Background	0.963	0.963	0.963	0.963	0.963	0.963
		Low	0.018	0.018	0.018	0.018	0.018	0.018
		High	0.018	0.018	0.018	0.018	0.018	0.018
	Female	Background	0.994	0.994	0.994	0.994	0.994	0.994
		Low	0.003	0.003	0.003	0.003	0.003	0.003
		High	0.003	0.003	0.003	0.003	0.003	0.003
EMR-D	Male	Background	0.962	0.962	0.962	0.962	0.962	0.962
		Low	0.019	0.019	0.019	0.019	0.019	0.019
		High	0.019	0.019	0.019	0.019	0.019	0.019
	Female	Background	0.985	0.985	0.985	0.985	0.985	0.985
		Low	0.008	0.008	0.008	0.008	0.008	0.008
		High	0.008	0.008	0.008	0.008	0.008	0.008

Table 21.16 Proportions of the population exposed to asbestos by subregion, age, sex and level of exposure (*continued*)

			Age group (years)					
Subregion	*Sex*	*Exposure level*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
EUR-A	Male	Background	0.971	0.971	0.971	0.971	0.971	0.971
		Low	0.026	0.026	0.026	0.026	0.026	0.026
		High	0.003	0.003	0.003	0.003	0.003	0.003
	Female	Background	0.991	0.991	0.991	0.991	0.991	0.991
		Low	0.008	0.008	0.008	0.008	0.008	0.008
		High	0.001	0.001	0.001	0.001	0.001	0.001
EUR-B	Male	Background	0.962	0.962	0.962	0.962	0.962	0.962
		Low	0.019	0.019	0.019	0.019	0.019	0.019
		High	0.019	0.019	0.019	0.019	0.019	0.019
	Female	Background	0.982	0.982	0.982	0.982	0.982	0.982
		Low	0.009	0.009	0.009	0.009	0.009	0.009
		High	0.009	0.009	0.009	0.009	0.009	0.009
EUR-C	Male	Background	0.949	0.949	0.949	0.949	0.949	0.949
		Low	0.025	0.025	0.025	0.025	0.025	0.025
		High	0.025	0.025	0.025	0.025	0.025	0.025
	Female	Background	0.975	0.975	0.975	0.975	0.975	0.975
		Low	0.013	0.013	0.013	0.013	0.013	0.013
		High	0.013	0.013	0.013	0.013	0.013	0.013
SEAR-B	Male	Background	0.959	0.959	0.959	0.959	0.959	0.959
		Low	0.020	0.020	0.020	0.020	0.020	0.020
		High	0.020	0.020	0.020	0.020	0.020	0.020
	Female	Background	0.981	0.981	0.981	0.981	0.981	0.981
		Low	0.010	0.010	0.010	0.010	0.010	0.010
		High	0.010	0.010	0.010	0.010	0.010	0.010
SEAR-D	Male	Background	0.959	0.959	0.959	0.959	0.959	0.959
		Low	0.021	0.021	0.021	0.021	0.021	0.021
		High	0.021	0.021	0.021	0.021	0.021	0.021
	Female	Background	0.978	0.978	0.978	0.978	0.978	0.978
		Low	0.011	0.011	0.011	0.011	0.011	0.011
		High	0.011	0.011	0.011	0.011	0.011	0.011
WPR-A	Male	Background	0.967	0.967	0.967	0.967	0.967	0.967
		Low	0.030	0.030	0.030	0.030	0.030	0.030
		High	0.003	0.003	0.003	0.003	0.003	0.003
	Female	Background	0.988	0.988	0.988	0.988	0.988	0.988
		Low	0.011	0.011	0.011	0.011	0.011	0.011
		High	0.001	0.001	0.001	0.001	0.001	0.001
WPR-B	Male	Background	0.955	0.955	0.955	0.955	0.955	0.955
		Low	0.022	0.022	0.022	0.022	0.022	0.022
		High	0.022	0.022	0.022	0.022	0.022	0.022
	Female	Background	0.974	0.974	0.974	0.974	0.974	0.974
		Low	0.013	0.013	0.013	0.013	0.013	0.013
		High	0.013	0.013	0.013	0.013	0.013	0.013

Table 21.17 Examples of sources used to assess the risk factor–disease relationship for selected occupational carcinogens

Selected risk factor	*Health outcome*	*Examples of key sources of evidence of causality*
Lung carcinogens	Cancer of the trachea, bronchus or lung	Nurminen and Karjalainen (2001); Steenland et al. (1996, 2003)
Leukaemogens	Leukaemia	Lynge et al. (1997); BEIR V (1990); IARC (1997)

- that relative risks are the same for men and women;
- that relative risk values are constant with age; and
- that the relative risks apply equally to the risk of developing the malignant condition (incident cases) and to the risk of dying from the condition (fatal cases); where relative risk values were based on disease incidence studies, the incidence ratio was comparable to the corresponding mortality risk ratio.

Steenland et al. (1996) estimated for the United States the relative risk of exposure to nine lung carcinogens (arsenic, asbestos, beryllium, cadmium, chromium, diesel fumes, nickel, silica and radon). They did not consider agents to which relatively few workers were exposed (BCME, coke oven and coal gasification fumes and soot) and they did not consider smoking, beyond the selection where possible of relative risk factors that had been adjusted for smoking. Combined relative risk values (ranging from 1.31 to 3.69) were calculated for all but radon, using inverse variance and a random-effects model and relying on major cohort studies of the specific agents. The authors estimated that 9000–10 000 men and 900–1900 women develop lung cancer annually in the United States owing to past exposure to occupational carcinogens (except radon). This would account for approximately 9% of lung cancer deaths in males and 2% in females, or 0.5% of all deaths annually in the United States.

Steenland et al. (2003) examined the population-attributable risk (PAR) from several studies (including Steenland et al. 1996). They applied the PAR to deaths occurring in 1997 in the United States to determine occupational deaths from lung cancer, among other outcomes. The authors determined a PAR for lung cancer in the range 6.1–17.3% for men and 2% for women. For overall cancer they determined a PAR of 7–19% for men and 11% for women. For leukaemia, a combined PAR for men and women of 0.8–2.8% was calculated.

Nurminen and Karjalainen (2001) estimated the proportion of fatalities related to occupational factors in Finland. The average number of exposed workers in Finland was estimated from census data by sex, age, occupation and industry, and the FINJEM national job-exposure matrix.

Relative risks were obtained from a review of epidemiological studies, focusing on risk estimates that were most valid for the Finnish exposure circumstances. The attributable fraction methodology was used to determine the proportion of deaths in the population attributable to occupational factors. The authors reported that 30% of deaths due to occupational disease in Finland in 1996 were caused by cancer. Occupational lung cancer accounted for 0.9% of all deaths. They attributed 24% of cancer of the bronchus and lung (29% for men and 5.3% for women) to occupational exposure to combined risk factors. The attributable fractions for urinary cancer were 10.3% overall—14.2% for men and 0.7% for women. Combined occupational risk factors resulted in 10.9% (18.5% for males, 2.5% for females) of leukaemia deaths being attributed to occupational exposures, the majority (17.8% and 2.3%, respectively) from electrical occupations, in contrast to 0.7% and 0.2%, respectively, from benzene. An average of 71.3% (90% for males, 25% for females) of malignant mesothelioma was attributed to occupation.

The three review papers described above (Nurminen and Karjalainen 2001; Steenland et al. 1996, 2003) provided summary measures, or information that can be used to determine summary measures, of relative risk for one or more of the main agents and outcomes of interest. The study by Nurminen and Karjalainen (2001) focused on Finland, and preferentially used studies based in Scandinavia or thought to be most relevant to Finland. Most of its relative risk estimates relate to lung cancer, with attributable fractions presented for leukaemia. The paper by Steenland et al. (1996), although focused on the United States, was more inclusive of studies of suitable quality. The other paper by Steenland et al. (2003) provided information on relative mortality risks similar to the first (1996) paper. All papers provided similar summary measures of relative risk for lung cancers, but the Steenland et al. (1996) results were used preferentially because they are generally based on a broader range of studies. However, the Steenland paper provided information only on lung cancer. Table 21.18 gives a summary of the risk measures for each of the carcinogens and the relevant outcomes. The basis for these risk estimations is described in more detail below.

Lung cancer

The evidence for substance-specific relative risk values, which were used to calculate the overall relative risk for the eight lung carcinogens, is briefly discussed below, relying heavily on the review paper by Steenland et al. (1996). The data in the paper provide a summary relative risk of 1.6 for occupational exposure to the set of lung carcinogens considered here.

Smoking is the main important potential confounder of lung cancer, and potentiates the effect of some exposures (notably with asbestos and lung cancer). In this analysis, where possible, studies were used that produced risk estimates for lung cancer after controlling for smoking.

Table 21.18 Summary of risk measures (relative risk and mortality rates) for occupational carcinogens

Health outcome	*Risk measure*	*Estimate and 95% confidence interval (CI)*	*Comments*	*Primary data sources*
Cancer of the trachea, bronchus and lung	Relative risk	Low exposure: 1.22 (1.09–1.35) to 1.32 (1.17–1.48) High exposure: 1.79 (1.59–1.97) to 1.93 (1.71–2.16)	Composite relative risk based on individual relative risk of arsenic, asbestos, beryllium, cadmium, chromium, diesel exhaust, nickel and silica	Steenland et al. (1996)
Leukaemia	Relative risk	Low exposure: 1.67 (1.51–2.00) to 1.93 (1.76–2.17) High exposure: 3.06 (2.64–3.80) to 3.86 (3.48–4.32)	Composite relative risk based on individual relative risk of benzene, ionizing radiation and ethylene oxide	BEIR V (1990); IARC (2000); Lynge et al. (1997); Steenland et al. (2003)

Arsenic

Arsenic is accepted as a Group 1 carcinogen (IARC 1980, 1987a). The six principal epidemiological studies (covering nearly 18 000 workers) reviewed by Steenland et al. (1996) indicated a combined relative risk of 3.69, with a range of 1.31–15.2 reported for individual studies and a clear dose–response relationship. The lowest relative risk arose from a study in which exposures mostly ranged from 7 to 13 $\mu g/m^3$, compared to the OSHA level of 10 $\mu g/m^3$ (Enterline et al. 1987). Excess cancers in other studies were probably due to high exposures that occurred largely in the past. A combined relative risk of 3.69 (95% CI 3.06–4.46) was determined by the Steenland et al. (1996) review, whereas 3.2 was used by Nurminen and Karjalainen (2001).

Asbestos

Both serpentine and amphibole asbestos have been shown to cause lung cancer in humans, with a clear dose–response relationship and a synergy between asbestos and tobacco (Lee 2001). Over 100 cohort studies and many case-referent studies, plus animal and cellular studies, provide ample evidence for causation. In six cohort studies of nearly 6000 asbestosis patients, the standardized mortality rate ranged from 3.5 to 9.1, with a combined relative risk of 5.9. In 20 studies of over 100 000 asbestos workers, the standardized mortality rate ranged from 1.04 for chrysotile workers to 4.97 for amosite workers, with a combined relative risk of 2.00. It is difficult to determine the exposures involved because few of the studies reported measurements, and because it is a problem to convert historical asbestos measurements in millions of dust particles per cubic foot to gravimetric units. Nevertheless, little excess lung cancer is expected from low exposure levels. These studies have been the subject of several reviews (IARC 1977; IPCS 1998; Nurminen and Karjalainen 2001; Steenland et al. 1996). The main papers provided a range of relative risks (1.04–7.4), with summary relative risks of 2.0 (Steenland et al. 1996) and 2.3 (Nurminen and Karjalainen 2001) cited in the two most recent reviews. The lower value (2.0, 95% CI 1.90–2.11), which is based on a wider range of studies, is accepted for this analysis.

Beryllium

Beryllium is an IARC Group 1 carcinogen (IARC 1993), although epidemiological evidence is rather limited. A standardized mortality rate for lung cancer of 2.0 was determined from a registry cohort of 689 women and men (Steenland and Ward 1991), and an overall standardized mortality rate of 1.24 was found in a study of 9225 male workers from seven beryllium plants (1.49 at plants with higher exposure) (Ward et al. 1992). Steenland et al. (1996) utilized a smoking-adjusted relative risk of 1.49 (no 95% CI reported), based on a beryllium plant with high exposures.

Cadmium

Cadmium is an IARC Group 1 carcinogen (IARC 1993). The best epidemiological evidence of its relationship to lung cancer comes from a cohort study by Stayner et al. (1992), although the evidence for carcinogenicity is stronger in animals and has recently been questioned in humans (Jarup and Nordberg 1998). The most recent follow-up study suggests a relative risk of 1.49 (95% CI 0.96–2.22) (Steenland et al. 1996). Nurminen and Karjalainen (2001) used 1.2, based on a Scandinavian study.

Chromium

Chromium is an IARC Group 1 carcinogen (IARC 1990a). There is ample epidemiological evidence of its causal association with lung cancer, with many cohort studies showing a dose–response relationship. Based on the largest and best designed studies of chromium production workers, producers of chromate paints and chromate plating workers, the overall relative risk is 2.78 (95% CI 2.47–3.52) (Steenland et al. 1996). Nurminen and Karjalainen (2001) used a lower relative risk of 1.4 from a hospital-based case-referent study.

Diesel exhaust

Polycyclic aromatic hydrocarbons comprise the main components of diesel exhaust, which contains a mixture of substances. Diesel exhaust has been accepted as a Group 2A carcinogen (IARC 1989) and was scheduled for further review in 2001. Owing to limitations in exposure assessment to diesel exhaust, human epidemiology has been difficult to conduct. However, cohort studies and meta-analyses confirm a relationship between diesel exhaust exposure and lung cancer, with summary relative risks in the range 1.3–1.5 (Bhatia et al. 1998; Lipsett and Campleman 1999). Based on six relatively consistent recent studies with good documentation of exposure to diesel exhaust, in which the number of cases ranged from 50 to 1256, Steenland et al. (1996) determined a combined relative risk of 1.31 (95% CI 1.13–1.44), and Nurminen and Karjalainen (2001) used the same estimate.

Nickel

Nickel is an IARC Group 1 carcinogen (IARC 1990a). Based on data from the 1990 report of the International Committee on Nickel Carcinogenesis in Man (ICNCM 1990), Steenland et al. (1996) calculated a combined relative risk of 1.56 (95% CI 1.41–1.73). Nurminen and Karjalainen (2001) used an estimate of 1.4 based on a Finnish study.

Silica

On the basis of detailed reviews, silica has been classified as an IARC Group 1 carcinogen (IARC 1987b, 1997). Several cohort studies in silica-

exposed and silicosis patients showed a dose–response relationship between silica exposure and lung cancer relative risk, and this was confirmed by meta-analyses and a pooled study (Steenland and Sanderson 2001). Animal and cellular studies provided supporting evidence. Controversy remains as to whether silicosis is a necessary precursor for the development of lung cancer, but this does not affect the underlying status of silica as a carcinogen (Checkoway 2000; Hnizdo and Sluis-Cremer 1991; Soutar et al. 2000). Steenland et al. (1996) based their combined relative risk of 1.33 (95% CI 1.21–1.45) on 13 large cohort and case–control studies of silica-exposed workers. These studies included granite workers, stone workers, pottery workers, brick workers, gold miners and diatomaceous earth miners, and covered a range of workers generally numbering from almost 1000 to over 5000. Half of the studies controlled for smoking. Nurminen and Karjalainen (2001) used a slightly higher estimate of 1.4.

Combined estimates

A common methodology, similar to that used by Steenland et al. (1996) and Nurminen and Karjalainen (2001), was used in this analysis for all lung carcinogens, in that occupational exposure to carcinogens was estimated and applied to relative risk estimates to enable the determination of attributable fractions. A mean relative risk of 1.63 was determined for eight lung carcinogens (not including radon), using data reported by Steenland et al. (1996). This was done by calculating a weighted average of the substance-specific relative risks, and weighting the substance-specific relative risks by the proportion of workers exposed to each substance to determine a mean relative risk for workers exposed to the eight lung carcinogens. This was done separately for each subregion, using the proportion of workers in each subregion exposed to specific agents to weight the relative risk for each of the agents. However, the resulting average relative risks were not clearly different from each other (all were close to 1.6).

In addition, to estimate an uncertainty range for the initial mean relative risk, a weighted average was calculated of the lower and upper 95% CI values of the relative risk reported for each substance (except beryllium, for which there were no estimated CI). These values (not to be confused with the partitioned relative risk values for low- and high-level exposure) were within 15% of the mean relative risk values. This is demonstrated for the AMR-A subregion (Table 21.19).

To produce relative risk estimates for low and high exposure, it was necessary to partition the mean relative risks into values that correspond to low- and high-level exposure. A mean relative risk (of 1.6) was determined for the United States. Based on the estimates of 90% of American workers exposed at or below about one fifth of the PEL values and 10% exposed at or above the PEL values, and an estimate of the American population-attributable fraction of lung cancer due to occupation

Table 21.19 Lung cancer relative risk, substance-specific and weighted average, for the AMR-A subregion

Carcinogen	*Combined relative risk*[a] *(95% CI)*	*Proportion of workers exposed*
Silica	1.33 (1.21–1.45)	0.0248
Cadmium	1.49 (0.96–2.22)	0.0015
Nickel	1.56 (1.41–1.73)	0.0039
Arsenic	3.69 (3.06–4.46)	0.0011
Chromium	2.78 (2.47–3.52)	0.0055
Diesel fumes	1.31 (1.13–1.44)	0.0217
Beryllium	1.49	0.0005
Asbestos	2.00 (1.90–2.11)	0.0094
Total[b]	1.59 (1.41–1.77)	

[a] Derived from major epidemiological studies.

[b] Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen.

Source: Steenland et al. (1996).

of 9% (Steenland et al. 1996), the mean relative risk of 1.6 was partitioned into a relative risk of 1.3 for low-level exposure to lung carcinogens, and 1.9 for high-level exposure. The United States ratios of the lower (1.3/1.6) and the higher (1.9/1.6) relative risks to the average relative risk were then applied to the average relative risks estimated for each subregion to produce estimated relative risks at low and high exposures for each subregion. In the same manner, upper and lower 95% CI were produced for these relative risks, based on the limits estimated for the average relative risks. The results of this process are shown in Table 21.20.

Leukaemia

Leukaemia has been linked to exposure to benzene, ionizing radiation and ethylene oxide, all of which are IARC Group 1 carcinogens (IARC 2001; WHO 1999). There is also some evidence that exposure to low-frequency electric fields may be leukaemogenic (Nurminen and Karjalainen 2001; WHO 2001). However, as this physical agent has not been included in CAREX, it has been excluded from this study.

Benzene

The causal relationship between leukaemia and benzene is well recognized, including data from cohort studies in China and the United States covering workers in chemical plants, refineries, machine production, and textile and cloth factories. Excesses of nonlymphocytic, myelogenous and acute myeloid leukaemias occurred. There is also limited evidence in

Table 21.20 Weighted summary relative risks for lung cancer for all subregions

Subregion	Summary relative risk[a]	Low exposure: Combined relative risk[a] (95% CI)	High exposure: Combined relative risk[a] (95% CI)
AFR-D	1.61	1.31 (1.17–1.45)	1.91 (1.72–2.11)
AFR-E	1.62	1.32 (1.18–1.45)	1.92 (1.72–2.12)
AMR-A	1.59	1.29 (1.14–1.44)	1.88 (1.67–2.11)
AMR-B	1.58	1.28 (1.14–1.42)	1.87 (1.67–2.08)
AMR-D	1.56	1.26 (1.13–1.41)	1.85 (1.64–2.05)
EMR-B	1.56	1.26 (1.13–1.40)	1.85 (1.64–2.05)
EMR-D	1.61	1.31 (1.18–1.45)	1.92 (1.72–2.10)
EUR-A	1.62	1.32 (1.17–1.48)	1.93 (1.71–2.16)
EUR-B	1.59	1.29 (1.15–1.44)	1.89 (1.69–2.10)
EUR-C	1.50	1.22 (1.09–1.35)	1.79 (1.59–1.97)
SEAR-B	1.58	1.28 (1.15–1.42)	1.88 (1.68–2.07)
SEAR-D	1.61	1.31 (1.17–1.45)	1.91 (1.70–2.09)
WPR-A	1.57	1.27 (1.13–1.42)	1.86 (1.65–2.08)
WPR-B	1.58	1.28 (1.14–1.42)	1.87 (1.67–2.07)

[a] Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen in each subregion.

mammals (Hayes et al. 1997; IARC 1990b). A recent review (Lynge et al. 1997) provides a low-exposure relative risk of 2.0 (95% CI 1.8–2.2) and a high-exposure relative risk of 4.0 (3.6–4.4).

Ionizing radiation

The causal relationship between ionizing radiation and leukaemia is well recognized. There is consistency across numerous studies, strong association between exposure and outcome, and evidence of a dose–response gradient. Excess leukaemia has been observed in survivors of the atomic explosions at Hiroshima and Nagasaki, and also among patients medically treated with X-rays or γ-rays. The risk of leukaemia increases over fivefold at sufficiently high doses (BEIR V 1990; IARC 2000; ICRP 1991). Models describing risk have been proposed as: Linear RR model $1+5.5$ _ dose in Sv, quadratic $RR=1+0.24$ dose $+0.27$ dose^2 (dose in Sv) (BEIR V 1990). Relative risks of 1.22 (1.07–1.70) for low exposure and 1.57 (1.18–2.88) for high exposure are accepted as the best available estimates (BEIR V 1990; IARC 2000).

Ethylene oxide

Workers have exposure to ethylene oxide either as a sterilant or as a chemical intermediary or final product. In a study in the United States,

ethylene oxide used as a sterilant was associated with lymphatic leukaemia and non-Hodgkin's lymphoma, with a rate ratio of 1.2 estimated for 45-year exposure to 1 ppm. Other studies in Sweden and the United Kingdom of Great Britain and Northern Ireland showed nonsignificant excesses of these cancers. Of six studies of chemical plant workers (two in Sweden and one each in Germany, Italy, the United Kingdom and the United States), two found significant excesses, two found nonsignificant excesses and two found expected rates (IARC 1997). Relative risk was found to range from 1.1 to 3.5 (Steenland et al. 2003).

An approach similar to that used for lung carcinogens was applied to the leukaemogens. The separate relative risks for the development of leukaemia arising from exposures to the main relevant occupational carcinogens were combined into single summary relative risks, one for low exposure and one for high exposure. This was done separately for each subregion, using the exposure prevalence of the workforce in each subregion to weight the exposure-specific risks. However, the resulting average relative risks were not clearly different from each other. CI were estimated in the same manner, weighting the estimated CI for benzene and ionizing radiation (there were no estimated CI for ethylene oxide). Unlike lung cancer, the low- and high-exposure relative risks were available for each exposure, and these were directly incorporated into low- and high-exposure summary measures through the weighting process. An example of this approach is shown in Table 21.21, using the WPR-B subregion. The results of this approach for each subregion are shown in Table 21.22.

Table 21.21 Leukaemia relative risk, substance-specific and weighted average, for the WPR-B subregion

	Low exposure	*High exposure*	
Carcinogen	*Combined relative risk*[a] *(95% CI)*	*Combined relative risk*[a] *(95% CI)*	*Proportion of workers exposed*
Benzene	2.0 (1.8–2.2)	4.0 (3.6–4.4)	0.0051
Ionizing radiation	1.22 (1.07–1.7)	1.57 (1.18–2.88)	0.0010
Ethylene oxide	1.1	3.5	0.0003
Total[b]	1.84 (1.68–2.12)	3.60 (3.20–4.16)	

[a] Derived from major epidemiological studies.

[b] Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen.

Table 21.22 Weighted summary relative risks for leukaemia for all subregions

	Low exposure	*High exposure*
Subregion	*Combined relative risk[a] (95% CI)*	*Combined relative risk[a] (95% CI)*
AFR-D	1.88 (1.72–2.15)	3.73 (3.34–4.24)
AFR-E	1.89 (1.73–2.15)	3.75 (3.37–4.25)
AMR-A	1.91 (1.74–2.16)	3.80 (3.41–4.28)
AMR-B	1.77 (1.61–2.07)	3.38 (2.98–4.01)
AMR-D	1.91 (1.74–2.16)	3.78 (3.39–4.27)
EMR-B	1.87 (1.70–2.13)	3.66 (3.26–4.19)
EMR-D	1.89 (1.72–2.15)	3.73 (3.33–4.23)
EUR-A	1.93 (1.76–2.17)	3.86 (3.48–4.32)
EUR-B	1.83 (1.67–2.11)	3.57 (3.18–4.13)
EUR-C	1.67 (1.51–2.00)	3.06 (2.64–3.80)
SEAR-B	1.89 (1.73–2.15)	3.76 (3.37–4.26)
SEAR-D	1.81 (1.65–2.10)	3.51 (3.11–4.09)
WPR-A	1.90 (1.73–2.15)	3.77 (3.38–4.26)
WPR-B	1.84 (1.68–2.12)	3.60 (3.21–4.16)

[a] Weighted summary relative risk, weighted by the proportion of workers exposed to each contributing carcinogen in each subregion.

Estimates of risk reversibility

There are limited data on risk reversibility from occupational exposure to carcinogens. The studies from which the estimated risks arise are based on cohorts of people exposed for different periods of time, followed up for various periods of time and with various periods of time between exposure cessation and follow-up, with follow-up periods varying between zero (still exposed) and many decades. Therefore, most of the absolute and relative risks produced by the studies already depend on whatever change in risk might occur once exposure ceases. However, some indication of the extent of risk reduction that might occur is given by a recent paper by Peto et al. (2000), which examined changes in the risk of developing lung cancer as a result of stopping smoking. The study estimated that, compared to the risk in persons who continued to smoke, the risk of lung cancer in males declined to about 0.66 within 10 years, to 0.44 between 10 and 20 years, to 0.2 between 20 and 30 years, and to 0.1 after 30 years.

3. Occupational airborne particulates

There are a vast number of respiratory conditions that can arise directly or indirectly from work. However, estimating exposures, risks and attrib-

utable proportions is not possible for many of these on an international (or even national) scale, because of lack of appropriate data sources. Therefore, only the more important of the work-related respiratory conditions, in terms of the total number of cases or the risks arising from exposure, are included here. All of these arise from exposure to particulates. Malignant respiratory disease is not included here because it is described in section 2.

Nonmalignant respiratory disease arises as a result of the exposure of workers to airborne agents, mostly in the form of particulates or dusts.[2] The primary route of exposure is inhalation, whereby these agents gain access to the respiratory system and are either deposited (in the case of dusts) or enter the circulatory system. For some exposures, there is a very clear connection between the exposure and the disease (for example, silicosis is only caused by exposure to silica). Some exposures cause more than one type of disease, and even more than one type of respiratory disease. For example, asbestos can result in malignant conditions of the lung and the pleura (the inside lining of the chest), malignant conditions of the peritoneum (the inside lining of the abdomen) and nonmalignant conditions of the lungs (asbestosis and COPD). Other exposures have not been well characterized, but are believed to result in certain conditions (such as some forms of occupational asthma).

3.1 Exposure variable

Causative agents of asthma

Asthma, which is a narrowing of the upper respiratory passages resulting in difficult breathing and wheezing, has both nonoccupational and occupational causes. Many hundreds of occupational agents, including some inorganic and organic dusts, have been associated with occupational asthma (Balmes et al. 2003; Chan-Yeung and Malo 1994; Venables and Chan-Yeung 1997). Biological agents include grains, flours, plants and gums, fur, feathers and other animal parts, insects and fungi, drugs and enzymes and various types of wood. Chemical agents include chlorofluorocarbons, alcohols, metals and their salts, and welding fumes (CCOHS 1997). These agents are found in a variety of workplaces, including food and natural products processing, animal handling facilities, manufacturing and construction.

It would not be possible to conduct exposure assessments and to obtain relative risk data for all the factors contributing to this important occupational disease, especially since they often occur in combination. We therefore used occupation as a proxy for exposure to agents that are associated with occupational asthma. The basis for this approach was the work of Karjalainen et al. (2001, 2002), who conducted extensive epidemiological studies of the entire Finnish workforce and developed relative risks for specific occupations. A similar but less extensive study based in 12 industrialized countries was also used (Kogevinas et al.

1999). Relative risks were applied to these occupational data to produce estimates of the number of deaths due to work-related asthma.

Causative agents of COPD

The causative agents of COPD are non-specific dust and fumes, with dusts showing a more consistent relationship than fumes (Becklake 1989). Because of a lack of worldwide data on the prevalence of occupational exposure to dusts and their combinations, work in specific economic subsectors was used as a surrogate for dust exposure. Relative risks were applied to these workforce data to produce estimates of the number of deaths from COPD arising from work-related exposures.

3.2 Estimating risk factor levels

The general exposure assessment methodology was described earlier. Occupation was used for asthma (Equation 2) and economic sector for COPD (Equation 1). The theoretical minimum risk corresponds to no occupational exposure above background levels to airborne particulates or other agents that cause nonmalignant respiratory disease.

Agents causing asthma

The proportion of the total population with occupational exposure to asthmagens was estimated using Equation 2. Estimates were made for each occupational category by determining the proportion of the population working in occupations that matched as closely as possible to those identified by Karjalainen et al. (2001, 2002) and for which relative risk values were provided (Table 21.23). Those not working and those employed in administration were together considered to be the nonexposed reference category (relative risk = 1). These calculations were done separately for men and women for each subregion of the world. Relative risks and the proportions exposed by occupational category were applied across all age groups from age 15 to ≥80 years.

Table 21.24 summarizes the age-adjusted distribution of the labour force into occupations matching the categories for which relative risks were identified by Karjalainen et al. (2002).

Agents causing COPD

It is not possible to estimate the proportion of the world's population exposed to the large number of agents identified in occupation-specific and agent-specific studies. Community-based studies have therefore been preferred. The most common exposure in these studies is exposure to dust and/or gas/fumes (e.g. Korn et al. 1987; Kryzanowski et al. 1986; Xu et al. 1992). Unfortunately, there are also no data to estimate the proportion of the world's workers exposed to dust and/or gas/fumes. The study by Korn et al. (1987) provides a link between self-reported exposure to dust (current and past exposure) and some categories of economic activity[3] among the currently employed. Categories of economic activity

Table 21.23 Comparison of Finnish occupational categories with 1968 ISIC codes

Finnish classification[a]	*Description*	*Examples*[a,b]	*1968 ISIC*[c]	*Description*	*Examples*	*Comments*
1	Administrative, managerial and clerical workers	No examples given	2	Administrative and managerial workers	Government officials, managers (upper level)	Combine Categories 2 and 3 (ISIC) for all economic subsectors
			3	Clerical and related workers	Office managers, government workers, secretaries, bookkeepers, stock clerks	
0	Technical, physical science, social science, humanistic and artistic workers	Engineers, medical personnel, child care givers, religious and social workers	0/1	Professional, technical and related workers	Scientists, technicians, engineers, medical and related workers, mathematicians, teachers, religious workers, artists	Use Category 0/1 (ISIC) for all economic subsectors
2	Sales workers	Wholesale and retail dealers, other sales workers	4	Sales workers	Working proprietors, sales managers, sales workers, insurance agents	Use Category 4 (ISIC) for all economic subsectors
3	Agriculture, forestry, commercial fishing	Farmers and managerial workers in agriculture, forestry and horticulture, agricultural and horticultural workers, animal husbandry workers	6	Agricultural, animal husbandry and forestry workers, fishermen, hunters	Farm managers and supervisors, agriculture and animal husbandry workers, forestry workers, fishermen	Use Category 6 (ISIC) for all economic subsectors
4	Mine and quarry workers	Miners, quarrymen				Use Category 7/8/9 for mining economic subsector

5	Transport and communications workers	Road transport workers and supervisors, transport service workers, postal services and couriers, engine room crews, motor vehicle and tram drivers, railway and station personnel, telephone switchboard operators, newspaper delivery workers, office receptionists, messengers				Use Categories 7/8/9 in ISIC in transport economic subsector (see below); does not include communications workers
6/7	Manufacturing and related workers	Workers in the following industries: textiles; smelting, metallurgical and foundry; iron and metalware; electrical; wood; painting and lacquering; other construction; food and beverage; chemical processing; packing and wrapping; stationary engine and machine; dock and warehouse; other manual	7/8/9	Production and related workers, transport equipment operators and labourers	Production supervisors, miners, quarrymen; workers in metal, wood, chemicals, textiles, food and beverages, tobacco, leather, stone, rubber, paper and construction industries, labourers, machine operators, electricians, material handlers	Use Categories 7/8/9 in all economic subsectors except mining and transport
			7/8/9	Specific transport equipment operators: motor vehicle drivers, bus, truck and tram drivers		

continued

Table 21.23 Comparison of Finnish occupational categories with 1968 ISIC codes (*continued*)

Finnish classification[a]	*Description*	*Examples*[a,b]	*1968 ISIC*[c]	*Description*	*Examples*	*Comments*
8	Service work	Firefighters and police, watch and security guards, cooks, housekeepers, domestic workers, building caretakers and cleaners, hygiene and beauty operators, launderers, dry cleaners and pressers	5	Service workers	Hotel managers, cooks, waiters, housekeepers, caretakers, beauty operators, firefighters, police	Use Category 5 (ISIC) for all economic subsectors

[a] Source: Karjalainen et al. (2001).

[b] Source: Karjalainen et al. (2002).

[c] Source: UN (2000).

Table 21.24 Proportion of the population in occupational categories based on exposure to agents causing asthma, by subregion and sex

		Proportion exposed by occupation								
Subregion	*Sex*	*Background*	*Administration*	*Technical*	*Sales*	*Agricultural*	*Mining*	*Transport*	*Manufacturing*	*Services*
AFR-D	Male	0.1595	0.0498	0.0562	0.0513	0.4612	0.0081	0.0289	0.1342	0.0510
	Female	0.4645	0.0249	0.0303	0.0324	0.3551	0.0012	0.0145	0.0482	0.0288
AFR-E	Male	0.1510	0.0524	0.0618	0.0447	0.4662	0.0082	0.0219	0.1377	0.0563
	Female	0.3498	0.0360	0.0497	0.0328	0.4171	0.0010	0.0114	0.0547	0.0475
AMR-A	Male	0.2746	0.1971	0.1080	0.0875	0.0327	0.0035	0.0196	0.1940	0.0830
	Female	0.4079	0.1772	0.1223	0.0789	0.0134	0.0005	0.0080	0.0928	0.0991
AMR-B	Male	0.1896	0.1124	0.0794	0.0665	0.1592	0.0077	0.0310	0.2225	0.1317
	Female	0.5823	0.0662	0.0671	0.0410	0.0531	0.0013	0.0032	0.0878	0.0979
AMR-D	Male	0.1785	0.1894	0.0346	0.0912	0.0521	0.0020	0.0386	0.3115	0.1021
	Female	0.6110	0.1033	0.0198	0.0466	0.0119	0.0001	0.0026	0.1328	0.0719
EMR-B	Male	0.2134	0.1042	0.1499	0.0854	0.1194	0.0046	0.0409	0.2072	0.0749
	Female	0.6903	0.0523	0.0866	0.0443	0.0293	0.0001	0.0101	0.0463	0.0407
EMR-D	Male	0.1805	0.0419	0.0490	0.1742	0.3584	0.0014	0.0000	0.1465	0.0480
	Female	0.6303	0.0110	0.0105	0.0509	0.2427	0.0002	0.0000	0.0401	0.0142

continued

Table 21.24 Proportion of the population in occupational categories based on exposure to agents causing asthma, by subregion and sex (*continued*)

		Proportion exposed by occupation								
Subregion	*Sex*	*Background*	*Administration*	*Technical*	*Sales*	*Agricultural*	*Mining*	*Transport*	*Manufacturing*	*Services*
EUR-A	Male	0.3227	0.1176	0.2145	0.0252	0.0420	0.0040	0.0000	0.1934	0.0807
	Female	0.5296	0.0942	0.1889	0.0131	0.0254	0.0008	0.0033	0.0733	0.0713
EUR-B	Male	0.2593	0.0747	0.0680	0.0400	0.2133	0.0136	0.0226	0.2490	0.0595
	Female	0.4624	0.0453	0.0552	0.0131	0.2352	0.0020	0.0041	0.1379	0.0449
EUR-C	Male	0.2700	0.0946	0.0532	0.0317	0.1536	0.0212	0.1097	0.2239	0.0421
	Female	0.4269	0.0818	0.0542	0.0512	0.0919	0.0138	0.1041	0.1239	0.0522
SEAR-B	Male	0.1756	0.0599	0.0482	0.0729	0.3664	0.0051	0.0336	0.1930	0.0452
	Female	0.4240	0.0397	0.0368	0.0812	0.2487	0.0013	0.0032	0.1244	0.0407
SEAR-D	Male	0.1502	0.0645	0.0550	0.0150	0.4634	0.0149	0.0392	0.1535	0.0444
	Female	0.5298	0.0134	0.0125	0.0028	0.3781	0.0026	0.0000	0.0509	0.0098
WPR-A	Male	0.2447	0.2058	0.1023	0.0787	0.0336	0.0014	0.0340	0.2270	0.0723
	Female	0.4795	0.1410	0.0832	0.0755	0.0281	0.0002	0.0094	0.1118	0.0713
WPR-B	Male	0.1600	0.1023	0.0655	0.0454	0.3659	0.0194	0.0370	0.1399	0.0645
	Female	0.2901	0.0928	0.0588	0.0753	0.2812	0.0066	0.0290	0.0925	0.0738

Note: See also Table 21.26.

among the currently employed are available worldwide, and can provide a broad approximation to the proportion of the world's population with current or past exposure to dust and/or gas/fumes. We based our estimates of exposed populations on data on employment in economic sectors of agriculture, industry and service from the World Bank (2001), supplemented by data from ILO (2000) on employment in economic activities. The proportions of the population with occupational exposure at medium and high levels to agents causing COPD were estimated using Equation 1.

Korn et al. (1987) defined as low-exposed those in finance, as medium-exposed those in the manufacture of non-durable goods, transport, utilities and the wholesale and retail trades, and as highly exposed those in the manufacture of durable goods, agriculture, mining and construction. Exposure was to "dusts" and to "gases", without these being further defined. We adopted these categories with some modification to account for our lack of data on the type of manufacturing industry and for the fact that agriculture in developed and developing countries probably involves different types of exposure to respirable dust. Lacking data that would have permitted us to divide manufacturing into medium and high potential for dust exposure, we have classified it as having potentially high dust exposure, given that in much of the world manufacturing involves more dust exposure than is typical in the United States where the Korn et al. study (1987) was done (Chien et al. 2002; Gomes et al. 2001). We have defined as nonexposed those not in the workforce and those in utility trade, finance and services. Those in agriculture, manufacturing and transportation were defined as having low exposure, while those in mining and construction were defined as having high exposure. Many workers in the medium and highly exposed economic activities are in fact not exposed to dusts, but on the whole the proportions in these industries are taken to represent the approximate proportion of those ever exposed to low and high levels of dusts in the general population. In Korn et al. (1987), the proportion of workers currently employed in the medium- and high-exposure industries listed above corresponded approximately to the proportion of those reporting ever having been occupationally exposed to dust in that study. This approach was followed in our study, in which it was assumed that the number of currently employed in specific industries corresponds roughly to the number ever occupationally exposed to dusts. The proportion exposed in different economic activities in each subregion was adjusted to account for an average labour force participation among the currently exposed in that subregion, which was applied across all ages. The results are presented in Table 21.25.

Table 21.25 Proportion of the population exposed to agents causing COPD, by subregion, sex and level of exposure

		Proportion ever exposed	
Subregion	*Exposure level*	*Male*	*Female*
	Background	0.3722	0.5920
AFR-D	Low	0.5086	0.3776
	High	0.1192	0.0305
	Background	0.3744	0.5386
AFR-E	Low	0.5051	0.4365
	High	0.1204	0.0249
	Background	0.6879	0.9056
AMR-A	Low	0.0879	0.0314
	High	0.2242	0.0630
	Background	0.5653	0.8908
AMR-B	Low	0.2336	0.0553
	High	0.2011	0.0539
	Background	0.6465	0.9337
AMR-D	Low	0.1253	0.0169
	High	0.2281	0.0494
	Background	0.5829	0.9256
EMR-B	Low	0.2007	0.0441
	High	0.2164	0.0303
	Background	0.5818	0.7780
EMR-D	Low	0.2204	0.1776
	High	0.1978	0.0444
	Background	0.6819	0.8965
EUR-A	Low	0.0565	0.0253
	High	0.2616	0.0781
	Background	0.5096	0.6598
EUR-B	Low	0.2636	0.2469
	High	0.2268	0.0933
	Background	0.4312	0.6463
EUR-C	Low	0.3273	0.2409
	High	0.2415	0.1128
	Background	0.4190	0.6694
SEAR-B	Low	0.4112	0.2384
	High	0.1698	0.0922
	Background	0.3965	0.5723
SEAR-D	Low	0.4822	0.3869
	High	0.1213	0.0408
	Background	0.5994	0.8387
WPR-A	Low	0.1200	0.0531
	High	0.2806	0.1082
	Background	0.3700	0.5244
WPR-B	Low	0.4474	0.3807
	High	0.1826	0.0949

3.3 Risk factor–disease relationships

Asthma

Occupational asthma is a condition characterized by variable airflow limitation or bronchial hyper-responsiveness related to workplace exposure. However, the precise definition of occupational asthma has been widely debated. The most controversial issue concerns whether only immunologically-mediated asthma should be considered to be occupational asthma or whether asthma arising as result of workplace exposure to irritants, or exacerbation of pre-existing asthma by workplace irritants, should also be considered in the definition (Lombardo and Balmes 2000; Malo and Chan-Yeung 2001; Wagner and Wegman 1998). Recently, consensus seems to have been reached in favour of a broad definition (American Thoracic Society review: Balmes et al. 2003). A broader approach has been supported by others (Blanc and Toren 1999; Karjalainen et al. 2001; Kogevinas et al. 1999; Milton et al. 1998; Toren et al. 1999), and recent studies of occupational asthma have tended to use a more inclusive approach (Karjalainen et al. 2001, 2002; Milton et al. 1998).

Occupational asthma is probably the most common work-related respiratory disorder in industrialized countries (Kogevinas et al. 1999), and is either stable (Singh and Davis 2002) or increasing in incidence (Sears 1997). Many hundreds of occupational agents, including some inorganic and organic dusts, have been associated with occupational asthma (Balmes et al. 2003; Chan-Yeung and Malo 1994; Venables and Chan-Yeung 1997).

Until recently, there has been limited information on the total risk of developing asthma from workplace exposure. The United States magnitude of mortality study (Steenland et al. 2003) estimated that about 5% of mortality from nonmalignant work-related respiratory disease was due to asthma. Studies of substance-specific risks have helped to identify or implicate particular substances as likely causative agents (e.g. Monso et al. 1998), but these studies have generally focused on agents thought to be sensitizers, and usually on only a limited number of these. They are therefore not useful for determining the true extent of asthma occurring as a result of work-related exposure. Several population-based studies have partially rectified this problem (Karjalainen et al. 2001, 2002; Kogevinas et al. 1996, 1999; Ng et al. 1994; Toren 1996; Toren et al. 1999), focusing on occupation-specific rather than substance-specific risks because of the plethora of potential causative exposures and the difficulty in characterizing them. These studies provided measures of relative risk and/or population-attributable fractions. Recent studies in Finland have estimated population-attributable fractions for occupational asthma of 18% (Nurminen and Karjalainen 2001) and of 17% (for women) and 29% (for men) (Karjalainen et al. 2002). A comprehensive review undertaken before these two Finnish studies found a

median value for population-attributable fraction of 9% for all relevant studies, and a median value of 15% for the highest-quality studies (Blanc and Toren 1999). The American Thoracic Society (Balmes et al. 2003) has recently reviewed the literature and estimated that approximately 15% of asthma is attributable to occupational exposure, based largely on studies in developed countries.

Of these studies, only that by Karjalainen et al. (2001, 2002) provides useable risk information to cover the whole workforce, while that by Kogevinas et al. provides useful information for agriculture. The study by Karjalainen et al. (2001, 2002) was a longitudinal study over 13 years covering the entire Finnish population, and provided relative risks for a large number of broad occupational categories. In that study, asthma was defined by the occurrence of clinically diagnosed asthma (n=49 575) during the follow-up period; national medical records were linked to census data on an individual's occupation. The study population was composed of all those currently employed, aged 25–59 years at baseline, without prior history of asthma. Relative risks were calculated by comparing the occupation-specific incidence to the incidence of occupational asthma in administrative, managerial and clerical workers, whose risk was assumed to be similar to the background population risk. The relative risks were adjusted for age, and separate risks were available for males and females, although these were very close to each other and certainly within the limits of random variation. The study by Kogevinas et al. (1999) was a cross-sectional study of asthma involving 15 000 people in 12 European countries. In both studies, relative risks of asthma morbidity were assumed to apply for asthma mortality. This assumption is likely to be reasonable in most circumstances, but may lead to some underestimation or overestimation of asthma mortality, depending on whether exposure results in asthma incidence or exacerbation.

The approach used here was based on the work of Karjalainen et al. (2001, 2002). The work by Kogevinas et al. (1999) was also used for the relative risk of asthma due to occupational exposure in agriculture. While the Finnish study was large, prospective and covered all occupations, there was concern that Finnish exposures within specific occupations might be atypical of the rest of the world. In particular, this was considered likely to be true for agriculture, since Finnish agriculture might involve more indoor work where the relative risks for asthma were relatively high. Therefore the Kogevinas et al. results were used for agriculture since they were believed to be more generalizable to agriculture in the rest of the world, especially the developing world.

We assumed that the relative risk of asthma morbidity owing to employment in occupational categories was approximately equal to the relative risks of asthma mortality. Those not working and those employed in administration were together considered to be the nonexposed reference category (relative risk=1). These calculations were done separately for men and women for each subregion of the world. Rela-

Table 21.26 Relative risks for occupational asthma by original occupation and economic subsector, and sex, age-adjusted

Occupation	*Relative risk (males)*	*Relative risk (females)*	*Source*
Background	1.00	1.00	Non-working population, used as reference
Administration	1.00	1.00	Karjalainen et al. (2002), also used as reference
Technical	1.05	1.06	Karjalainen et al. (2002)
Sales	1.14	1.13	Karjalainen et al. (2002)
Agricultural	1.41	1.41	Kogevinas et al. (1999)
Mining	1.95	1.00	Karjalainen et al. (2002)
Transport	1.31	1.22	Karjalainen et al. (2002)
Manufacturing	1.56	1.33	Karjalainen et al. (2002)
Services	1.53	1.41	Karjalainen et al. (2002)

tive risks and the proportions exposed by occupational category were applied across all age groups from age 15 to ≥80 years. The relative risks by occupation are shown in Table 21.26.

Chronic obstructive pulmonary disease (COPD)

Tobacco smoking is clearly the most important risk factor, but many work-related exposures have been demonstrated to cause COPD (Hendrick 1996). A recent United States study estimating the magnitude of mortality due to occupational exposure (Steenland et al. 2003) used an estimate of 14% for the population-attributable fraction for COPD due to occupational dust exposure (based on a community study of severe COPD) (Korn et al. 1987), and found that COPD represented 87% of all fatal work-related nonmalignant respiratory disease, although some of the other types of respiratory disease may have been underestimated. A review of Finnish data also used a population-attributable fraction of 14% for men (and 5% for women) (Nurminen and Karjalainen 2001), and a similar figure (15%) was recently used in a review by the American Thoracic Society (Balmes et al. 2003).

As for asthma, difficulties arise from the vast array of definite, probable and possible causes of work-related COPD. The role of smoking, particularly in causing possible confounding effects, makes interpretation of studies difficult. Apparently significant individual differences in susceptibility, and uncertainty about pathological mechanisms, also cause problems. This area has been the subject of several reviews (Attfield and Wagner 1998; Balmes et al. 2003; Becklake 1989, 1994; Hendrick 1996; NIOSH 1996; Oxman et al. 1993), some covering all exposures and some concentrating on mineral dusts.

As a result of difficulties in characterizing all the likely causative occupational exposures, few published papers provide information that com-

prehensively describes the risk of developing COPD as a result of work. The paper by Korn et al. (1987) has been used in this analysis, as it provides relative risk information covering all workplace exposures. This study (the methods of which were described in more detail in an earlier study by Ferris et al. 1979) used data from a random sample of white adults aged 25–74 years from six United States cities and their surrounding areas (8515 people were included in the final sample). The definition of COPD was $FEV_1/FVC < 0.6$, representing reasonably severe disease. Logistical regression analyses were undertaken, determining the odds ratios for various respiratory conditions and controlling for age, sex, current and lifetime smoking history and city of residence. These odds ratios for COPD morbidity from Korn et al. (1987) were assumed to apply to COPD mortality.

The study by Korn et al. (1987) provides a strict definition of COPD and relative risks for both men and women, and was based on a large number of participants. This study was therefore used as the basis of the relative risk and attributable fraction estimates presented here. Relative risks for COPD prevalence were used as an approximation of the relative risks for COPD mortality.

Korn et al. (1987) found relative risks of COPD of 1.62 for men and 1.24 for women for a history of exposure to dusts. We partitioned these relative risks into high- and low-exposure categories, and also used slightly different relative risks for low exposure in developed and developing countries. In developing countries the great majority of low-exposure employment is in agriculture, where much dust is nonrespirable. In developed countries much of the exposure in the low categories is in industries other than agriculture, where a higher percentage of dust exposure may be respirable and toxic. It was assumed that the relative risks applied across all age categories. The estimated relative risks are shown in Table 21.27.

Risk reversibility

As for carcinogens, there are limited data on risk reversibility. The studies from which the estimated risks arise are based on cohorts of people

Table 21.27 Annual risks of COPD mortality

	Developing countries		*Developed countries (AMR-A, EUR-A, WPR-A)*	
Relative risk	*Men*	*Women*	*Men*	*Women*
Nonexposed	1.0	1.0	1.0	1.0
Low	1.2	1.1	1.4	1.2
High	1.8	1.4	1.8	1.4

Source: Korn et al. (1987).

exposed for different periods of time, followed up for various periods of time and with various periods of time between exposure cessation and follow-up, with follow-up periods varying between zero (still exposed) and many decades. Therefore, most of the absolute and relative risks produced by the studies are already dependent on whatever change in risk might occur once exposure ceases. Indications of risk reversibility for COPD may be obtained from the literature on smoking.

4. Occupational noise

Noise is a common occupational hazard. The unit for sound (noise) level, whether measuring noise exposure or hearing loss, is the decibel (dB). Noise exposure levels as used in this document have the unit dBA.[4] Noise-induced hearing loss is reported in dBHL. There is variability in the literature in the use of the terms to describe hearing ability. As used here, hearing loss refers to a decline in an individual's hearing ability. Hearing impairment refers to the effect of hearing loss on the individual's ability to function. The U.S. National Institute for Occupational Safety and Health (NIOSH) uses the term "material hearing impairment"[5] to describe a hearing loss greater than 25 dB, and most occupational studies refer to 25 dBHL. The WHO definition used in this study is hearing loss greater than or equal to 41 dBHL. Therefore, extrapolations were made from the occupational studies to fit the requirements of the WHO study.

4.1 Exposure variable

The exposure variable used in this analysis is a direct measure of the risk factor, i.e. occupational exposure to noise, which is the causative agent of noise-induced hearing loss. As global data on the frequency of occurrence, duration and intensity of noise exposure do not exist, it was necessary to model this exposure for workers employed in various occupational categories. The theoretical minimum is based on expected background levels of noise, and consistency with national and international standards. Most experts agree that levels below 80 dBA would result in minimal risk of developing hearing loss.

For workers in various occupational categories, three levels of exposure were estimated:

- minimum exposure, less than 85 dBA;
- moderately high noise, ≥85–90 dBA; and
- high noise, >90 dBA.

The choice of these levels was based on the recommended exposure limits (RELs) for occupational noise exposure around the world. In most developed countries the REL is 85 dBA as an eight-hour time-weighted average without hearing protection. In the United States the PEL is

90 dBA for an eight-hour day, although a hearing conservation programme is required for all employees exposed above 85 dBA for an eight-hour day. In developing countries, the REL is usually 90 dBA (Ahmed et al. 2001; Alidrisi et al. 1990; Hernandez-Gaytan et al. 2000; Hessel and Sluis-Cremer 1987; Osibogun at al 2000; Shaikh 1996; Sriwattanatamma and Breysse 2000).

Although the theoretical minimum exposure to noise was determined to be 80 dBA, it was not possible to estimate frequency of exposure by occupational category to occupational noise between 80 and 85 dBA. Therefore, persons with occupational exposure <85 dBA were included with the background population.

4.2 Estimating risk factor levels

Data sources

Potentially useful studies were identified using the various approaches described in section 1. The key terms used were "occupational noise" and "occupational hearing impairment". Relevant studies were identified by critically appraising the references obtained. This included consideration of the approaches to selection, measurement, analysis and control of confounding. Potential confounders of noise-induced hearing loss include nonoccupational exposure to noise, undocumented occupational noise levels, use of personal protective equipment, use of some medicines, and outer- and middle-ear pathology. Recent review articles were used where available, and the main articles were obtained and appraised.

The main reason for excluding studies was that they did not contain data appropriate for determining risk of noise-induced hearing loss. Problems included an inappropriate (for this purpose) exposure measurement (such as reporting for only one or a few occupational groups or tasks); inappropriate (for this purpose) outcome measurement (such as dB per year loss with age or no data as to the number of cases vs total population); poorly characterized exposure or self-reported hearing loss; and inadequate control of confounding.

In the United States, about 9 million workers are exposed to time-weighted average sound levels of 85 dBA and above (Simpson and Bruce 1981, quoted in Suter 2000), and about 10 million have noise-induced hearing loss >25 dB (USDOL OSHA 2002b). About 17% of American "production workers" are exposed to average noise levels at or above 85 dBA (NIOSH 1998). In the European Union, 28% of workers surveyed reported that for at least 25% of the time they were occupationally exposed to noise loud enough to cause them to raise their voices during conversation (corresponding to approximately 85–90 dBA) (EASHW 2000). The highest percentages of exposed workers were reported for mining, quarrying, manufacturing and construction. Australia compensates about 10 000 people each year for noise-induced

hearing loss; evidence indicates that only one third of workers with noise-induced hearing loss file compensation claims (NOHSC 1993). Summary statistics on noise exposure are not available for most industrializing and nonindustrialized countries. However, most published reports indicate that average noise exposure levels are well above the recommended occupational level in many industrialized countries, which is generally established at 85–90 dB for an eight-hour work day (Suter 2000; WHO/FIOSH 2001).

Information on noise exposures and noise-induced hearing loss in developing countries is given in Table 21.28. These studies are characterized by high occupational noise exposure levels, and many report hearing losses in exposed workers. The authors generally recommended engineering controls and hearing conservation programmes, including hearing protection, indicating that hearing protection is not widely used. Seventeen studies conducted in 12 countries in South America, Africa and Asia reported noise levels in a wide range of workplaces, including mining and the manufacture of food, fabrics, printed materials, metal products, drugs and watches. Most studies provided ranges of sound levels, with the lowest reported noise levels often below 80 dBA and the upper levels always above 90 dBA. All the studies that examined the hearing ability of workers revealed increased rates of hearing impairment in noise-exposed workers compared to nonexposed controls.

Exposure estimation

Occupational exposure to elevated noise levels depends on a variety of factors, including (i) occupation and industry and (ii) workplace-specific factors such as type of facility and process, raw materials, machinery, tools, the existence of engineering and work practice controls, and the existence, condition and use of personal protective devices. Thus exposure assessment was conducted using the occupational category approach (Equation 2), modified to reflect different noise exposures in occupations in different economic subsectors.

Our estimation of the proportion of workers in each occupational category with exposure to noise at or above 85 dBA (PEW(oc(r,g)i)) was based on United States data on the prevalence of noise exposure at or above 85 dBA among production workers in nine economic subsectors (NIOSH 1998; USDHHS 1986) (see Table 21.29).

The prevalence values among production workers were calculated from the US National Occupational Exposure Survey conducted during 1981–1983 (NIOSH 1998), which estimated the number of production workers exposed to noise at or above 85 dBA, by economic subsector. All other prevalence values were estimated by us, based on the NIOSH values for production workers. The value of 0.20 calculated for production workers in agriculture was extrapolated to all agricultural workers in all economic subsectors. Similarly, the value of 0.12 for

Table 21.28 Studies of noise exposures and hearing impairment in selected developing countries

Country or area	*Facility/job*	*Sound levels*	*Hearing loss*	*Notes*	*Source*
Brazil	Rotogravure printing workers	Continuous noise levels from 71 to 93 dBA	Some 49% of 124 workers exposed to noise and organic solvents had hearing loss (>25 dB) in the high frequencies, significantly associated with age		Morata et al. (1997)
Egypt	Road traffic policemen in Cairo	Average 97 dBA with horns, 85 without; 97 at railway crossings	About 20-dB loss at all frequencies compared to office policemen		Kamal et al. (1989)
Egypt	Textile factory	78–91 dBA in wool sorting and combing units	Compared to nonexposed controls, workers exposed to <85 dBA had only 1% increase in hearing impairment after 12 years. In workers exposed to >85 dBA the increased risk was 9.6%	Hearing impairment was defined as average of left and right ear thresholds at 0.5, 1 and 2 kHz, >25 dB	Moselhi et al. (1979)
Hong Kong SAR	Five industries: weaving, bottling, metal working, spinning, airport	L_{eq} (8-hour time-weighted average, dBA): weaving 102; bottling 94; metal working 96; spinning 97; airport 80–90	Compared to controls, noise-exposed workers had significantly higher thresholds in most age groups and in all five industries, closely matching predicted values	No evidence was found for any ethnic differences between western groups and Cantonese Chinese, either in general hearing ability or in response to long-term noise exposure	Evans and Ming (1982)
India	Heavy engineering industry: machine shop and press divisions	Ranged from 83–116 dBA. At selected work sites: press 94–110; machine shop 83–92; foundry 86–116	Mean hearing threshold: 40 controls 4–24 dB; 53 machine shop employees 14–40 dB; 60 press employees 19–70 dB	Hearing impairment was progressive with age for all groups. Use of hearing protection was recommended	Raja and Ganguly (1983)

India	Textile mill weavers	102–104 dBA	120 weavers, exposed 1–15 years. In the age range 30–34 years, median threshold of audibility in the right/left ear was 55/55 compared to 15/15 for controls; for 35–39-year olds the threshold was 60/55 compared to 15/15 for controls		Bhattacharya et al. (1981)
India	Drug and pharmaceutical company	Noise levels in dBA: fermentation 100–105; air compressor 95–102; ammonia compressor 93–97; primary air filter 104–106. Night shift levels were 1–3 dBA higher	—	Authors recommended engineering controls and hearing conservation programme, including use of hearing protection	Bhattacharya et al. (1990)
India	Watch factory in Bangalore	Maximum noise levels ranged from 74 in assembly to 99 dBA in the diesel generator room	—		Mukherjee et al. (1995)
Nigeria	Car assembly	94–108 dB	Hearing thresholds of 165 workers were significantly higher than nonexposed controls, and correlated significantly with employment duration		Oleru et al. (1990)

continued

Table 21.28 Studies of noise exposures and hearing impairment in selected developing countries (*continued*)

Country or area	*Facility/job*	*Sound levels*	*Hearing loss*	*Notes*	*Source*
Nigeria	Textile workers in five factories in Lagos	Continuous noise levels of 95–115 dBA	Hearing thresholds of 61 noise-exposed workers were significantly higher than 90 nonexposed controls. After 7 years of employment, exposed workers lost 2–12 dB per year, compared to 0.6–1.8 dB per year in controls	No hearing protection worn. Exposed workers did not display 4000-Hz notch, and the shape of the audiograms[a] was convex upwards, indicating lower losses at the middle frequencies. (Typical audiograms with noise-induced hearing loss display a convex downwards shape, indicating higher losses at the middle frequencies)	Oleru (1980)
Pakistan	Polyester fibre plant	Average noise levels: filament take-up unit 93.2 dBA; texturizing unit 94.8 dBA; compressor house 99.5 dBA	—	Typical exposure is 48 hours per week in these areas. Author recommended engineering controls and hearing conservation, including use of hearing protection	Shaikh (1996)
Saudi Arabia	78 factories producing food, chemicals, plastics, metals, paper and other products	86% exceeded 85 dBA, at least in part of the factory. In 12% all of the factory exceeded 85 dBA	—	None of the factories practised noise protection	Alidrisi et al. (1990)
Singapore	Audiometric testing of noise-exposed workers is mandatory in Singapore. Most cases of noise-induced	Noise dosimetry on 46 of these cases showed a mean time-weighted exposure of 90 dBA	127 cases of NID identified from 1985–1994. On average, after 24 years of exposure, the mean hearing threshold at 1, 2 and 3 kHz was 62 dB	Author stated that NID is the leading occupational disease in Singapore, with >500 new cases per year	Tay (1996)

	deafness (NID) are in those employed in shipping and metal manufacturing, the remainder in transport, quarrying and other manufacturing				
South Africa	Gold mining (cross-sectional survey of 2667 workers in Johannesburg)	Authors quoted a noise survey in which the majority of underground and surface gold mining occupations were exposed above 85 dBA	Hearing impairment was defined as average hearing loss of >25 dB for 500, 1000 and 2000 Hz, with 5 times weighting of better ear. None of the miners <22 years had hearing impairment, rising progressively to 22% of those ≥58 years old	Use of hearing protection increased from 13% in 1979 to 17% in 1982	Hessel and Sluis-Cremer (1987)
Sudan	Cotton ginning	99–107 dB	—	Newly mechanized facility	Khogali (1970)
United Arab Republic	Textile industry (El-Mehalla El-Kobra)	Average of 98 dB in 1200–4800 Hz range; up to 103 dBA	92% (60/73) of workers exposed to noise for ≥10 years in weaving departments had mean hearing impairment of 60 dB compared to 20 dB for control group	Audiometric test methods not described; hearing impairment not defined	Noweir et al. (1968)
Zambia	Copper mines (based on author's experiences as ear, nose and throat specialist in Zambian copper belt in 1975–1977)	"Continuous noise"	100 miners tested audiometrically. Of those with over 20 years, 23% were completely deaf	No hearing protection worn by miners	Obiako (1979)

— No data.

[a] Measurement of hearing loss using an audiometer that produces sounds at specific frequencies and sound pressure levels. The hearing threshold level is a function of frequency, indicating how a person hears at a given time.

Table 21.29 Prevalence of noise exposure ≥85 dBA

Economic subsector	Occupational category						
	Professional	*Administrative*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*[a]
Agriculture	0.05	0.05	0.05	0.12	0.12	0.20	0.20
Mining	0.05	0.05	0.05	0.12	0.12	0.20	0.85
Manufacturing	0.05	0.05	0.05	0.12	0.12	0.20	0.22
Electricity	0.05	0.05	0.05	0.12	0.12	0.20	0.15
Construction	0.05	0.05	0.05	0.12	0.12	0.20	0.18
Trade	0.02	0.02	0.02	0.12	0.12	0.20	0.13
Transport	0.02	0.02	0.02	0.12	0.12	0.20	0.12
Finance[b]	0.02	0.02	0.02	0.12	0.12	0.20	0.02
Services	0.02	0.02	0.02	0.12	0.12	0.20	0.03

[a] Source: NIOSH (1998).

[b] Based on 1.5% of workers exposed to noise in "business services".

production workers in transportation was extrapolated to all sales and service workers. The value for professional, administrative and clerical workers was extrapolated from 0.02 indicated for production workers in business services. The remaining value, 0.05 for professional, administrative and clerical workers in agriculture, mining, manufacturing, electricity and construction, was based on expert judgement.

The prevalence values were then partitioned into moderately high and high noise exposures, i.e. ≥85–90 dBA and >90 dBA, to estimate the proportions of workers exposed to moderately high and high levels of noise (EPF(r), exposure partitioning factor). Data from the United States (USDHHS 1986), taken from the 1981 Occupational Safety and Health Administration Final Regulatory Analysis for the Hearing Conservation Amendment, provide the distribution of noise exposure of over nine million American production workers (see Table 21.30). Of the 6 063 000 production workers with exposure at or above 85 dBA, slightly over half (3 407 000 or 56%) were exposed above 90 dBA. The distribution of noise exposure levels among workers exposed over 90 dBA was also used to determine that 95 dBA is a reasonable level of noise to estimate risks among the workers in the high-exposure group (>90 dBA).

The partitioning of workers by occupational category and noise level was assigned as follows, based on data in Table 21.30. Among production workers exposed at or above 85 dBA, half were considered to be exposed at ≥85–90 dBA and half exposed at >90 dBA. (Note that these partitioning values do not consider the use of personal protective equipment.) Of the agricultural workers and sales and service workers exposed at or above 85 dBA it was assumed, based on expert judgement, that approximately 70% are exposed at ≥85–90 dBA and 30% at >90 dBA.

Table 21.30 Distribution of 9 368 000 United States production workers who had noise exposure levels of 80 dBA or greater

Noise-exposure level (dBA)	*Number of workers*
80–85	3 305 000
86–90	2 656 000
91–95	1 936 000
96–100	965 000
>100	506 000
Total >85	6 063 000
Total >90	3 407 000

Source: USDOL OSHA 1981, cited in NIOSH (1991).

Table 21.31 Prevalence of noise exposure 85–90 dBA in A subregions

Economic subsector	*Occupational category*						
	Professional	*Administrative*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*
Agriculture	0.05	0.05	0.05	0.09	0.09	0.14	0.10
Mining	0.05	0.05	0.05	0.09	0.09	0.14	0.43
Manufacturing	0.05	0.05	0.05	0.09	0.09	0.14	0.11
Electricity	0.05	0.05	0.05	0.09	0.09	0.14	0.08
Construction	0.05	0.05	0.05	0.09	0.09	0.14	0.09
Trade	0.02	0.02	0.02	0.09	0.09	0.14	0.07
Transport	0.02	0.02	0.02	0.09	0.09	0.14	0.06
Finance	0.02	0.02	0.02	0.09	0.09	0.14	0.01
Services	0.02	0.02	0.02	0.09	0.09	0.14	0.02

All professional, administrative and clerical workers with noise exposure at or above 85 dBA were assumed to be at the ≥85–90-dBA level. Tables 21.31 and 21.32 present the distribution of noise exposure levels among workers in the A subregions by occupational category within economic sectors.

In the absence of global data, it was assumed that the same proportion of workers in these occupational categories in the developing countries would be exposed to noise levels at or above 85 dBA (B, C, D, and E subregions). Given the rarity of hearing conservation programmes in the developing subregions, it was assumed that 5% of production workers would be exposed in the ≥85–90 dBA category and 95% in the >90 dBA category (as opposed to 50/50 for the A subregions). Additionally, because mechanization is not widespread for D and E subregions, the majority (95%) of the agricultural workers exposed at or above 85 dBA were assigned to the ≥85–90-dBA level. Assignment of all

Table 21.32 Prevalence of noise exposure >90 dBA in A subregions

Economic subsector	Occupational category						
	Professional	*Administrative*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*
Agriculture	0	0	0	0.03	0.03	0.06	0.10
Mining	0	0	0	0.03	0.03	0.06	0.43
Manufacturing	0	0	0	0.03	0.03	0.06	0.11
Electricity	0	0	0	0.03	0.03	0.06	0.08
Construction	0	0	0	0.03	0.03	0.06	0.09
Trade	0	0	0	0.03	0.03	0.06	0.07
Transport	0	0	0	0.03	0.03	0.06	0.06
Finance	0	0	0	0.03	0.03	0.06	0.01
Services	0	0	0	0.03	0.03	0.06	0.02

Table 21.33 Prevalence of noise exposure 85–90 dBA in B and C subregions

Economic subsector	Occupational category						
	Professional	*Administrative*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*
Agriculture	0.05	0.05	0.05	0.09	0.09	0.14	0.01
Mining	0.05	0.05	0.05	0.09	0.09	0.14	0.04
Manufacturing	0.05	0.05	0.05	0.09	0.09	0.14	0.01
Electricity	0.05	0.05	0.05	0.09	0.09	0.14	0.04
Construction	0.05	0.05	0.05	0.09	0.09	0.14	0.01
Trade	0.02	0.02	0.02	0.09	0.09	0.14	0.01
Transport	0.02	0.02	0.02	0.09	0.09	0.14	0.01
Finance	0.02	0.02	0.02	0.09	0.09	0.14	0.00
Services	0.02	0.02	0.02	0.09	0.09	0.14	0.00

other occupational categories was the same as for the A subregions. Tables 21.33–21.36 reflect the different partitioning for the B + C and D + E subregions.

Table 21.37 presents the proportions of workers exposed to moderately high and to high noise levels by subregion, age and sex. The proportions of males exposed to these noise levels were consistently higher than those of females, owing both to higher rates of participation in the labour force and to higher rates of females working in the services sector.

4.3 Risk factor–disease relationships

High noise levels in the workplace may cause elevated blood pressure, sleeping difficulties, annoyance and stress. Excessive noise can interfere

Table 21.34 Prevalence of noise exposure >90 dBA in B and C subregions

Economic subsector	*Occupational category*						
	Professional	*Administrative*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*
Agriculture	0.00	0.00	0.00	0.03	0.03	0.06	0.19
Mining	0.00	0.00	0.00	0.03	0.03	0.06	0.81
Manufacturing	0.00	0.00	0.00	0.03	0.03	0.06	0.21
Electricity	0.00	0.00	0.00	0.03	0.03	0.06	0.14
Construction	0.00	0.00	0.00	0.03	0.03	0.06	0.17
Trade	0.00	0.00	0.00	0.03	0.03	0.06	0.12
Transport	0.00	0.00	0.00	0.03	0.03	0.06	0.11
Finance	0.00	0.00	0.00	0.03	0.03	0.06	0.02
Services	0.00	0.00	0.00	0.03	0.03	0.06	0.03

Table 21.35 Prevalence of noise exposure 85–90 dBA in D and E subregions

Economic subsector	*Occupational category*						
	Professional	*Administrative*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*
Agriculture	0.05	0.05	0.05	0.09	0.09	0.19	0.01
Mining	0.05	0.05	0.05	0.09	0.09	0.19	0.04
Manufacturing	0.05	0.05	0.05	0.09	0.09	0.19	0.01
Electricity	0.05	0.05	0.05	0.09	0.09	0.19	0.01
Construction	0.05	0.05	0.05	0.09	0.09	0.19	0.01
Trade	*0.02*	*0.02*	*0.02*	0.09	0.09	0.19	0.01
Transport	*0.02*	*0.02*	*0.02*	0.09	0.09	0.19	0.01
Finance	*0.02*	*0.02*	*0.02*	0.09	0.09	0.19	0.00
Services	*0.02*	*0.02*	*0.02*	0.09	0.09	0.19	0.00

with communications in the workplace, resulting in property damage or personal injury. Tinnitus[6] and temporary threshold shift[7] may also occur. However, the most serious effect is irreversible hearing impairment, resulting from damage to the delicate hearing mechanisms of the inner ear. Noise-induced hearing loss typically begins in the frequency range (pitch) of human voices and thus interferes with spoken communication.

Noise-induced hearing loss is caused by exposure to loud noises, such as those produced by woodworking equipment, chain saws, heavy machinery, gunfire, aircraft or amplified music. Permanent hearing loss from exposure to noise may happen quite early and an audiometric notch, or initial loss at or around 4000 Hz, may be noticeable within six

Table 21.36 Prevalence of noise exposure >90 dBA in D and E subregions

Economic subsector	*Occupational category*						
	Professional	*Administrative*	*Clerical*	*Sales*	*Service*	*Agriculture*	*Production*
Agriculture	0.00	0.00	0.00	0.03	0.03	0.01	0.19
Mining	0.00	0.00	0.00	0.03	0.03	0.01	0.81
Manufacturing	0.00	0.00	0.00	0.03	0.03	0.01	0.21
Electricity	0.00	0.00	0.00	0.03	0.03	0.01	0.14
Construction	0.00	0.00	0.00	0.03	0.03	0.01	0.17
Trade	0.00	0.00	0.00	0.03	0.03	0.01	0.12
Transport	0.00	0.00	0.00	0.03	0.03	0.01	0.11
Finance	0.00	0.00	0.00	0.03	0.03	0.01	0.02
Services	0.00	0.00	0.00	0.03	0.03	0.01	0.03

months to one year from starting a job with a hazardous noise exposure. There is significant variation in the susceptibility to noise damage, so that two workers with the same exposure may not experience the same hearing impairment. With prolonged exposure to the same noise, hearing loss continues to worsen. For a given noise environment, most of the hearing loss occurs in the first few years, although there is a slower continuing progression as long as the noise exposure continues.

When a person is removed from the noise, hearing loss does not worsen but does remain permanent. Any additional hearing loss after termination of work in a noisy environment is due to other causes, most often presbycusis (age-related hearing loss). Most people are subject to presbycusis, which is the most common form of sensorineural hearing impairment. Data show that from as early as 30 years of age, and gradually increasing in later years, some hearing loss occurs in the general population. Individual variation is great, with around 50% of the population maintaining good hearing into old age. Other factors, such as ear infection secondary to airborne contaminants, mechanical injury or chemical substances can lead to or aggravate hearing impairment in the workplace.

Description of studies

In the literature review, only three studies were found that indicated the frequency of hearing impairment at different thresholds of hearing (Davis 1989; Malchaire 2000; Waitzman and Smith 1999). Malchaire compared the expected percentage of subjects who, at age 55, presented with hearing impairment at 25 and 50 dBHL with the personal level of exposure (Lpe) to noise in dBA, in the absence of noises >140 dB. The exposure time frame was 35 years (Table 21.38).

Table 21.37 Proportions of the working-age population occupationally exposed to different noise levels, by sex and subregion

Subregion	Sex	Exposure level	Age group (years)					
			15–29	30–44	45–59	60–69	70–79	≥80
AFR-D	Male	<85 dBA	0.87	0.84	0.84	0.86	0.89	0.95
		85–90 dBA	0.09	0.12	0.11	0.10	0.08	0.04
		>90 dBA	0.04	0.04	0.04	0.04	0.03	0.01
	Female	<85 dBA	0.92	0.90	0.90	0.92	0.96	0.98
		85–90 dBA	0.07	0.09	0.09	0.07	0.04	0.02
		>90 dBA	0.01	0.01	0.01	0.01	0.01	0.00
AFR-E	Male	<85 dBA	0.87	0.84	0.84	0.86	0.89	0.95
		85–90 dBA	0.09	0.12	0.11	0.10	0.08	0.04
		>90 dBA	0.04	0.04	0.04	0.04	0.03	0.02
	Female	<85 dBA	0.92	0.90	0.90	0.92	0.96	0.98
		85–90 dBA	0.07	0.08	0.08	0.07	0.04	0.02
		>90 dBA	0.01	0.01	0.01	0.01	0.01	0.00
AMR-A	Male	<85 dBA	0.92	0.90	0.90	0.91	0.93	0.97
		85–90 dBA	0.05	0.06	0.06	0.06	0.04	0.02
		>90 dBA	0.03	0.04	0.04	0.03	0.02	0.01
	Female	<85 dBA	0.96	0.95	0.95	0.96	0.98	0.99
		85–90 dBA	0.03	0.03	0.03	0.03	0.01	0.01
		>90 dBA	0.01	0.01	0.01	0.01	0.01	0.00
AMR-B	Male	<85 dBA	0.90	0.87	0.88	0.89	0.91	0.96
		85–90 dBA	0.05	0.06	0.06	0.05	0.04	0.02
		>90 dBA	0.06	0.07	0.07	0.06	0.05	0.02
	Female	<85 dBA	0.95	0.94	0.94	0.95	0.97	0.99
		85–90 dBA	0.03	0.03	0.04	0.03	0.02	0.01
		>90 dBA	0.02	0.03	0.03	0.02	0.01	0.01
AMR-D	Male	<85 dBA	0.91	0.89	0.89	0.90	0.93	0.96
		85–90 dBA	0.03	0.04	0.04	0.04	0.03	0.01
		>90 dBA	0.05	0.07	0.07	0.06	0.05	0.02
	Female	<85 dBA	0.95	0.94	0.94	0.96	0.97	0.99
		85–90 dBA	0.02	0.03	0.03	0.02	0.01	0.01
		>90 dBA	0.03	0.03	0.03	0.02	0.01	0.01
EMR-B	Male	<85 dBA	0.91	0.88	0.89	0.90	0.92	0.96
		85–90 dBA	0.04	0.05	0.05	0.04	0.03	0.02
		>90 dBA	0.05	0.07	0.07	0.06	0.05	0.02
	Female	<85 dBA	0.96	0.95	0.95	0.96	0.98	0.99
		85–90 dBA	0.02	0.03	0.03	0.02	0.01	0.01
		>90 dBA	0.02	0.02	0.02	0.02	0.01	0.01
EMR-D	Male	<85 dBA	0.88	0.85	0.85	0.86	0.90	0.95
		85–90 dBA	0.09	0.11	0.11	0.10	0.07	0.04
		>90 dBA	0.04	0.04	0.04	0.04	0.03	0.02
	Female	<85 dBA	0.91	0.89	0.89	0.92	0.95	0.98
		85–90 dBA	0.07	0.09	0.09	0.07	0.04	0.02
		>90 dBA	0.02	0.02	0.02	0.02	0.01	0.00
EUR-A	Male	<85 dBA	0.92	0.90	0.90	0.91	0.93	0.97
		85–90 dBA	0.05	0.06	0.06	0.06	0.04	0.02
		>90 dBA	0.03	0.04	0.04	0.03	0.02	0.01

continued

Table 21.37 Proportions of the working-age population occupationally exposed to different noise levels, by sex and subregion (*continued*)

Subregion	Sex	Exposure level	Age group (years)					
			15–29	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
	Female	<85 dBA	0.96	0.96	0.95	0.97	0.98	0.99
		85–90 dBA	0.03	0.03	0.03	0.02	0.01	0.01
		>90 dBA	0.01	0.01	0.01	0.01	0.01	0.00
EUR-B	Male	<85 dBA	0.88	0.85	0.85	0.87	0.90	0.95
		85–90 dBA	0.05	0.06	0.06	0.05	0.04	0.02
		>90 dBA	0.07	0.09	0.09	0.08	0.06	0.03
	Female	<85 dBA	0.93	0.91	0.91	0.93	0.96	0.98
		85–90 dBA	0.04	0.05	0.05	0.04	0.02	0.01
		>90 dBA	0.03	0.04	0.04	0.03	0.02	0.01
EUR-C	Male	<85 dBA	0.88	0.85	0.85	0.87	0.90	0.95
		85–90 dBA	0.04	0.05	0.05	0.04	0.03	0.02
		>90 dBA	0.08	0.10	0.10	0.09	0.07	0.04
	Female	<85 dBA	0.93	0.92	0.92	0.94	0.96	0.98
		85–90 dBA	0.02	0.03	0.03	0.02	0.01	0.01
		>90 dBA	0.04	0.05	0.05	0.04	0.02	0.01
SEAR-B	Male	<85 dBA	0.88	0.84	0.84	0.87	0.91	0.95
		85–90 dBA	0.06	0.08	0.08	0.06	0.04	0.02
		>90 dBA	0.06	0.08	0.08	0.07	0.05	0.03
	Female	<85 dBA	0.92	0.88	0.88	0.91	0.96	0.98
		85–90 dBA	0.05	0.08	0.08	0.05	0.02	0.01
		>90 dBA	0.04	0.04	0.04	0.03	0.02	0.01
SEAR-D	Male	<85 dBA	0.87	0.79	0.80	0.84	0.88	0.94
		85–90 dBA	0.09	0.15	0.15	0.11	0.08	0.04
		>90 dBA	0.04	0.05	0.05	0.05	0.04	0.02
	Female	<85 dBA	0.91	0.94	0.95	0.96	0.98	0.99
		85–90 dBA	0.07	0.04	0.03	0.02	0.01	0.01
		>90 dBA	0.02	0.02	0.02	0.02	0.01	0.00
WPR-A	Male	<85 dBA	0.92	0.90	0.90	0.92	0.96	0.98
		85–90 dBA	0.04	0.06	0.06	0.04	0.02	0.01
		>90 dBA	0.03	0.04	0.04	0.03	0.03	0.01
	Female	<85 dBA	0.95	0.94	0.94	0.96	0.97	0.99
		85–90 dBA	0.03	0.04	0.04	0.03	0.02	0.01
		>90 dBA	0.02	0.02	0.02	0.01	0.01	0.00
WPR-B	Male	<85 dBA	0.87	0.84	0.84	0.86	0.89	0.95
		85–90 dBA	0.06	0.08	0.08	0.07	0.05	0.03
		>90 dBA	0.07	0.08	0.08	0.07	0.06	0.03
	Female	<85 dBA	0.93	0.91	0.91	0.93	0.96	0.98
		85–90 dBA	0.04	0.05	0.05	0.04	0.02	0.01
		>90 dBA	0.03	0.04	0.04	0.03	0.02	0.01

Table 21.38 Expected percentages of workers with hearing loss of >25 dB or >50 dB after 35 years of exposure, by personal level of exposure

	Personal exposure (dBA)							
Hearing loss	*85*	*90*	*92*	*94*	*97*	*98*	*99*	*100*
50 dB	4	5	7	9	15	18	21	26
25 dB	29	36	40	46	59	65	70	75

Source: Malchaire (2000).

Waitzman and Smith (1999) analysed data from the United States Health Examination Survey (1960–1961) and the National Health and Nutrition Examination Survey (NHANES I, 1971–1975). Hearing loss was rated by a scheme developed by Klockhoff et al. (1974). The Klockhoff analysis used four hearing loss levels (Slight, Moderate, Severe1, Severe2), which were related in this analysis to the standard hearing impairment scales by using typical noise-induced hearing loss curves. The categories developed by Klockhoff et al. were based on the frequencies used in the National Health Examination Survey (500, 1000, 2000, 3000, 4000 and 6000 Hz), and the NHANES I study (500, 1000, 2000 and 4000 Hz). As an example of this conversion procedure, the criteria of Klockhoff et al. for slight hearing loss were based on a 30-dB loss at 4000 Hz. Using audiometric data presented by Klockhoff et al. (1974), we estimated the equivalent hearing losses for each category as follows:

- Slight: >21 dBHL
- Moderate: >38 dBHL
- Severe1: >41 dBHL
- Severe2: >50 dBHL

Waitzman and Smith reported odds ratios calculated by multivariate regressions for two age groups for workers in construction, manufacturing/mining and other subsectors (see Table 21.39). Blue-collar construction workers experienced between 2 and >3.5 times the risk experienced by white-collar workers in "other industries". The pattern of their hearing loss at normal speech frequencies significantly disrupted their ability to communicate.

Davis (1989) reported on the prevalence of hearing loss as a function of age in the adult population of Great Britain. Audiometric analyses on adults ranging in age from 17 to ≥80 years were conducted in four cities. Hearing impairment was reported for >25, >45 and >65 dBHL. Davis found a "significant" level of hearing loss (>25 dBHL) in 16% of the adult population (17–≥80 years).

Table 21.39 Odds ratios from logistic multivariate regressions on audiometric measures of hearing loss, by age group

	Ages 25–44 years				*Ages 45–65 years*			
Type of worker/industry	*Slight >21 dB*	*Moderate >38 dB*	*Severe 1 >41 dB*	*Severe 2 >50 dB*	*Slight >21 dB*	*Moderate >38 dB*	*Severe 1 >41 dB*	*Severe 2 >50 dB*
White collar								
Construction	0.85	0.00	0.00	0.00	2.50	2.17	1.68	2.18
Manufacturing/ mining	0.91	1.45	1.17	1.14	1.02	1.43	0.98	1.25
Other industry groups	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Blue collar								
Construction	2.79	3.50	2.34	2.65	3.08	3.54	1.98	2.38
Manufacturing/ mining	2.01	3.03	1.94	2.40	2.33	1.86	1.88	1.90
Other industry groups	1.38	2.42	1.92	1.95	1.84	1.94	1.40	1.69

Source: Waitzman and Smith (1999).

Several studies were found that presented relative risk estimates by specific occupation. The risk estimates were higher than those determined in this analysis for exposed workers, as they focused on occupations with generally high noise exposures. These studies were based in Canada (Hessel 2000), Germany (Arndt et al. 1996) and Great Britain (Palmer et al. 2001) (see Table 21.40).

In Germany, a prevalence ratio for hearing loss of 1.5 was found in construction workers vs white-collar employees. Hearing impairment was defined as the sum of thresholds at 2000, 3000 and 4000 Hz greater than 105 dB at least in one ear. Hessel (2000), in a similar study in Canada, found that construction workers (with the exception of boilermakers) had a lower prevalence of hearing loss than the Germans. Also, prevalence of hearing impairment in the comparison group in Canada was lower than in Germany. According to the author, the differences found may be due to the year of the study. The Canadian study was carried out 7–9 years later than the German study, and there may have been lower occupational noise levels and/or use of personal hearing protection. These potentially confounding factors were not described in the German study.

In contrast to the Canadian and German studies, prevalence ratios determined in Great Britain were based on self-reported hearing impairment. Prevalence of "ever employed in a noisy job" was compared against "never exposed in a noisy job". A noisy place was defined as one "where there was a need to shout to be heard". The questions used to define hearing impairment were modelled on those used in the Institute

Table 21.40 Prevalence ratio of occupational noise-induced hearing impairment in available studies

Country	*Reference*	*Definition of hearing impairment*	*Occupation*	*Prevalence ratio*	*95% CI*
Canada (Edmonton)	Hessel (2000)	Greater than 105 dB hearing loss at 2, 3, 4 kHz (corresponds to >35 dBHL)	Plumbers	2.91	—
			Boilermakers	3.88	—
			Electricians	1.46	—
Germany	Arndt et al. (1996)	Greater than 105 dB healing loss at 2, 3, 4 kHz (corresponds to >35 dBHL)	Carpenters	1.77	1.48–2.12
			Unskilled workers	1.75	1.47–2.09
			Plumbers	1.49	1.19–1.85
			Painters	1.2	0.96–1.49
			Plasterers	1.29	1.05–1.59
			Overall	1.5	1.29–1.82
Great Britain	Palmer et al. (2001)	Severe: wearing a hearing aid or having great difficulty in both ears on hearing conversation in a quiet room >45 dBHL	Male	2.9	—
			Female	1.8	—
		Moderate and worse: reported moderate difficulty in hearing conversation in a quiet room equivalent to 45 dBHL	Male	3.6	—
			Female	2.9	—

— No data.

of Hearing Research national survey of hearing, in which those who reported moderate or worse hearing impairment were found to have a mean hearing loss of 45 dBHL. The mean hearing loss of 45 dBHL is similar to the cut-off of 41 dBHL used by WHO and the Global Burden of Disease study, whereas the cut-off used in the German and Canadian studies corresponded to slight hearing impairment (>35 dBHL).

NIOSH, in a re-analysis of the data from its Occupational Noise and Hearing Survey (Prince et al. 1997) derived excess risk[8] estimates with a model that used the average of 1000, 2000, 3000 and 4000 Hz and a hearing loss >25 dBHL. We used this information to develop excess risk estimates for workers exposed at 85–90 dBA (defined by us as moderately high exposure) and >90 dBA (defined by us as high noise exposure, equivalent to 95 dBA). To estimate excess risks for workers exposed to moderately high noise, we used the observed exposure–response relationships developed by NIOSH (1998) for workers of different ages who were exposed at 80, 85 and 90 dBA for various numbers of years. The data show that at any noise level, hearing impairment increases with age and/or length of exposure. Also, the highest risk is found at the highest levels of exposure. Prince et al. (1997) found a small increase in excess risk in workers exposed to 80–84 dBA vs a <80 dBA control group; however, these risk estimates are imprecise owing to the low numbers of workers in the study exposed to noise at these levels.

NIOSH (1998) also provides two data points of excess risk for workers exposed at 95 dBA for prolonged periods. Table 21.41 illustrates our estimation of excess risk of material hearing impairment at >25 dBHL for workers exposed at 95 dBA, based on these two data points. The excess risk value of 38.3% at 95 dBA for a 65-year-old worker after 10 or more years of exposure was taken from NIOSH (1998), Appendix Table IV. In addition, an excess risk value of 19.5% was taken from the table for 30-year-old workers exposed to noise at 95 dBA and for a duration of exposure of 5–10 years. The values for 30-, 40- and 50-year olds with >10 years of exposure were interpolated using the ratios of change of hearing loss with age at 90 dBA between each age group in Table 21.41. All calculations of NIOSH exposure–response relationships were based on material hearing impairment at >25 dBHL and were adjusted by us to reflect noise-induced hearing loss at ≥41 dBHL.

The International Organization for Standardization (ISO) has also developed procedures for estimating hearing loss due to noise exposure. Their most recent model (referred to as the "1990-ISO model") and the 1997 NIOSH model (NIOSH 1998) are reasonably similar. Table 21.42 summarizes the excess risk estimates developed separately by NIOSH and ISO for material hearing impairment >25 dBHL caused by occupational noise exposure.

Table 21.41 Excess risk estimates for material hearing impairment >25 dBHL, by age and duration of exposure

	Excess risk (%)			
Average daily exposure (dBA)	*Age 30*	*Age 40*	*Age 50*	*Age 60*
5–10 years of exposure				
95[a] (estimated)	19.5			
90	5.4	9.7	14.3	15.9
85	1.4	2.6	4	4.9
80	0.2	0.4	0.6	0.8
>10 years of exposure				
95[a] (estimated)	24	31	38	38.3
90	10.3	17.5	24.1	24.7
85	2.3	4.3	6.7	7.9
80	0.3	0.6	1	1.3

[a] Estimates for 95 dBA were developed from NIOSH 1998 using methods described in the text.

Source: NIOSH (1998).

Table 21.42 NIOSH and ISO estimated excess risk of incurring material hearing impairment (>25 dBHL at 1, 2, 3 and 4 kHz) over a 40-year working lifetime and at various average noise exposures

	Excess risk (%)	
Average daily noise exposure (dBA)	*ISO*	*NIOSH*
90	17	25
85	6	8
80	1	1

Source: NIOSH (1998).

EXTRAPOLATING FROM RISKS AT >25 dBHL TO RISKS AT ≥41 dBHL

Occupational hearing loss is usually denoted as >25 dBHL but WHO uses ≥41 dBHL as a cut-off point to estimate prevalence of hearing loss. Therefore, extrapolations were made from studies of occupational risks at >25 dBHL to estimate risk to workers of hearing loss at 41 dBHL or greater. Data from the USDOL OSHA 1981 Final Regulatory Analysis for the Hearing Conservation Amendment (NIOSH 1991) provided a means of adjusting the various reports based on material hearing impairment >25 dBHL to >40 dBHL, a level of hearing loss assumed for this project to be equivalent to the WHO definition of ≥41 dBHL. As presented in Table 21.43, OSHA estimated the number of workers with various levels of hearing loss or impairment. The number of expected

Table 21.43 Hearing levels (dBHL) of 9368000 United States production workers with noise exposure levels of ≥80 dBA

Hearing threshold level (1, 2 and 3 kHz)	*Cumulative cases*	*Expected cases*	*Excess cases*
>15 dB (mild hearing loss)	3735000 (40%)	2111000 (23%)	1624000 (17%)
>25 dB (material hearing impairment)	2025000 (22%)	965000 (10%)	1060000 (11%)
>40 dB (moderate to severe hearing impairment)	718000 (8%)	245000 (3%)	473000 (5%)

Source: USDOL OSHA, 1981, cited in NIOSH (1991).

Table 21.44 Estimated excess risk of incurring hearing impairment at 41 dBHL or greater over a 40-year working lifetime and at various average noise exposures

	Excess risk (%)	
Average daily noise exposure (dBA)	*ISO*	*NIOSH*
90	7.6	11.2
85	2.7	3.6
80	0.4	0.4

cases (based on hearing levels of a nationwide sample of adults in U.S. Public Health Service hearing surveys) was subtracted to derive the number of excess cases at various levels of hearing loss or impairment in United States production workers (OSHA 1981, reported in NIOSH 1991).

Using the data in Table 21.43, a factor to correct excess risk data at >25 dBHL to WHO's excess risk at ≥41 dBHL was determined as a ratio of the number of excess cases at >40 dB divided by the number of excess cases at >25 dBHL; i.e. 473000 divided by 1060000 yields a ratio of 0.446. This correction factor of 0.446 was used to correct excess risk at >25 dBHL from the reported excess risk in Table 21.42 to the excess risk at ≥41 dBHL as presented in Table 21.44. As no additional data were available, the hearing impairment of production workers at >40 dBHL was assumed to be equivalent to the WHO definition of hearing loss of 41 dBHL or greater used in this project.

Excess risk estimates for hearing impairment at ≥41 dBHL are presented in Table 21.45. They were generated by applying the same correction factor of 0.446 to Table 21.41.

Relative risk estimation for noise induced hearing loss at ≥41 dBHL

The relative risk values were extrapolated using the following formula:

Table 21.45 Estimated excess risk for hearing impairment at ≥41 dBHL, by age and duration of exposure

	Excess risk (%)			
Average daily exposure (dBA)	*Age 30*	*Age 40*	*Age 50*	*Age 60*
5–10 years of exposure				
95	8.7			
90	2.4	4.3	6.4	7.1
85	0.6	1.2	1.8	2.2
80	0.1	0.2	0.3	0.4
>10 years of exposure				
95	10.7	13.8	16.9	17.0
90	4.6	7.8	10.7	11.0
85	1.0	1.9	3.0	3.5
80	0.1	0.3	0.4	0.6

$$\text{Relative risk} = 1 + (\text{excess risk} / \text{expected risk})$$

Excess risk is defined in this study as "the percentage of workers with a hearing impairment in an occupationally noise-exposed population, minus the percentage who would normally incur such impairment from aging in an unexposed population". The expected risk is the prevalence for the general unexposed population. While the NIOSH document provides the excess risk of the exposed population, the expected risk is not reported by NIOSH. Data from Davis (1989) estimates prevalence as a function of age in the adult population of Great Britain. The average prevalence for both ears for a noise-induced hearing loss of >45 dBHL was calculated by us for the general population, using the data from Davis (1989) and the methods described above to adjust NIOSH data for noise-induced hearing loss at >25 dBHL to generate Table 21.45. The results are reported in Table 21.46.

In our study, relative risks for the age groups 0–4 and 5–14 years were not estimated, as occupational risks are not present and/or data are unavailable on levels or length of exposure. In the calculation of excess risk, the age group 15–29 years was assigned the lowest excess risk value of 8.7 in Table 21.45 for age 30 with 5–10 years of exposure. For the category of workers with moderately high noise exposure (85–90 dBA), the excess risk estimate is the geometric mean of the excess risk estimates for 85 dBA and 90 dBA for each age group (see Table 21.45).

The older age groups (30–44 years, 45–59 years, etc.) did not neatly fit the age categories in Table 21.45, so worker-population-weighted averages were constructed for excess risk values at the required ages. For example, the excess risk estimate for the age group 30–44 years, at 85–90 dBA, was calculated by first taking the geometric mean of the excess risk

Table 21.46 Prevalence of hearing loss at >45 dBHL for the general population of Great Britain

Age group (years)	*Prevalence*
17–30	1.25
31–40	1.90
41–50	4.75
51–61	6.40
61–70	9.35
71–80	16.55
≥81	25.35

Source: Davis (1989).

Table 21.47 Relative risks by age group and level of exposure

	Age group (years)					
Level	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
<85 dBA	1.00	1.00	1.00	1.00	1.00	1.00
85–90 dBA	1.96	2.24	1.91	1.66	1.12	1.00
>90 dBA	7.96	5.62	3.83	2.82	1.62	1.00

estimates at 85 and 90 dBA, for people with >10 years of exposure in the 30- and 40-year-old categories. These values were then weighted by the worker population in the age groups 30–39 and 40–44 years. In a similar procedure, the prevalence values in Table 21.46 for the general population were adjusted to the same age groups. Relative risks for the age groups 70–79 years and ≥80 years were calculated from figures in Prince et al. (1997) and the prevalence data from Davis (1989). Table 21.47 presents the relative risks by age group and level of exposure.

Risk reversibility

It was assumed that risk reversibility following exposure removal was immediate. In other words, for those previously exposed who have not developed noise-induced hearing loss yet, the risk is removed after exposure removal (no new cases). Those with the condition, however, will continue to be affected by it.

5. Occupational ergonomic stressors

The physical ergonomic features of work that are most frequently cited as risk factors for MSDs include a rapid pace of work and repetitive motion patterns; insufficient recovery time; heavy lifting and other forceful manual exertions; non-neutral body postures (either dynamic or

static); mechanical pressure concentrations; vibration (both segmental and whole-body); and low temperature. For the present analysis, the risk factor is exposure to the combination of occupational exposures that are implicated in the etiology of low back pain, including physical stressors and possibly psychosocial or work organization features as well.

5.1 Exposure variable

Assessing the fraction of back pain disorders that can be attributed to occupation requires that an indicator be identified that can be measured on a global scale and that can also be matched with data on known exposure–risk relationship(s). The various reviews of low back pain epidemiology have implicated an overlapping set of occupational exposures, such as lifting, forceful movements, awkward postures, whole-body vibration and perhaps psychosocial stressors. However, such exposures are rarely assessed in surveillance activities on a large scale, and thus data are not available for risk assessment calculations at the global level.

In contrast, low back pain (and other MSD morbidity) is commonly reported by broad industrial or occupational groupings. Thus, even though occupation is a less precise indicator of risk than a specific exposure, its widespread availability in administrative data sets and some epidemiological studies makes it useful in this context. Some epidemiological studies have also provided sufficient data to relate back pain to the same occupational categories. Occupation was therefore considered as a proxy for specific risk factors. The exposure variable is an occupational category with its assigned level of risk (low, medium or high rate of low back pain). This method thus required the assumption that the distribution of the combined individual risk factors is similar within each occupational group across geographical regions. This argument applies to psychosocial as well as physical exposures.

Given that differences can occur within occupations, the assumed homogeneity of occupational groups in their total exposure to ergonomic risk factors implies that differences in exposure among occupations are substantially larger than differences among workers within the occupation. Although this assertion appears self-evident, only a few investigators to date have examined it explicitly. Burdorf (1992) evaluated the homogeneity of the exposure to postural load on the back within and among four occupational groups, and reported that the exposure variability within occupational groups was small compared with differences among groups. The estimated contribution of the variance for postural load between occupational groups was the largest source of variance. Hollman et al. (1999) and Paquet et al. (1999) have similarly shown that, even within one subsector (health care and construction, respectively), differences in ergonomic exposures among jobs can be large relative to the variability within jobs. These studies provide strong evidence in support of the approach taken in this analysis. Although the literature is less conclusive regarding their effects, to the extent that any specific

ergonomic factor is etiologically important it is assumed to be internalized in the relative risks by occupation. At the same time, specific analyses at the national or local level could increase the precision by assessing specific physical risk factors at the workplace.

5.2 Theoretical minimum risk

For low back pain, "theoretical minimum risk" is considered to represent the level of disease that would occur in the population if all excessive physical workload were abated by effective implementation of ergonomic control measures. While interventions to reduce ergonomic stressors have not yet been widely implemented on a global scale, studies are available from selected settings demonstrating the great potential of exposure (and disease) reduction in this area. Certain interventions have shown that removal of ergonomic stressors can practically lead to the removal of back pain (or its reduction to negligible levels), which justifies the choice of theoretical minimum. For example, in jobs where the entire work activity consists of manually handling materials, lifting equipment can successfully reduce both the biomechanical exposure to the lower back and the risk of low back disorders (Marras et al. 2000).

Westgaard and Winkel (1997) reviewed 89 studies on ergonomic intervention studies, of which 20 were classified as mechanical exposure interventions and 32 as production system interventions. Most mechanical exposure interventions target the ergonomic exposure level directly, through redesign of the work station and work methods. The reviewers concluded that in work situations whose mechanical exposure level is initially high, a reduction in the level of mechanical exposure may be beneficial for musculoskeletal health. For comparison, several other risk factors considered in this chapter have not yet shown that interventions can have such highly effective results when applied to selected population groups.

Since occupation represents a proxy for the composite of etiological exposures, rather than being the exposure *per se*, it is not necessary that persons leave one occupation for another in order to achieve the theoretical minimum risk. Reduction in relative risk would occur through improved job design to reduce exposures within each occupational category. The number of individuals working in each category could remain constant, even though the nature of the risk in each category would change.

5.3 Estimating risk factor levels

Since data were not available worldwide on the prevalence of specific ergonomic exposures, occupations were grouped here by their risk of low back pain. Thus, for this outcome, estimation of risk factor levels is not independent of the levels of disease assigned to them in the next step (see section 5.4). The exposure assessment by occupation was utilized, as described in the Introduction. Using managers and professionals as a

baseline for comparison, epidemiological studies have indicated that clerical and sales workers have a slightly elevated risk, operators and service workers have a moderate risk, and farmers have the highest risk of low back pain (see section 5.4 for details).

The basic approach was to determine economic activity by subregion, age and sex, and then to determine the distribution of the working population in the various occupational categories. As each occupational category was assigned a single relative risk factor (based on methodology described below), it was not necessary to partition exposures into "high" or "low" levels.

The estimates of occupations with risk of low back pain were based on the published literature. The 1968 International Standard Classification Codes for Occupations utilizes the term "production workers", whereas the epidemiological studies refer to "operators". Based on the literature, we made the following assignments.

- Background exposure: professional and administrative workers
- Low exposure: clerical and sales workers
- Moderate exposure: operators (production workers) and service workers
- High exposure: farmers

The non-working population is not considered in this analysis, and is attributed the same relative risk as the background exposure category. It is likely that younger workers are represented more often in the production occupations, and that older workers have more opportunities to move into management and administration positions.

As seen from Table 21.48, the majority of the working-age population is employed in occupations with exposure to factors linked to low back pain. Males have higher exposure in general, owing to higher rates of participation in the labour force. Exposures are higher for men in the less developed subregions, owing to a higher proportion of workers in agriculture than in the more developed subregions.

5.4 Risk factor–disease relationships

Pain in the soft tissues of the back is extremely common among adults. In the United States, the National Arthritis Data Workgroup reviewed national survey data showing that each year some 15% of adults report frequent back pain or pain lasting more than two weeks (Lawrence et al. 1998). Back pain is widespread in many countries, and is associated with substantial financial costs and loss of quality of life. In Canada, Finland and the United States, more people are disabled from working as a result of MSDs—especially back pain—than from any other group of diseases (Badley et al. 1994; Pope et al. 1991; Riihimäki 1995a). MSDs also constitute a major proportion of all registered and/or

Table 21.48 Proportions of the working-age population occupationally exposed to different levels of ergonomic stressor, by sex and subregion

Subregion	*Sex*	*Exposure category*	*Age group (years)*					
			15–29	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	Background	0.29	0.11	0.13	0.22	0.40	0.70
		Low	0.09	0.11	0.10	0.09	0.07	0.04
		Moderate	0.20	0.25	0.25	0.22	0.17	0.09
		High	0.42	0.53	0.52	0.46	0.35	0.18
	Female	Background	0.53	0.42	0.41	0.55	0.73	0.87
		Low	0.05	0.06	0.06	0.05	0.03	0.01
		Moderate	0.09	0.11	0.11	0.08	0.05	0.02
		High	0.34	0.41	0.42	0.32	0.19	0.09
AFR-E	Male	Background	0.29	0.12	0.14	0.22	0.40	0.70
		Low	0.08	0.10	0.10	0.09	0.07	0.03
		Moderate	0.20	0.25	0.25	0.22	0.17	0.09
		High	0.42	0.53	0.52	0.47	0.36	0.18
	Female	Background	0.41	0.34	0.37	0.50	0.67	0.83
		Low	0.06	0.07	0.07	0.05	0.04	0.02
		Moderate	0.11	0.13	0.12	0.10	0.06	0.03
		High	0.41	0.46	0.44	0.35	0.23	0.12
AMR-A	Male	Background	0.49	0.33	0.37	0.64	0.91	0.95
		Low	0.18	0.24	0.23	0.13	0.03	0.02
		Moderate	0.29	0.39	0.36	0.21	0.05	0.03
		High	0.03	0.04	0.04	0.02	0.01	0.00
	Female	Background	0.58	0.47	0.53	0.79	0.95	0.98
		Low	0.19	0.24	0.21	0.09	0.02	0.01
		Moderate	0.22	0.28	0.24	0.11	0.02	0.01
		High	0.01	0.02	0.02	0.01	0.00	0.00
AMR-B	Male	Background	0.32	0.16	0.23	0.43	0.71	0.86
		Low	0.15	0.19	0.17	0.13	0.06	0.03
		Moderate	0.37	0.46	0.43	0.32	0.16	0.08
		High	0.15	0.19	0.17	0.13	0.06	0.03
	Female	Background	0.63	0.57	0.69	0.84	0.94	0.97
		Low	0.10	0.12	0.09	0.04	0.02	0.01
		Moderate	0.21	0.24	0.18	0.09	0.03	0.02
		High	0.06	0.07	0.05	0.03	0.01	0.00
AMR-D	Male	Background	0.41	0.19	0.20	0.29	0.49	0.75
		Low	0.15	0.21	0.20	0.18	0.13	0.06
		Moderate	0.39	0.55	0.53	0.48	0.34	0.17
		High	0.04	0.06	0.06	0.05	0.04	0.02
	Female	Background	0.70	0.62	0.69	0.77	0.86	0.93
		Low	0.09	0.11	0.09	0.07	0.04	0.02
		Moderate	0.20	0.26	0.21	0.16	0.09	0.05
		High	0.01	0.01	0.01	0.01	0.01	0.00
EMR-B	Male	Background	0.48	0.23	0.27	0.41	0.64	0.82
		Low	0.15	0.22	0.21	0.17	0.10	0.05
		Moderate	0.27	0.40	0.38	0.31	0.19	0.09
		High	0.10	0.15	0.14	0.11	0.07	0.03

Table 21.48 Proportions of the working-age population occupationally exposed to different levels of ergonomic stressor, by sex and subregion (*continued*)

Subregion	Sex	Exposure category	Age group (years)					
			15–29	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
	Female	Background	0.77	0.74	0.82	0.87	0.94	0.97
		Low	0.10	0.11	0.08	0.05	0.03	0.01
		Moderate	0.10	0.12	0.08	0.06	0.03	0.01
		High	0.03	0.04	0.02	0.02	0.01	0.00
EMR-D	Male	Background	0.32	0.10	0.13	0.29	0.59	0.80
		Low	0.19	0.25	0.24	0.20	0.11	0.06
		Moderate	0.17	0.23	0.22	0.18	0.10	0.05
		High	0.32	0.43	0.41	0.33	0.19	0.10
	Female	Background	0.64	0.59	0.64	0.76	0.88	0.94
		Low	0.06	0.07	0.06	0.04	0.02	0.01
		Moderate	0.05	0.06	0.05	0.04	0.02	0.01
		High	0.24	0.28	0.24	0.16	0.08	0.04
EUR-A	Male	Background	0.58	0.38	0.46	0.78	0.97	0.98
		Low	0.11	0.16	0.14	0.06	0.01	0.00
		Moderate	0.27	0.40	0.35	0.14	0.02	0.01
		High	0.04	0.06	0.05	0.02	0.00	0.00
	Female	Background	0.67	0.59	0.69	0.92	0.99	0.99
		Low	0.11	0.14	0.11	0.03	0.00	0.00
		Moderate	0.19	0.23	0.18	0.04	0.01	0.00
		High	0.03	0.04	0.03	0.01	0.00	0.00
EUR-B	Male	Background	0.36	0.15	0.29	0.64	0.80	0.90
		Low	0.10	0.13	0.11	0.06	0.03	0.01
		Moderate	0.33	0.45	0.37	0.19	0.10	0.05
		High	0.21	0.28	0.23	0.12	0.06	0.03
	Female	Background	0.51	0.32	0.48	0.80	0.89	0.95
		Low	0.05	0.07	0.05	0.02	0.01	0.01
		Moderate	0.20	0.27	0.21	0.08	0.04	0.02
		High	0.25	0.34	0.26	0.10	0.05	0.03
EUR-C	Male	Background	0.35	0.13	0.20	0.73	0.90	0.95
		Low	0.11	0.14	0.13	0.04	0.02	0.01
		Moderate	0.39	0.52	0.48	0.16	0.06	0.03
		High	0.15	0.20	0.19	0.06	0.02	0.01
	Female	Background	0.47	0.18	0.35	0.85	0.96	0.98
		Low	0.13	0.20	0.15	0.04	0.01	0.01
		Moderate	0.31	0.48	0.38	0.09	0.03	0.01
		High	0.10	0.15	0.12	0.03	0.01	0.00
SEAR-B	Male	Background	0.32	0.09	0.13	0.33	0.59	0.80
		Low	0.11	0.14	0.13	0.10	0.06	0.03
		Moderate	0.25	0.33	0.32	0.25	0.15	0.07
		High	0.33	0.44	0.42	0.32	0.20	0.10
	Female	Background	0.49	0.36	0.40	0.60	0.81	0.90
		Low	0.11	0.13	0.13	0.08	0.04	0.02
		Moderate	0.16	0.21	0.19	0.13	0.06	0.03
		High	0.24	0.30	0.28	0.19	0.09	0.05

continued

Table 21.48 Proportions of the working-age population occupationally exposed to different levels of ergonomic stressor, by sex and subregion (*continued*)

Subregion	Sex	Exposure category	Age group (years)					
			15–29	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
SEAR-D	Male	Background	0.29	0.10	0.13	0.34	0.51	0.76
		Low	0.06	0.08	0.08	0.06	0.04	0.02
		Moderate	0.23	0.29	0.28	0.21	0.16	0.08
		High	0.42	0.53	0.52	0.39	0.29	0.14
	Female	Background	0.57	0.45	0.52	0.69	0.85	0.92
		Low	0.01	0.02	0.01	0.01	0.00	0.00
		Moderate	0.06	0.08	0.07	0.04	0.02	0.01
		High	0.36	0.46	0.40	0.26	0.13	0.06
WPR-A	Male	Background	0.51	0.29	0.30	0.49	0.78	0.89
		Low	0.16	0.24	0.23	0.17	0.07	0.04
		Moderate	0.30	0.43	0.42	0.31	0.13	0.07
		High	0.03	0.04	0.04	0.03	0.01	0.01
	Female	Background	0.59	0.50	0.52	0.74	0.91	0.95
		Low	0.16	0.20	0.19	0.10	0.04	0.02
		Moderate	0.21	0.26	0.25	0.13	0.05	0.02
		High	0.03	0.04	0.04	0.02	0.01	0.00
WPR-B	Male	Background	0.27	0.11	0.17	0.45	0.74	0.87
		Low	0.13	0.16	0.15	0.10	0.05	0.02
		Moderate	0.25	0.30	0.28	0.19	0.09	0.04
		High	0.35	0.43	0.40	0.27	0.13	0.06
	Female	Background	0.30	0.20	0.39	0.74	0.92	0.96
		Low	0.17	0.20	0.15	0.07	0.02	0.01
		Moderate	0.22	0.25	0.19	0.08	0.03	0.01
		High	0.30	0.35	0.27	0.11	0.04	0.02

compensable work-related diseases in many countries, representing a third or more of all registered occupational diseases in the United States, the Nordic countries and Japan (Bernard 1997; Pope et al. 1991; Vaaranen et al. 1994).

Guo et al. (1995) estimated that 65% of cases of low back pain in the United States are attributable to occupational activities. To date, there have been no other published estimates of the fraction of back pain (specifically) in the total population that is occupationally induced. However, low back pain was identified by the Pan American Health Organization as one of the top three occupational health problems to be targeted by surveillance within the WHO Region of the Americas (Choi et al. 2001).

Among MSDs caused by occupational ergonomic stressors, only low back pain is currently a separate category in the GBD database and could be assessed. For purposes of this chapter, low back pain is defined as all pain in the back that is not secondary to another disease or injury cause (such as cancer or a motor vehicle accident). This includes disk prob-

lems (displacement, rupture), sciatica and other sources of back pain. Cervical spine problems, such as neck pain or neck torsion problems, are excluded. This category of conditions is considered equivalent to what others have termed non-traumatic MSDs affecting the lower back.

In the epidemiological literature, MSDs of the back are often defined on the basis of pain reported on interview, usually with standardized study criteria referring to time of onset, frequency and/or severity of pain. Physical examinations have sometimes been used to supplement questionnaires, particularly to help localize the symptoms reported on interview and to rule out other causes of those symptoms. However, an important proportion of epidemiologically relevant (exposure-related) back disorders are negative on physical examination (e.g. Punnett et al. 1991) as well as on X-ray (e.g. Riihimäki et al. 1990). Most cases of low back pain cannot be diagnosed by objective criteria and are typically designated idiopathic or non-specific (Frank et al. 1995; Riihimäki 1991, 1995b), even if there are findings on examination or severe symptoms and loss of function.

It is difficult to measure directly the validity of questionnaire responses, since no consensus exists regarding a single "gold standard" against which all other measurements could be compared. The sensitivity and reliability of physical examination manoeuvres for identifying low back pain range from good to poor; not all pain results from known mechanisms for which there is a corresponding objective test (Deyo et al. 1992; Viikari-Juntura and Riihimäki 1999). Deyo et al. (1992) suggested that as many as 85% of cases of low back pain cannot be diagnosed because of the poor performance of examination and imaging tests.

A recent review by NIOSH (Bernard 1997) (see also below) emphasized that health outcomes defined subjectively should be included in any consideration of work-related back disorders. In 24 of the 42 epidemiological studies on low back pain reviewed, the health outcome was defined by reported symptoms on questionnaires or interview, ranging from any back pain to specific symptoms such as those consistent with sciatica. In several studies, MSD cases defined by symptoms alone and those defined by findings on physical examination have shown very similar associations with the ergonomic characteristics of subjects' jobs. Symptom-based case definitions generally appear to be both unbiased and more sensitive than those that require documented abnormalities on physical examination (e.g. Bernard et al. 1993; Punnett 1998; Punnett et al. 1991; Silverstein et al. 1986, 1987).

Other case definitions sometimes used epidemiologically include low back impairment or disability, typically indicated by reduced ability to perform activities of daily living or occupational tasks, work absenteeism and the seeking of medical care for back pain. Such behavioural measures, however, are less desirable than low back pain *per se*, because they are more distal from the direct morbidity and more likely to be affected by interpersonal variability (e.g. tolerance of pain before seeking medical

attention) or by differences in job demands (e.g. pain of the same severity may cause more low back disability in persons whose jobs are more demanding).

At the same time, there is a strong correlation between the frequency of musculoskeletal symptoms by occupation and the frequency of workers' compensation claims and recorded work-related repetitive trauma disorders in those same occupations (Fine et al. 1986; Silverstein et al. 1997). Outcomes such as days of restricted activity, long-term disability, health care utilization and use of medication are very common among people with back pain, indicating the public health importance and cost of these disorders even when self-reported pain is not confirmed objectively (Badley et al. 1994, 1995; Guo et al. 1999; Miedema et al. 1998; Punnett 1999; Westgaard and Jansen 1992).

Back pain has been defined operationally in various ways in epidemiological studies, including both prevalent and incident conditions. Variations in the definition are related to the recall period (e.g. pain now, or within the last week or the past year), the frequency or duration (e.g. at least three times in the past year, or lasting at least one week) and the severity (e.g. at least a "4" on a 7-point pain intensity scale), among others. Even among studies that use similar definitions, prevalence estimates can vary substantially (Loney and Stratford 1999). However, for the purpose of evaluating the exposure–response relationship, as long as comparisons are made within a study population that has been evaluated with a consistent case definition, estimates of relative risk do not appear to be greatly affected (Ozguler et al. 2000).

Evidence of causality

The evidence on low MSDs, including back pain, in relation to workplace factors has been thoroughly reviewed by NIOSH (Bernard 1997). The National Research Council, with the Institute of Medicine, has also published a comprehensive review of the evidence on MSDs in the workplace (National Research Council 2001). Strong or sufficient evidence was found for a number of risk factors at the workplace to be associated with back pain (Table 21.49). The National Research Council report (2001) summarized ranges of risk estimates for specific occupational stressors (Table 21.50).

In addition to these two comprehensive reviews from the United States, numerous other authors from Europe, Asia and Canada reviewed the same epidemiological literature or variously defined subsets, and most reached similar conclusions (e.g. Burdorf and Sorock 1997; Frank et al. 1996; Garg 1992; Hagberg et al. 1995; Hales and Bernard 1996; Hoogendoorn et al. 1999; Hulshof et al. 1987; Jensen 1988; Jin et al. 2000; Johanning et al. 1991; Lagerström et al. 1998; Nachemson and Jonsson 2000; Riihimäki 1991, 1995a; Viikari-Juntura 1997; Wikstrom et al. 1994). For example, in a systematic literature review that focused on 28 cohort and three case–control studies of highest methodological

Table 21.49 Rating of evidence for causal relationships between specific occupational stressors and back disorders according to the NIOSH review

Strength of evidence	*Specific stressor*
Strong evidence[a]	Lifting and forceful movements Whole-body vibration
Evidence[b]	Awkward postures Heavy physical work
Insufficient evidence[c]	Static work postures
Evidence of no effect[d]	Other stressors

[a] *Strong evidence.* A causal relationship is shown to be very likely between intense or long-duration exposure to the specific risk factor(s) and MSD of the back when the epidemiological criteria of causality are used. A positive relationship has been observed between exposure to the specific risk factor and MSD of the back where chance, bias and confounding factors could be ruled out with reasonable confidence in at least several studies.

[b] *Evidence.* Some convincing epidemiological evidence shows a causal relationship when the epidemiological criteria for causality for intense or long-duration exposure to the specific risk factor(s) and MSD of the back are used. A positive relationship has been observed between exposure to the specific risk factor and MSD of the back in studies in which chance, bias and confounding factors are not the likely explanation.

[c] *Insufficient evidence.* The available studies are of insufficient number, quality, consistency or statistical power to permit a conclusion regarding the presence or absence of a causal association.

[d] *Evidence of no effect.* Adequate studies consistently show that the specific workplace risk factor is not related to development of MSD of the back.

Note: In this review, 42 epidemiological studies were selected on the basis of the following criteria: (i) exposed and reference populations were well defined; (ii) MSDs of the back were measured by well defined, explicit criteria determined before the study; (iii) exposure was evaluated so that some inference could be drawn regarding repetition, force, extreme joint position, static loading or vibration and lifting tasks; (iv) study participation of more than 70%. Thirty studies used a cross-sectional design and five a prospective cohort, four were case–control studies and two were retrospective cohorts. Full descriptions of these studies appear in Table 6-6 of the NIOSH review. Criteria for causality were based on strength of association, consistency, specificity of effect or association, temporality, exposure–response relationship and coherence of evidence.

Source: Bernard (1997).

quality, Hoogendoorn et al. (1999) found strong evidence for manual material handling, bending and twisting and whole-body vibration as risk factors for back pain. They found moderate evidence for patient handling and heavy physical work, and no evidence for standing or walking, sitting, sporting activities and total leisure-time physical activity. Specific psychological stressors, supported by weaker evidence, were reviewed by Burdorf and Sorock (1997), and are shown in Table 21.51.

Some of the results of these reviews on specific stressors were obtained from studies that were conducted within various occupational groups, such as operators, mine workers, dentists, office workers and nurses. The evidence presented above implies that preventive interventions reducing the exposure to these risk factors would decrease the occurrence of back disorders considerably, even within an occupation.

Table 21.50 Summary of epidemiological studies with risk estimates of null and positive associations of work-related risk factors and the occurrence of back disorders

	Risk estimate expressed in relative risk or odds ratio				*Attributable fraction (%)*	
	Null association[a]		*Positive association*			
Work-related risk factor	*N*[b]	*Range of odds ratio*	*N*[b]	*Range of odds ratio*	*N*[b]	*Range*
Manual material handling	4	0.90–1.45	24	1.12–3.54	17	11–66
Frequent bending and twisting	2	1.08–1.30	15	1.29–8.09	8	19–57
Heavy physical load	0	NA	8	1.54–3.71	5	31–58
Static work posture	3	0.80–0.97	3	1.30–3.29	3	14–32
Repetitive movements	2	0.98–1.20	1	1.97	1	41
Whole-body vibration	1	1.10	16	1.26–9.00	11	18–80

NA Not applicable.

[a] Confidence intervals of the risk estimates included the null estimate (1.0). In only 12 of 16 null associations was the magnitude of risk estimate presented.

[b] Number of associations presented in the epidemiological studies.

Source: National Research Council (2001).

Table 21.51 Summary of epidemiological studies on associations between specific occupational psychological risk factors and the occurrence of back disorders

	Risk estimate		*Attributable fraction*	
Specific risk factor	*N*[a]	*Range*	*N*[a]	*Range (%)*
Mental stress	4	1.30–2.08	4	23–44
Job dissatisfaction	5	1.39–2.40	4	21–41
Work pace	1	1.12	NA	NA
Monotonous work	5	1.25–2.34	4	20–44

NA Not applicable.

[a] Number of associations presented in the epidemiological studies.

Source: Burdorf and Sorrock (1997).

Furthermore, there is sufficient evidence on biomechanical risk factors to conclude that many cases of low back pain could be prevented by workplace changes. For example, Marras et al. (2000) showed that lifting equipment and other engineering controls had demonstrable effects on lowering both biomechanical exposure to the lower back (compressive force, torque, etc.) and reported rates of low back disorders in 36 repetitive manual material handling jobs at 16 different companies.

The reviews by the National Research Council (2001), Westgaard and Winkel (1997) and Frank et al. (1996) each cited a number of well designed studies that identified opportunities to prevent risk of low back pain by reducing exposure to biophysical and psychosocial factors. To illustrate the improvements that can be obtained by reducing ergonomic stressors at work, selected interventions and their impact are shown in Table 21.52. Certain interventions have practically completely removed ergonomic stressors from the workplace.

Despite this extensive literature, some still dispute the importance of these factors, especially in relation to nonoccupational causes (e.g. Battie and Bigos 1991; Nachemson 1999; Waddell 1991). There are probably several interrelated reasons for the continuing controversy, many of which have been discussed by others (Frank et al. 1995, 1996; Viikari-Juntura and Riihimäki 1999). The available epidemiological evidence still consists largely of cross-sectional and retrospective case–control investigations. With regard to assessment of morbidity, the use of self-reported symptoms for end-points has also generated discussion, as described above. Sparse longitudinal data leave important gaps in knowledge concerning latency of effect, natural history, prognosis and potential for selection bias in employed populations (e.g. the "healthy worker effect"). Few studies have attempted systematically to describe these features of back disorders in populations with defined levels of exposure to ergonomic stressors. Specific, quantitative comparisons of conclusions based on prevalence and incidence data within the same population are rare, and knowledge is still sparse as to the factors that predict recovery or persistence among workers who continue in their jobs after the onset of a disorder. No study has been identified that compares current and former workers with reference to both prior MSD morbidity and exposure status.

In addition, there are many known or suspected nonoccupational risk factors; some study populations have not provided enough statistical power to address potential confounding and effect modification of exposure–response relationships. Exposure assessment has often been limited, with too few exposures characterized to rule out confounding, or the use of crude exposure indicators leading to potential misclassification and unreliable conclusions. Attempts to partition risk between physical and psychosocial domains have obscured the overlap between the two and the distinction between preventable and nonpreventable risk factors (Bongers et al. 1993; MacDonald et al. 2001; Volinn and Punnett 2001; Volinn et al. 2001).

In the light of these issues, it is important to restate that the issue is not whether all back disorders are caused by work; there is a clear consensus that this is not the case. Nevertheless, most investigators and reviewers have concluded, equally clearly, that a large proportion of back disorders *among persons with high exposure to ergonomic stressors at work* could be prevented by reducing those exposures.

Table 21.52 Selected intervention studies on occupational ergonomic stressors and reduction of occurrence of low back pain

Job title/activity	*Intervention: engineering/administrative control*	*Outcome*	*Study*[a]
Nursing assistants (in two nursing home units)	Introduction of walking belts, mechanical hoists and shower chairs	Back injury incidence reduced from 83 to 42 per 100 full-time equivalents	Garg and Owen (1994) [Ex. 26-1415]
Poultry processing employees	Introduction of workstation analysis and redesign, including altering heights of products, providing workstands, and installing tank tilters to reduce manual handling	Back injury rates declined from 4.4 to 3.0 per 200000 hours	Stuart-Buttle (1994) [Ex. 26-1045]
Office furniture manufacturing assembly workers	Scissor lifts installed to aid in packaging file cabinets of different sizes; small-assembly workstations altered to eliminate twisting and bending during lifting	Back injuries cut by 50%	LaBar (1991) [Ex. 26-1078]
Metal castings, unpacking operation	Crates modified by adding gates at each end and installing scissors lift to lift crates; changes made in the way castings were stacked in the crates to permit workers' arms to remain close to the body while unpacking	Back injuries associated with this operation eliminated	Oxenburgh (1994) [Ex. 26-1041]
Palletizing operation	Scissors lift table with turntable tops installed alongside each packing station	Five out of six back injuries eliminated	Benson (1987) [Ex. 26-1062]
Lamp manufacturing	Addition of vacuum hoist; reduced equipment height; reduced box size and weight; introduction of back awareness programme for employees	Back and upper extremity disorders eliminated in the last four years	Carreau and Bessett (1991) [Ex. 26-1071]
Unpacking car parts	Plywood sheets modified to reduce weight and permit them to slide more easily in the grooves	Back injuries associated with this operation eliminated	Oxenburgh (1991) [Ex. 26-1041]

Manual handling of bulk paper	Manual lifting eliminated by installation of scissors lift; trolley runners replaced by roller bearings that enable the paper to be loaded onto the scissors lift without manual lifting	No back injuries in the three years since modifications made	Oxenburgh (1991) [Ex. 26-1041]
Railway repairmen	Storage of tools and materials off the ground between knee and shoulder height; winches to lift and handle heavy equipment; redesigned work tables, trolleys and carts to more easily handle train car parts	Low back injuries and lost work days eliminated	McMahan (1991) [Ex. 26-1083]
Nursing, hospital	Changes in procedures for moving patients, manoeuvring carts and equipment, using gall bladder boards, walking on wet floors, accessing power outlets	Back injury rates reduced by 25% in 18 months since programme implemented	Garg and Dockery (1995) [Ex. 26-1095]
Video display terminal operators	Reduction of keying to no more than five hours per day and evaluation of new chairs; new chairs installed in February 1991	Low back pain reduced from 8.3% to zero between November 1990 and February 1993	[Ex. 500-41-115]
Forestry workers	Forest tractor seats: spring tension and inclination adjustment; accessory lumbar support provided	Reduction of low back pain among forest tractor drivers: 24 of 50 drivers found the lumbar support beneficial	Perkio-Makela and Riihimäki (1997)

[a] All exhibit numbers refer to materials in USDOL OSHA (2000).

Source: Lahiri et al. (2002).

Other outcomes

Although the present analysis was limited to low back pain, the evidence on MSDs caused by occupational ergonomic stressors is broader. MSDs affecting the neck and the upper and lower limbs result from the same risk factors as are implicated in low back pain. For example, in a study of over 10 000 manufacturing employees, the effect of "greater physical demands" of the job on low-back musculoskeletal injuries (relative risk of 1.6, 95% CI 1.2–2.1) was only slightly higher than that for all other musculoskeletal injuries combined (relative risk of 1.4, 95% CI 1.1–1.7) (Tsai et al. 1992). De Zwart et al. (1997), studying over 7300 men in the Netherlands, found higher prevalences of shoulder disorders among employees in heavy physical work (e.g. heavy lifting and frequent stooping) and steeper increases over four years than among employees in less physically demanding jobs. The magnitude of these effects was very similar to those for low back injuries. The work-relatedness of upper and lower extremity MSDs has been discussed extensively, again by European as well as North American reviewers (e.g. Armstrong et al. 1993; Bernard 1997; Bongers et al. 1993; Buckle and Devereaux 1999; Hagberg et al. 1995; National Research Council 2001; Sluiter et al. 2000).

Also excluded here are other types of health effects related to ergonomic stressors, such as acute workplace injuries, cardiovascular disease, mental health and adverse reproductive effects (Punnett 2002). Thus, the total impact of excessively strenuous work activities on morbidity and related quality of life is greater than that estimated in this risk assessment.

Hazard estimates

Data sources

For the purposes of this analysis, studies were sought that compared the risk of low back pain among broad occupational groups (defined similarly to the economic subsectors explained above) and comprehensively enough to cover the range of paid occupations. Smaller, more specific studies limited to relatively narrow occupational groups (e.g. nurses, dock workers, drivers) were checked for consistency with the larger data sets. Studies where the reference groups were also engaged in substantial physical activity (e.g. house painters) were excluded. The most recent literature (1997–2001) was searched for exposure data and exposure–risk relationships, and published statistics of national occupational health and safety institutes were consulted.

In addition to this systematic search, a number of reviews and studies were identified to provide evidence supporting the selected approach. Search strategies were described in the Introduction. Medline was searched for articles more recent than 1985, using any of the keywords back pain, back disorder, back or musculoskeletal combined with any of

the following: occupation, occupational, workplace, work, workers, risk factors, risk.

Description of literature

The studies specifically referred to in this section are summarized in Table 21.53.

The report of the National Research Council (2001) stated that the *occupations* with the highest risk among men were construction labourers, carpenters and industrial truck and tractor equipment operators, while among women they were nursing aides/orderlies/attendants, licensed practical nurses, maids and janitor/cleaners. Other high-risk occupations were hairdressers and automobile mechanics, often employed in small businesses or self-employed. No rates were listed against occupations in the report. The report stated that the highest-risk *industries* for men were lumber and building material retailing, crude petroleum and natural gas extraction and sawmills/millwork. Among women, the highest-risk industries were nursing and personal care facilities, beauty shops and motor vehicle equipment and manufacturing. No rates were listed for industries in the report.

Leigh and Sheetz (1989) measured low back pain on the basis of a national survey and a self-reported statement regarding "trouble with back or pain during the last year". They estimated relative risks by comparing the outcome frequency among occupational groups, using managers as a reference group (Table 21.54). This chapter places great weight on this study, because it was relatively large (1404 participants) compared to many others, it covered a comprehensive sample of occupations, and the results were adjusted for potential confounding variables. One important limitation, however, is that the multivariate analyses simultaneously included two ratings of physical exposure, socioeconomic status and occupational title. Since physical exposure is hypothesized to be the primary pathway through which occupational differences are manifested, these analyses would certainly lead to an underestimation of the effect of occupation on MSDs.

Although operators and service workers have very similar relative risks, it is common that intervention strategies differ among these occupational settings. For that reason, the relative risks and exposure assessments for those two occupational groups are presented separately throughout this analysis.

Within the limits of the available literature, the results of the Leigh and Sheetz analyses appear to be generally consistent with other reported relative risk values (Table 21.55). Since many other studies used office workers or other sedentary occupations as the reference group, it is appropriate to adjust the Leigh and Sheetz findings for comparative purposes. This can be done by dividing the relative risks for categories 3, 4, and 5 by 1.38 (the relative risk for clerical or sales work), in order to estimate a relative risk with clerical jobs as the reference group. The new

Table 21.53 Key studies and reviews on work-related back pain

Study population (source) or literature reviewed	*Population size*	*Outcome measured/reported*	*Magnitude or relative risk*	*Comments*	*Reference*
Reviews					
Back pain in relation to heavy physical work, lifting and forceful movements, non-neutral postures, and/or whole-body vibration: Belgium, Finland, Italy, Japan, Netherlands, Russian Federation, Sweden, USA	42 studies, evaluated according to 4 criteria for methodological quality	Back pain (multiple definitions among the studies)	Heavy physical work: [a] 1.0–12.1 Lifting and forceful movements: [a] 0.9–10.7 Non-neutral postures: 1.2–10.7 Static work postures: 0.8–24.6 Whole-body vibration: 1.0–39.5 Conclusions: "strong evidence" for causal relationship with lifting and whole-body vibration; "evidence" for causal relationship with awkward posture and heavy physical work	Literature review: etiological studies	Bernard (1997)
Important risk factors for work-related back disorders, and strength of the association between the consistent risk factors and back disorder were identified	35 epidemiological studies published from 1980 to 1996	Back disorders (multiple definitions among the studies)	Manual materials handling: 1.12–3.07 Frequent bending and twisting: 1.29–8.09 Heavy physical load: 1.54–3.71 Static work movement: 1.30–3.29 Repetitive movement: 1.97 Whole-body vibration: 1.47–9.00 Mental stress: 1.30–2.08 Job satisfaction: 1.39–2.40 Pace of work: 1.21 Job decision latitude or monotonous work: 1.25–2.34 Conclusion: lifting or carrying loads, whole-body vibration and frequent bending and twisting consistently associated with work-related back disorder. Job dissatisfaction and low job decision latitude also important	Illustration on individual risk factors, such as age, smoking and education	Burdorf and Sorock (1997)

Intensive and systematic review on physical load during work or leisure time as risk factors for back pain. Literature sources: Medline (1966–1997), Embase (1988–1997), Psyclit (1974–1997), NIOSHTIC, CISDOC and HSELINE (1977–1997) and Sportdiscus (1949–1997)	31 publications in Dutch, English, French and German	Inclusion criteria (health outcome): back pain based on symptoms or signs of non-specific back pain, self-reported or measured (via sick leave, medical consultation, treatment and disability due to back pain, etc.)	Strong evidence for manual material handling, bending and twisting, and whole-body vibration as risk factors. Moderate evidence for patient handling and heavy physical work. No evidence for standing or walking, sitting, sports and total leisure-time physical activity	Cross-sectional studies were excluded	Hoogendoorn et al. (1999)
Review of studies of work-related low back pain in China among literature issued from 1983 to 1997	16 epidemiological studies selected; quality inclusion/ exclusion criteria were applied; all studies included were cross-sectional	Occurrence of work-related low back pain (multiple definitions among the studies)	*Prevalence ratios:* Bending and twisting: 2.0–8.5 Static posture: 1.5–14.3 Whole-body vibration: 1.9–5.5	Suggestion of potentially greater exposure in China's work environment and potential systematic bias	Jin et al. (2000)
		Back pain (multiple definitions among the studies), muscle activity, tissue load and tissue damage	"linkages … between external loads and biomechanical loading of the spine, biomechanical loading and internal tolerances of the spine, and internal tolerances and outcomes (from pain through disability) are well established…The literature relating to causal factors in work-related low back disorders is coherent and provides ample evidence on how adverse work situations can lead to them." (pp. 357–358)	Literature review: etiological, experimental and intervention studies	National Research Council (2001)
Systematic literature review of articles published in English, 1980–1995, 25 most populous countries	NA	Back pain (multiple definitions among the studies); point prevalence of low back pain/ annual or lifetime incidence	2–4 times higher rates of back pain in high-income than in low-income countries. Within low-income countries, rates are higher among urban than among rural populations	Comparison of low-income and high-income countries	Volinn (1997)

continued

Table 21.53 Key studies and reviews on work-related back pain (*continued*)

Study population (source) or literature reviewed	*Population size*	*Outcome measured/reported*	*Magnitude or relative risk*	*Comments*	*Reference*
Original studies					
Current full-time employees at one site of large chemical manufacturing company, USA, 1987–1989	5 903 employees	Back or joint pain; back pain >30 days; visit to physician for back pain	35.4% reported back or joint pain during the past year, 5.3% back pain lasting >30 days; managers and technicians had highest prevalences of back pain >30 days	Adjusted for age, sex and ethnicity	Burchfiel et al. (1992)
Random sample of retired workers living in the Paris area, members of a supplementary interprofessional retirement pension fund	993 people in total (626 in first survey in 1982–1983, 464 in second survey in 1987–1988)	Osteoarticular disorders: presence of pain with/ without restricted joint movement for at least 6 months before interview	Increased frequency of osteoarticular pain during the 5 years between the two interviews; from 52% to 65% in men and 72% to 82% in women A significance increase in frequency of osteoarticular pain for those exposed to heavy weights and awkward postures (from 68% to 77% and 56% to 76%, respectively)	Significantly higher frequencies of osteoarticular pain for women than men in both interviews	Derriennic et al. (1993)
Probability sample of total working population in USA, 1988	30 074 workers	Back pain every day for a week or more during 12 months before interview	National estimate of back pain: 22.4 million cases and a prevalence of 17.6% in 1988 65% of cases were attributable to occupational activities The risk was highest for construction labourers among males (prevalence 22.6%) and nursing aides among females (18.8%)		Guo et al. (1995, 1999)
Random sample: all small and medium-sized factories listed with Labour Department in Delhi, India	60 of 6 076 factories; 631 workers selected randomly from 60 factories	Lumbar pain diagnosed by medical practitioner	Buffing workers, operators and assembly workers had highest pain prevalences (30.4%, 28.7% and 27.2%, respectively)		Joshi et al. (2001)

Workers' compensation claims for back injuries in 26 states (among 1 705 674 workers' compensation claims), all workers employed in 26 states, USA, 1979	329 474 claims for back injuries	Compensation claims filed for back injury or strains/sprains	Back injury constitutes 19.3% of compensation claims in 26 states Construction and mining industries have largest incidence: 1.6 and 1.5 claims/100 workers, respectively Occupations with the largest incidence compared to total workers: miscellaneous labourers 12.3 claims/100 workers, garbage collectors 11.1 claims/100 workers	Percentage of workers employed in 26 states in 1979, obtained from 1970 census	Klein et al. (1984)
Diagnosed low back pain cases, among all employed persons in Japan, 1986 and 1988	13 166 low back pain cases	Accidental low back pain caused by the action of a sudden force; required absence of ≥4 days	Transport occupations: 8 cases per 10 000 workers Construction: 4 cases per 10 000 workers Sales/service: 0.5 cases per 10 000 workers Mining and cargo handling: 13.9 and 12.3 cases per 10 000 workers	National survey data	Kuwashima et al. (1997)
Probability sample of United States workers employed ≥20 hours a week, 1972–1973	959 working males and 455 working females	Low back pain: experience of trouble with back or spine in past year	1-year low back pain past prevalence: 19.4% males, 20.7% females Relative risk: farmers 5.17, operators 2.39, service 2.67, clerical 1.38 compared to managers and professionals Relative risk for high job demands: 1.68	Health outcome did not distinguish between upper and lower back pain Over-adjustment by including socioeconomic status, occupation and physical risk factors in same multivariate analyses	Leigh and Sheetz (1989)
Three random samples of the Finnish population, 20–64 years of age, 1988–1990	7 544 people	Any back pain in past 30 days/back pain in past 12 months	Odds ratios of back pain: 2.1 in farmers, 1.8 in manual workers, 1.7 in entrepreneurs, 1.4 in lower white-collar workers	Results adjusted for age, height, marital status, smoking, body mass index, physical activity, mental distress	Leino-Arjas et al. (1998)
Workers at 8 aluminium plants in Norway, 1998	5 654 people	One-year prevalence of musculoskeletal symptoms by Standardized Nordic Questionnaire	Odds ratio for low back symptoms among operator vs office workers: 1.8 (range 1.5–2.1)	Adjusted for sex, weight, height, smoking, and nonoccupational physical activity	Morken et al. (2000)

continued

Table 21.53 Key studies and reviews on work-related back pain (*continued*)

Study population (source) or literature reviewed	*Population size*	*Outcome measured/reported*	*Magnitude or relative risk*	*Comments*	*Reference*
Active workers from 4 occupational sectors (office, hospital, warehouse, airport registration), France, 1991	725 people	Six definitions of low back pain (pain or discomfort in a lumbar area in the previous 6 months/pain at least one day/pain >30 days/intensity of pain above 3/physician visit/ treatment for low back pain)	Prevalence of low back pain varied from 8% to 45% according to the case definition Carrying heavy loads and bending postures showed consistently high odds ratios (1.88–2.14) for most low back pain definitions	Over-adjustment by including both occupation and physical risk factors in multivariate analyses; also adjusted for sex, age and body mass index	Ozguler et al. (2000)
All back disorders reported to clinic in one automobile assembly plant (referents from same production departments)	95 assembly workers with back disorders (cases) and 124 without (referents)	Back cases: workers who sought medical attention for "new" back disorders during a 10-month period	Exposed workers: 84% Bending and twisting (100% vs 0%): odds ratio 8.09 (range 1.5–44.0) Lifting (>44.5 Newtons/min): odds ratio 2.2 (range 1.0–4.7)	Adjusted for age, sex and sports activity	Punnett et al. (1991)
Employed persons in Washington State, USA, 1990–1998	Approximately 1.23 million full-time equivalent workers per year	Workers' compensation claims for nontraumatic soft tissue disorders of the back	Rates by industry sector ranged from 43.5 to 280.0 per 10 000 full-time equivalents	Surveillance data	Silverstein et al. (2002)
Population-based survey of approx 9.9 million adults (15 years or older), Belgium, 1991	3 829 people	Reported symptoms: low back pain, history of low back pain, first low back pain and daily low back pain	Current low back pain: 33% of population Work dissatisfaction associated with low back pain history (odds ratio >2.4)	Adjusted for age, sex, language, residence, social class and job satisfaction	Skovron et al. (1994)
Full-time regular workers in Shell Oil Company's manufacturing facilities between 1987 and 1989, USA	10 350 people	Low-back injury (ICD-9 CM, 722, 724)	Physically demanding jobs have relative risk of 1.57 for low back injury and 1.35 for non-low back musculoskeletal injury Smoking and overweight showed high relative risks (1.54 and 1.42, respectively)	Job title used to identify potential for increased physical demand at work	Tsai et al. (1992)

NA Not applicable.

[a] Range of effects for studies that met at least one criterion for epidemiological quality.

Table 21.54 Relative risks of low back pain for occupational groups, with managers and professionals as the reference group

Occupational activity	*Relative risk (95% CI)*
Managers and professionals	1.0 (NA)
Clerical or sales worker	1.38 (0.85–2.25)
Operators	2.39 (1.09–5.25)
Service workers	2.67 (1.26–5.69)
Farmers	5.17 (1.57–17.0)

NA Not applicable.

Source: Based on data from Leigh and Sheetz (1989).

values would be 1.73, 1.93 and 3.75, respectively. Keeping in mind that these estimates represent the average values for the entire occupational category, it can be seen that the other studies cited fall within the CIs, with very few exceptions, and in fact generally have similar point estimates (Table 21.55).

The only study that can directly and numerically be compared to that of Leigh and Sheetz (1989) is that by Leino-Arjas et al. (1998). However, the only value corresponding directly to one of the categories of Leigh and Sheetz is that for farmers. The relative risk is lower (2.13) than the one put forward by Leigh and Sheetz (5.17), which may reflect better working conditions for farmers in Finland. For this analysis we therefore used the average of these two results (see below).

Also available are administrative statistics from several countries on the number of cases of back conditions. These are generally compiled from employers' surveys or compensation statistics and typically report lower rates than those assessed by population surveys. Rates for certain occupations, as compared to managers and professionals, can be estimated on the basis of these statistics. Table 21.56 summarizes administrative workplace statistics on conditions involving the back, based on reports by employers in the United States of work-related injuries (Bureau of Labor Statistics 2001) and compensation statistics of the Australian workforce (National Occupational Health and Safety Commission 2001) and of the German national workforce (Bundesverband der Betriebskrankenkassen 2001).

All three of these data sets show higher risks for occupations other than managers and professionals, although the point estimates vary somewhat. None of these frequency estimates is adjusted for potential confounding variables. The incidents assessed in the first two data sets are limited to cases that have been recognized as work-related cases and involve behaviour such as absence from work or filing a claim against the employer. In contrast, the German study sought to assess the health status of the population more comprehensively and these data are

Table 21.55 Relative risks of occupational groups by exposure level

	Source							
Occupation (exposure category)	*Leigh and Sheetz (1989)*[a]	*Astrand (1987)*[b]	*Bongers et al. (1990)*[b]	*Bovenzi and Betta (1994)*[b]	*Burdorf et al. (1993)*[b]	*Hildebrandt (1995)*[b]	*Johanning et al. (1991)*[b]	*Magnusson et al. (1996)*[b]
Managers and professionals	1.00/NA							
Professionals								
Managers								
Teachers								
Clerical or sales workers	1.38/1.00							
Office workers (sedentary)				1.00	1.00	1.00	1.00	1.00
Clerks		1.00						
Air force officers			1.00					
Civil servants								
Sales								
Operators	2.39/1.73						3.90	
Construction labourers								
Manual workers		2.28						
Pilots and aircrew			9.00					
Drivers (bus, truck, tractor)				1.83–5.49	2.51	1.32		1.55–2.10
Crane operators					3.29			
Dock workers								
Plumbers						1.32		
Carpenters								
Technicians								
Assembly, packing, food processing								
Automobile mechanics								
Maintenance								
Service workers	2.67/1.93							
Airport registration workers								
Hospital workers								
Warehouse workers								
Stock handlers, baggers								
Janitors, cleaners								
Waitresses								
Nurses								
Farmers	5.17/3.75							

NA Not applicable.

[a] Relative risks used in estimation of global burden of disease. The second set of relative risk values was estimated using clerical/sales jobs as the reference group, for the purpose of comparison with other studies.

Source									
Partridge and Duthie (1968)[b]	*Riihimäki et al. (1989)*[b]	*Riihimäki et al. (1994)*[b]	*Videman et al. (1990)*[b]	*Burchfiel et al. (1992)*	*Ozguler et al. (2000)*	*Joshi et al. (2001)*	*Guo et al. (1995)*[c] *[female]*	*Morken et al. (2000)*	*Leino-Arjas et al. (1998) (male)*
									1.00
				1.00					
				1.80					
							[1.2]		
	1.00	1.00	1.00	0.89	1.00	1.00		1.00	1.35
1.00									
				1.10					
	1.0–1.5	1.40		1.10		1.83		1.80	
							2.10		
			3.60	1.49					1.84
			2.90				2.00		
1.27									
							1.70		
		1.50					2.10		
				1.20		1.59			
						1.73			
							1.80		
						1.59	1.70		
				1.03					
					0.86				
					1.13				
					0.54				
							1.70		
							[2.0]		
							[1.6]		
							[1.5]		
							1.80		2.13

[b] Cited in Bernard (1997).

[c] Compared to all male or female workers.

Table 21.56 Relative risks of occupational conditions involving the back by occupational title, compared to managers and professionals, from national surveillance data

	Relative risk for back conditions		
Occupational activity[a]	*USA*[b]	*Australia*[c]	*Germany*[d]
Managers and professionals	1.0	1.0	1.0
Tradespeople	—	5.5	—
Clerks	—	1.1	1.5
Technical, sales and administrative support	2.2	—	—
Sales and service workers	—	2.2	2.9
Service workers	7.4	—	—
Operators	9.1	—	2.4
Farmers, fishermen and forestry workers	4.3	—	3.6
Operators and farmers	—	8.8	—

— No data.

[a] Owing to the different reporting schemes, some rows (occupational activities) represent the sum of several other rows.

[b] Bureau of Labor Statistics (2001), nonfatal occupational injuries and illnesses involving days away from work, for injuries involving the back.

[c] National Occupational Health and Safety Commission (2001), conditions affecting the upper and lower back.

[d] Bundesverband der Betriebskrankenkassen (2001), musculoskeletal illnesses of the lower back.

Table 21.57 Exposure categories and relative risks of low back pain for occupational groups selected for this analysis, with managers and professionals as the reference group

Exposure category	*Relative risk*	*95% CI*	*Occupational activity*
Background	1.0	NA	Managers and professionals
Low	1.38	0.85–2.25	Clerical and sales workers
Moderate	2.53	1.09–5.69	Operators and service workers
High	3.65	1.57–17.0	Farmers

NA Not applicable.

Source: Exposure level adapted from Leigh and Sheetz (1989).

therefore likely to be more comparable to those reported by Leigh and Sheetz. The values are, in fact, relatively close except for agricultural workers.

Given that the study by Leigh and Sheetz (1989) best fits the format required for this analysis, and the supporting evidence displaying very similar quantitative values, the proposed exposure categories and attributed relative risks are displayed in Table 21.57. The value for farmers is provided by an average of the relative risks for farmers in the Leigh and

Sheetz (5.17) and Leino-Arjas et al. (2.13) studies, resulting in a relative risk of 3.65. The CI, however, remains the same, because the CI from the Leigh and Sheetz study (1.57–17.0) includes the CI provided by Leino-Arjas et al. (1.6–2.9) and is wider, which is probably a truer representation of the statistical uncertainty of this estimate.

Methodological quality of the literature

Many of the reviews cited above used systematic criteria to evaluate the potential for selection bias, information bias and confounding in the individual investigations. Several of them identified the methodologically stronger studies and relied primarily or exclusively on those to draw conclusions about the strength of the evidence.

Potential confounding by nonoccupational factors such as sex, age, anthropometry, smoking, heredity and general medical history was extensively investigated in the great majority of studies cited above. All of the studies on which NIOSH relied most heavily, as being rigorous and methodologically sound, controlled for multiple potential confounding variables, permitting the conclusion that physical job factors cause low back pain independently of other factors. Burdorf and Sorock's (1997) review also summarized the associations between low back pain and specific occupational exposures, and relied more heavily on data with adjustment for important covariates. For example, Smedley et al. (1995) adjusted for age, height and nonmusculoskeletal symptoms (the only nonoccupational factors associated with low back pain—see below) in their analysis of low back pain and patient handling demands among female nurses. Tsai et al. (1992) examined the effect of greater vs less physical demands in the job, adjusting for six nonoccupational covariates.

Leigh and Sheetz (1989) adjusted for sex, race, education, height and smoking. In addition, they included terms for occupation *and* for physical effort and repetitive work; this means that the effect of occupation is likely to be underestimated, since the primary intermediate variable (physical effort) was also included. There is also a great deal of discussion in the epidemiological literature about the mechanisms of the effect of socioeconomic status (see below). It could easily be argued that the inclusion of terms for education also results in overadjustment, since a lower level of education is strongly associated with employment in "unskilled" jobs with higher physical exposures and is likely to act at least in part through such limited job opportunities.

Ozguler et al. (2000) analysed multiple low back pain case definitions. The same set of covariates was examined for each one, and all those nonoccupational factors (sex, age, obesity, psychosomatic "well-being") that were associated with low back pain were kept in the model. Like that of Leigh and Sheetz, this study overadjusted the estimates for occupation, because exposure variables such as carrying heavy loads and bending posture were entered in addition to the occupation indicators.

Many investigators have treated socioeconomic status and sex as potential confounding variables that require adjustment in statistical analysis of MSD etiology. However, to the extent that these factors act through or are surrogates for working conditions, both physical and psychosocial, such analyses may in fact serve to obscure the role of those exposures. Both the incidence and the severity of low back pain show an inverse gradient with socioeconomic status (blue collar vs white collar jobs, income, education level) in both men and women (Bergenudd and Nilsson 1988; Broersen et al. 1996; Heistaro et al. 1998).

It seems highly plausible that a large part of the gradient of socioeconomic status in MSDs is due to differences in the work performed, since jobs with lower socioeconomic status consistently involve more physically strenuous and repetitive work (Behrens et al. 1994; Hollman et al. 1999). In a large study of metal working employees, psychosocial conditions at work and physical load were generally correlated with each other and were worse for blue-collar than for white-collar employees, as well as for women compared with men (Leino and Hänninen 1995). In each of these subgroups, adverse working conditions predicted the development or worsening of MSDs over a 10-year follow-up period. The effect of social class was not explained by "lifestyle" factors such as smoking, leisure-time physical activity, body mass index, alcohol consumption or marital status (Leino-Arjas et al. 1998).

Sex differences

Sex is also often described as a "risk factor" for MSDs. In the great majority of studies relied on here, either the population was restricted to one sex or relative risks were adjusted for sex. However, in most studies of low back pain the prevalence was the same or only slightly higher in men than in women (e.g. Behrens et al. 1994; Guo et al. 1995; Morken et al. 2000; Tsai et al. 1992). Skovron et al. (1994) found a higher prevalence among men than among women aged 20–49 years, whereas from 50 years of age the prevalence in women gradually increased relative to that in men. Thus sex is not a strong risk factor for low back pain in any case, and confounding of these effect estimates is not of concern.

Women and men typically experience qualitatively and quantitatively different working conditions (Punnett and Herbert 2000). Women are overrepresented in "light", monotonous jobs that require precise, repetitive hand motions with less latitude for decision-making. Men are more often found in jobs with heavy whole-body workload, such as manual materials handling. In general, once job assignments and the consequent occupational exposures are taken into account, sex differences become negligible (Punnett and Herbert 2000).

Table 21.58 Prevalence and attributable risk of joint pain at 6 and 11 years after retirement, among workers with prior exposure to "heavy physical work"

	Interview 1 (after 6 years)	*Interview 2 (after 11 years)*
Men		
Exposed	58%	75%
Unexposed	49%	57%
Attributable risk	*15.5%*	*24.0%*
Women		
Exposed	87%	91%
Unexposed	68%	79%
Attributable risk	*21.8%*	*13.2%*

Source: Derriennic et al. (1993).

Estimates of risk reversibility

Although no explicit studies have been carried out on low back pain attributed to occupational factors before retirement, it has been assumed that leaving the job would reduce the risk of back pain. The burden of work-related back pain would diminish gradually once the theoretical minimum exposure was reached. Since the theoretical minimum is zero, no new cases would arise. However, morbidity from past exposure might persist or worsen after retirement (Derriennic et al. 1993; Holte et al. 2000; Sobti et al. 1997).

Derriennic et al. (1993) defined a closed cohort of retirees from mixed occupations in France, with an average elapsed period of six years from retirement (at age 63) to the baseline survey. A follow-up survey was conducted after five years. Joint pain was reported by 29% of men and 42% of women at baseline. At 11 years post-retirement, the attributable risk was even higher among men, although it decreased in women because of the high prevalence of joint pain among unexposed persons (Table 21.58).

In summary, there are few or no epidemiological data on whether or not new back disorders develop after leaving work that can be attributed to ergonomic stressors in previously high-exposed (vs low-exposed) workers. Thus, we have assumed the work-related incidence to be zero after retirement from paid employment. However, we do have the high impact of interventions on exposed workers, which supports a reversibility of 100%. This means that the incidence of low back pain is zero for chronic and acute cases of low back pain after exposure ceases. However, chronic low back pain will continue (i.e. the incidence will be zero, but those who have already developed it will continue to experience it).

Extrapolation of risk factor–disease relationships from one subregion to another

Because occupational group was used as an indicator of the average level of combined risk factors for low back pain within each occupation, differences in distribution of risk factors that might exist within occupations or between countries are an important consideration. Risk ratios among occupations vary somewhat from one country to another. This could be due to differences in distributions of risk factors for low back pain, or regional or cultural divergences in symptom reporting. These discrepancies become even more difficult to interpret when the comparisons are made between developing and developed countries. Unfortunately, the data are sparse regarding cross-national differences, both in exposure distributions (within similar types of job) and in reporting of low back pain.

One important element is the extent to which ergonomic interventions have been implemented in the various countries or regions. Although there are insufficient data to quantify the extent of effective ergonomic programmes in each region, it is generally true that occupational health and safety legislation, enforcement and adaptation of engineering controls (ergonomic changes) tend to be more widespread in developed countries, especially in northern Europe followed by North America. If this is correct, then application of occupation-specific relative risks from developed countries (e.g. Leigh and Sheetz 1989) to developing countries would produce conservative estimates.

Similarly, it would be easy to assume that, because of mechanization and other changes in production technology, more developed countries would typically have fewer ergonomic stressors in the same type of work than developing countries, even without intention to reduce ergonomic stressors. For example, Bao et al. (1997) compared shoulder-neck ergonomic exposures in a Chinese and a Swedish assembly line workplace. The Swedish workplace had a better ergonomic workstation design and was better balanced, as well as less sensitive to production irregularities, than the Chinese workplace. The Swedish operators were less exposed to awkward postures during work.

However, in contrast to the general assumption that low back pain rates should be higher in low-income than in high-income countries, a systematic review by Volinn (1997) showed 2–4 times higher rates among Belgian, German and Swedish general populations than among southern Chinese, Philippine, Indonesian and Nigerian farmers. Mentioning that the prevalence of low back pain is higher in the urban populations of low-income countries, and sharply higher in enclosed workshops in low-income countries compared to low-income rural populations, Volinn suggested that low back pain might be associated with urbanization and rapid industrialization, which imply more repetitive movements and loss of control over work pace and scheduling. The author noted that interpretation of the findings requires consideration of

Table 21.59 Comparison of ranges of effect estimates for selected risk factors for low back pain in some working populations of China, India and the Russian Federation

	China, India, Russian Federation		*Developed countries*[a]	
Risk factor	*Studies (n)*	*POR range*	*Studies (n)*	*POR range*
Bending and twisting[b]	4	3.1–16.5	9	1.3–8.1
Static posture[c]	5	2.0–19.9	3	1.3–3.3
Whole-body vibration[a]	4	2.5–14.2	14	1.5–9.0
Heavy manual lifting	2[c]	1.4 ~ 3.5	9	1.5 ~ 3.1

[a] Data taken from National Research Council (2001).

[b] Data taken from Jin et al. (2000) (China).

[c] Summary of data from Ory et al. (1997) (India) and Toroptsova et al. (1995) (Russian Federation).

the methodological quality of population surveys, such as sampling procedure, formulation of questions, procedures for administration of the survey, and nonresponse bias.

Little information on risk of low back pain by occupation is available from developing countries, in particular studies that would pass the quality criteria of the NIOSH or National Research Council reviews. One summary of the literature from China (Jin et al. 2000) reported risk factors for back pain similar to those reported in developed countries. However, in a comparison of the effect estimates for specific risk factors, the authors found slightly higher prevalence odds ratios (POR) in Chinese low back pain studies than other studies (Table 21.59). Alternative explanations would include unmeasured confounding or effect modification. Studies on occupational back pain performed in developing countries do generally report prevalences of back pain within specified occupations, but without comparing them to a reference group (Chiou and Wong 1992; Chiou et al. 1994; Joshi et al. 2001; Kumar et al. 1999; Muruka 1998; Omokhodion et al. 2000; Toroptsova et al. 1995; Yip 2001). Prevalences were generally high for the studied groups, but the lack of comparison to reference groups did not allow conversion into relative risk information, which was necessary for this analysis.

In summary, plausible arguments can be and have been advanced in favour of low back pain rates in specific occupational groups (farmers, factory workers, etc.) being both higher and lower in developing countries compared with developed countries, but the available data permit neither confirmation of this nor quantification of the differences in risk.

6. OCCUPATIONAL RISK FACTORS FOR INJURIES

Workplace injuries are a common hazard for workers. Deaths due to occupational injuries are defined as any potentially avoidable death due

to an external cause resulting from an exposure related to the person's work. The definition excludes death during commuting to or from the workplace. Workers travelling for work purposes are included.

Data in developed countries indicate that differential risks for injury exist by sector, being highest in agriculture and production, less in sales and service, and lowest in professional, administrative and clerical sectors. But similar data are unavailable for developing countries. At the same time, occupational registries provide some indication of injury outcomes—vs risk factor exposure—which can be used to assess the mortality associated with occupational factors. Applying the fatality rates due to occupational injuries per 100 000 insured workers (Table 21.60) to the number of persons in the EAP, as defined earlier in the chapter, gives an indication of total deaths from injuries among workers. The rates reported here for fatal injuries were reported in most countries only for insured populations. Thus, we made the assumption that the same rates applied to all in the EAP, whether or not they were insured, despite some evidence that fatality rates are higher in uninsured populations (Dror 2001; Forastieri 1999; Loewenson 1998). Unfortunately, there is a lack of adequate data on work-related injuries in developing countries to make it possible to generate plausible rates for economic sectors by age, sex and subregion. Mortality outcomes were distributed in the same age pattern as reported in the United States for unintentional injuries.

Because no risk factor exposure is defined in this approach, the counterfactual risk (e.g. theoretical minimum risk level) was defined based on the outcome rather than risk factor exposure. To approximate the safest working conditions observed where all avoidable injury hazards are controlled by effective preventive measures, we chose the injury fatality rate of 0.1 (per 100 000 workers) in the age group 16–17 years and in the occupation category "service" from the National Traumatic Occupational Fatalities surveillance system for the United States for the period 1980–1995 (Marsh and Layne 2001).

6.1 Outcomes considered

The outcomes considered were unintentional injuries, which include motor vehicle accidents, poisonings, falls, fires, drownings and the category "other unintentional injuries". Other unintentional injuries comprise exposure to inanimate mechanical forces, exposure to mechanical forces, other accidental threats to breathing, exposure to electric current, radiation and extreme ambient air temperature and pressure, contact with venomous animals and plants, exposure to forces of nature and accidental exposure to other and unspecified factors. Homicide at the workplace was not assessed owing to a complete lack of data from developing countries. To estimate the impact of the disability produced by nonfatal injuries, years lived with disability (YLD) were estimated using

Table 21.60 Fatality rates due to occupational injuries (per 100 000 insured workers) by country and year

Country	*Year(s)*	*Fatality rate per 100 000*	*Source*
Australia	1982–1984	8.06	Harrison et al. (1989)
	1998–1999	4.0	Worker's Compensation Cases (NOSHC 2002)
Austria	1998	5.3	ILO (2000)
Bolivia	1995	3.7	PAHO/WHO (1998)
Brazil	1995	13.3	PAHO/WHO (1998)
Canada	1970–1997	8.79	Human Resources Development Canada (2000)
China	1997	11.1	Kam Lam (2000)
	1991–1997	9.1 (1991–1997); 11.5 (1997)	Xia et al. (2000)
Costa Rica	1996	10.5	PAHO/WHO (1998)
Cuba	1996	4.2	PAHO/WHO (1998)
Czech Republic	1999	4.2	ILO (2000)
Denmark	1999	2	ILO (2000)
Dominican Republic	1996	6.3	PAHO/WHO (1998)
El Salvador	1996	4.7	PAHO/WHO (1998)
European Union	1998	5.03	Dupre (2001)
Finland	1997	3.1	ILO (2000)
Ireland	1999	4.21	ILO (2000)
Jamaica	1996	11.8	PAHO/WHO (1998)
Jordan	1980–1993	25.5	Atallah et al. (1998)
Mexico	1996	10.4	PAHO/WHO (1998)
Namibia	1998/1999	25	Amweelo (2000)
New Zealand	1985–1994	5.03	Feyer et al. (2001)
Panama	1996	14.5	PAHO/WHO (1998)
Peru	1996	190	PAHO/WHO (1998)
Philippines	1999	11	National Statistics Office Philippines (2000)
Poland	1999	4.5	ILO (2000)
Slovenia	1998	4	ILO (2000)
Singapore	2000	10.82	Singapore Government (2000)
Spain	2000	9.2	Ministerio de Trabajo y Asuntos Sociales (2002)
Sweden	1998	1.7	ILO (2000)
Thailand	1999	11.48	ILO (2000)
United Kingdom	1998	0.8	ILO (2000)
United States	1980–1995	4.25	NIOSH (2000a)
Venezuela	1997	0.58	ILO (2000)

the same attributable fractions as for mortality (age and sex) (i.e. it was assumed that an occupational injury had the same likelihood of being fatal as injuries caused by other factors).

6.2 Underreporting

Conventional sources of data on fatal injuries at work are compensation registries, insurance companies, death certificates and autopsy reports based on mortuary records. Data from compensation registries and insurance companies underestimate the magnitude of fatal injuries, either because they do not cover some sectors of the workforce or because they refer only to successful claims. To improve the accuracy of the reporting of fatal injuries at work, many countries gather data from different systems and data sources (death certificates, insurance companies, labour inspectorates, coroners' files, medical examiners' files) or develop specific projects. Wide disparities exist regarding the accuracy of these sources in identifying fatal injuries at work.

Despite the usefulness of the death records, data from the United States reveal that such records identify only between 67% and 90% of fatal injuries at work. A similar underreporting (72.3%) has been found in the Mortality Registry of Tuscany in Italy (Chellini et al. 2002). The only study in a developing country that analysed underreporting showed that 28% of occupational fatalities in Cape Town, South Africa had not been reported in terms of statutory regulations (Lerer and Myers 1994). The level of underreporting increases to between 78% and 85% in rural areas (Schierhout et al. 1997). On the other hand, special registries also underreport; the National Fund for Occupational Diseases in Italy, for example, has a reporting rate of only 56.4% (Chellini et al. 2002).

To our knowledge, the most accurate system currently in place that uses multiple data sources to identify and classify work-related injuries is in the United States. Data are gathered from death certificates in two surveillance systems: the National Traumatic Occupational Fatality System (NTOF) of NIOSH and in the Bureau of Labor Statistics CFOI system. Thus it would appear that the United States has fairly complete records of occupational deaths due to injury (CDC 2001). Although, owing to paucity of data, we did not use the estimates of underreporting to calculate the rates of fatal injuries due to risks at work, this does indicate likely underestimation.

7. Results

Tables 21.61–21.63 present the overall attributable fractions, mortality and burden of disease for the selected occupational risk factors considered here.

In total, occupational risk factors considered here were responsible for 775 000 deaths worldwide in 2000. There were five times as many deaths in males as in females: 647 000 vs 128 000. The leading occupational

Table 21.61 Attributable fractions (%) for the disease burden due to occupational exposure

Risk factor	*Outcome*	*Males*	*Females*	*Total*
Ergonomic stressors	Low back pain	41	32	37
Noise	Hearing Loss	22	11	16
Agents leading to COPD	COPD	18	6	13
Asthmagens	Asthma	14	7	11
Risk factors for injuries	Unintentional injuries	12	2	8
Beryllium, cadmium, chromium, diesel exhaust, nickel, arsenic, asbestos, silica	Trachea, bronchus or lung cancer	10	5	9
Benzene, ethylene oxide, ionizing radiation	Leukaemia	2	2	2

Table 21.62 Deaths (000s) due to occupational exposure[a]

				Total	
Risk factor	*Outcome*	*Males*	*Females*	*Deaths*	*% total from occupational risk factors*
Agents leading to COPD	COPD	240	78	318	41
Risk factors for injuries	Unintentional injuries	291	19	310	40
Beryllium, cadmium, chromium, diesel exhaust, nickel, arsenic, asbestos, silica	Trachea, bronchus or lung cancer	88	14	102	13
Asthmagens	Asthma	23	15	38	5
Benzene, ethylene oxide, ionizing radiation	Leukaemia	4	3	7	1
Total		647	128	775	100

[a] Asbestos exposure is the most important cause of mortality from mesothelioma. Cause-of-death statistics coded in ICD-10 allow direct estimation of the total number of mesothelioma deaths. Using this method, recent studies suggest that each year there are about 700 malignant mesothelioma deaths in Australia (Leigh and Driscoll 2003), 700 in Japan (Furuya et al. 2003), 2600 in the United States (Price and Ware 2004), and 4000 in Europe (Peto et al. 1999). A large proportion of these deaths are undoubtedly caused by asbestos exposure, primarily work-related. Combining estimates of asbestos exposure in all 14 subregions with hazards obtained from these studies would result in an estimate of more than 40 000 mesothelioma deaths caused by asbestos exposure in the world. Of these preliminary estimates, about 9000 occur in developed countries (AMR-A, EUR and WPR-A), 9 000 in SEAR-D, and 16 000 in WPR-B. These estimates are subject to uncertainty, especially in developing countries where ICD-10 cause-of-death data and detailed data on history of asbestos exposure are not available. These preliminary estimates are currently undergoing further refinement by authors. Preliminary estimates also indicate that there may have been approximately 9000 deaths from silicosis, 7000 deaths from asbestosis and 14 000 deaths from coal workers' pneumoconiosis as a result of exposure to occupational dusts (silica, asbestos and coal dust) in 2000.

Table 21.63 DALYs (000s) due to occupational exposure

				Total	
Risk factor	*Outcome*	*Males*	*Females*	*DALYs*	*% total from occupational risk factors*
Risk factors for injuries	Unintentional injuries	9779	718	10496	48
Noise	Hearing Loss	2788	1362	4150	19
Agents leading to COPD	COPD	3020	713	3733	17
Asthmagens	Asthma	1110	511	1621	7
Beryllium, cadmium, chromium, diesel exhaust, nickel, arsenic, asbestos, silica	Trachea, bronchus or lung cancer	825	144	969	4
Ergonomic stressors	Low back pain	485	333	818	4
Benzene, ethylene oxide, ionizing radiation	Leukaemia	66	35	101	0
Total		18073	3816	21889	100.0

cause of death was COPD (41%) followed by unintentional injuries (40%) and trachea, bronchus or lung cancer (13%). Workers who developed outcomes related to occupational risk factors lost about 22 million years of healthy life. By far the main cause of years of healthy life lost, within occupational diseases, was unintentional injuries (with 48% of the burden). This was followed by hearing loss due to occupational noise (19%) and COPD due to occupational agents (17%). Among the occupational factors analysed in this study, these three conditions accounted for 84% of years of healthy life lost. DALYs were almost five times greater in males than in females. Low back pain and hearing loss have in common the fact that they do not directly produce premature mortality, but substantial disability. This feature differentiates these conditions from the others analysed in the study. Results for specific risk factors are provided below.

7.1 Carcinogens

Tables 21.64–21.68 summarize the attributable fractions, mortality and burden of disease for the occupational carcinogens considered here.

For lung cancer, the attributable fraction varied from 5% in AMR-A to 14% in EUR-C, with overall attributable fractions for lung cancer estimated to be 10% for men and 5% for women (9% overall). For leukaemia, estimates of the attributable fraction varied from 1% in EMR-D to 3% in several subregions. There were estimated to be approximately 7000 deaths from leukaemia each year, with a much more even proportion between males and females than was seen for lung cancer, although approximately two thirds of the DALYs are due to male cases.

Table 21.64 Attributable fractions for lung cancer and leukaemia disease burden caused by workplace exposure

	Lung cancer			*Leukaemia*		
Subregion	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
AFR-D	9	4	7	3	1	2
AFR-E	9	4	7	3	2	3
AMR-A	6	2	5	3	3	3
AMR-B	11	3	8	2	2	2
AMR-D	12	2	8	3	2	3
EMR-B	12	2	9	3	2	2
EMR-D	9	3	7	2	1	1
EUR-A	7	2	6	3	3	3
EUR-B	12	4	10	3	2	3
EUR-C	15	9	14	2	2	2
SEAR-B	10	4	9	2	2	2
SEAR-D	11	4	9	2	0	2
WPR-A	8	3	6	2	2	2
WPR-B	12	7	10	2	2	2
World	10	5	9	2	2	2

Table 21.65 Deaths (000s) from lung cancer and leukaemia caused by workplace exposure

	Lung cancer			*Leukaemia*		
Subregion	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
AFR-D	1	0	1	0	0	0
AFR-E	1	0	1	0	0	0
AMR-A	7	2	8	0	0	1
AMR-B	4	0	4	0	0	0
AMR-D	0	0	0	0	0	0
EMR-B	1	0	1	0	0	0
EMR-D	1	0	1	0	0	0
EUR-A	11	1	12	1	0	1
EUR-B	6	0	6	0	0	0
EUR-C	12	1	14	0	0	0
SEAR-B	3	0	3	0	0	0
SEAR-D	11	1	12	0	0	1
WPR-A	3	0	4	0	0	0
WPR-B	27	7	34	1	1	2
World	88	14	102	4	3	7

Table 21.66 DALYs (000s) due to lung cancer and leukaemia caused by workplace exposure

	Lung cancer			*Leukaemia*		
Subregion	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
AFR-D	6	1	7	2	1	3
AFR-E	9	2	11	4	2	6
AMR-A	53	13	65	4	3	7
AMR-B	34	4	38	4	4	8
AMR-D	2	0	2	2	1	2
EMR-B	10	1	11	2	1	3
EMR-D	14	2	16	3	1	4
EUR-A	89	9	99	6	4	10
EUR-B	60	5	65	3	2	5
EUR-C	127	14	140	2	2	4
SEAR-B	32	3	34	3	2	5
SEAR-D	109	11	120	10	1	11
WPR-A	23	3	26	1	1	2
WPR-B	257	76	333	19	11	30
World	825	144	969	66	35	101

Table 21.67 Age-specific attributable fractions, deaths and DALYs for lung cancer and leukaemia, males

	Age group (years)						
	15–29	*30–44*	*45–59*	*60–69*	*70–79*	*80–89*	*All ages*
Attributable fractions (%)							
Lung cancer	11	11	10	10	10	9	10
Leukaemia	3	3	3	3	3	3	2
Deaths (000s)							
Lung cancer	0	3	20	30	26	8	88
Leukaemia	1	1	1	1	1	0	4
DALYs (000s)							
Lung cancer	10	76	306	279	136	18	825
Leukaemia	29	13	11	7	4	1	66

For each condition, deaths were predominantly among older persons up to 79 years, whereas DALYs tended to be highest in the younger age groups.

7.2 Nonmalignant respiratory diseases

Tables 21.69–21.74 summarize the attributable fractions, mortality and disease burden for asthma and COPD risk factors, each estimated as described earlier.

Table 21.68 Age-specific attributable fractions, deaths and DALYs for lung cancer and leukaemia, females

	Age group (years)						
	15–29	*30–44*	*45–59*	*60–69*	*70–79*	*80–89*	*All ages*
Attributable fractions (%)							
Lung cancer	5	5	5	5	4	4	5
Leukaemia	2	3	3	3	3	3	2
Deaths (000s)							
Lung cancer	0	1	3	4	4	2	14
Leukaemia	0	0	0	0	1	1	3
DALYs (000s)							
Lung cancer	3	19	52	41	25	4	144
Leukaemia	10	8	8	4	3	1	35

Table 21.69 Attributable fractions (%) for mortality from asthma and COPD caused by workplace exposure

	Asthma			*COPD*		
Subregion	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
AFR-D	21	15	18	16	5	11
AFR-E	23	18	20	16	5	11
AMR-A	15	9	11	18	3	11
AMR-B	20	8	13	17	3	11
AMR-D	19	7	13	15	2	9
EMR-B	18	5	12	17	2	11
EMR-D	20	10	16	17	3	11
EUR-A	16	7	11	19	4	13
EUR-B	22	14	18	19	6	14
EUR-C	21	12	18	21	6	16
SEAR-B	23	14	18	18	6	13
SEAR-D	23	14	18	16	5	11
WPR-A	17	9	13	21	5	16
WPR-B	22	16	19	19	7	12
World	21	13	17	18	6	12

It was estimated that 38 000 deaths (23 000 men and 15 000 women) and 1.6 million DALYs result from occupational asthma each year. One quarter to one third of the asthma deaths and DALYs occurred in SEAR-D. The attributable fraction for mortality from asthma varied between subregions from 11% in AMR-A and EUR-A to 20% in AFR-E, with worldwide attributable fractions estimated to be 21% for men and 13% for women (17% overall). The overall attributable

Table 21.70 Attributable fractions (%) for burden of disease (DALYs) for asthma and COPD caused by workplace exposure

	Asthma			*COPD*		
Subregion	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
AFR-D	11	7	10	16	5	11
AFR-E	13	9	11	16	5	12
AMR-A	9	4	7	18	3	11
AMR-B	12	4	8	17	3	10
AMR-D	11	3	7	13	1	7
EMR-B	11	2	7	17	2	12
EMR-D	14	6	10	17	3	11
EUR-A	11	4	8	19	4	12
EUR-B	15	8	12	19	6	13
EUR-C	18	8	14	21	6	14
SEAR-B	16	9	13	18	6	13
SEAR-D	17	10	13	16	5	11
WPR-A	12	5	9	21	5	14
WPR-B	15	9	12	19	7	14
World	14	7	11	18	6	13

Table 21.71 Numbers of deaths (000s) from asthma and COPD caused by workplace exposure

	Asthma			*COPD*		
Subregion	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
AFR-D	1	1	2	4	1	6
AFR-E	2	1	3	5	1	7
AMR-A	0	0	1	12	2	14
AMR-B	1	0	1	8	1	9
AMR-D	0	0	0	0	0	0
EMR-B	0	0	0	1	0	1
EMR-D	2	1	2	7	1	8
EUR-A	1	1	1	16	2	18
EUR-B	1	1	2	5	1	7
EUR-C	2	1	3	12	2	15
SEAR-B	2	2	4	8	1	9
SEAR-D	7	5	12	47	13	60
WPR-A	1	0	1	3	0	4
WPR-B	3	3	6	109	52	161
World	23	15	38	240	78	318

Table 21.72 DALYs (000s) from asthma and COPD caused by workplace exposure

	Asthma			*COPD*		
Subregion	*Males*	*Females*	*Total*	*Males*	*Females*	*Total*
AFR-D	63	27	90	43	10	53
AFR-E	84	56	141	57	12	69
AMR-A	37	15	51	147	21	168
AMR-B	98	27	125	115	17	132
AMR-D	16	4	19	6	0	6
EMR-B	18	3	21	20	1	20
EMR-D	74	27	100	75	13	87
EUR-A	41	14	55	176	29	205
EUR-B	30	13	43	75	19	94
EUR-C	32	9	41	135	34	169
SEAR-B	44	26	70	90	21	111
SEAR-D	310	166	476	552	149	701
WPR-A	23	9	33	44	9	53
WPR-B	241	115	356	1485	378	1862
World	1110	511	1621	3020	713	3733

Table 21.73 Age-specific mortality attributable fractions, deaths and DALYs for asthma and COPD, males

	Age group (years)						
	15–29	*30–44*	*45–59*	*60–69*	*70–79*	*80–89*	*All ages*
Attributable fractions (%)							
Asthma	23	23	23	22	22	21	21
COPD	17	18	18	18	18	19	18
Deaths (000s)							
Asthma	3	4	6	4	4	2	23
COPD	0	3	29	56	91	62	240
DALYs (000s)							
Asthma	670	228	144	43	20	5	1110
COPD	88	564	992	710	517	149	3020

fraction for asthma morbidity plus mortality was about two thirds of that for mortality, reflecting the fact that globally a great deal of asthma occurs at younger ages and is nonfatal and nonoccupational in origin.

For COPD mortality, the attributable fraction varied between subregions from 9% in AMR-D to 16% in EUR-C and WPR-A (Table 21.69).

Table 21.74 Age-specific mortality attributable fractions, deaths and DALYs for asthma and COPD, females

	Age group (years)						
	15–29	*30–44*	*45–59*	*60–69*	*70–79*	*80–89*	*All ages*
Attributable fractions (%)							
Asthma	13	14	14	13	13	12	13
COPD	6	5	5	6	6	6	6
Deaths (000s)							
Asthma	2	3	4	2	2	2	15
COPD	0	1	6	13	28	30	78
DALYs (000s)							
Asthma	228	95	81	28	15	5	511
COPD	45	133	149	152	166	69	713

Worldwide attributable fractions for COPD were estimated to be 18% for men and 6% for women (12% overall). Overall attributable fractions (based on DALYs and reflecting mortality and morbidity) were very similar to the mortality-based fractions (see Tables 21.69 and 21.70). The estimated number of deaths is almost an order of magnitude higher for COPD than for asthma, with an estimated 318 000 deaths (240 000 men and 78 000 women) and 3.7 million DALYs resulting from occupational COPD each year. Half of the COPD deaths and half of the DALYs occurred in WPR-B, owing in part to the large population of the subregion, high background COPD mortality rates and the relatively high employment in mining.

For both asthma and COPD, males predominated. Compared to females, males had nearly 50% higher attributable fraction for asthma mortality and three times that for COPD mortality. The ratio was about two for disease burden. Similar ratios were seen for the estimated numbers of deaths and DALYs due to these conditions. Asthma deaths were fairly evenly spread among all age groups from 30 to 79 years of age, whereas DALYs predominantly involved persons aged 30–59 years. For COPD, the majority of deaths occurred in persons aged ≥60 years, whereas DALYs were more evenly spread among all age groups from 30 to 79 years of age (see Tables 21.73 and 21.74).

7.3 Noise

Occupational noise-induced hearing loss accounted for more than four million DALYs, all of them produced by the disability associated with hearing loss (YLD). Worldwide, the burden of hearing loss attributed to occupational noise is 16%, ranging between 7% in WPR-A and 21% in WPR-B. By sex, the effects of exposure to occupational noise are larger for males than for females in all subregions (Table 21.75). Attributable fractions are related to age group and sex in all subregions. Males usually

Table 21.75 Attributable fractions of occupational noise-induced hearing loss, by sex and subregion

Subregion	*Males*	*Females*	*All*
AFR-D	23	11	17
AFR-E	23	12	18
AMR-A	12	5	9
AMR-B	19	9	15
AMR-D	18	9	14
EMR-B	20	9	15
EMR-D	20	13	16
EUR-A	13	5	9
EUR-B	24	13	19
EUR-C	24	13	18
SEAR-B	23	16	19
SEAR-D	24	9	16
WPR-A	9	6	7
WPR-B	26	15	21
World	22	11	16

Table 21.76 Attributable fractions (%) and DALYs (000s) for occupational noise-induced hearing loss, by age group

	Age group (years)[a]												
	15–29		*30–44*		*45–59*		*60–69*		*70–79*		*Total*		
	Males	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*All*
Attributable fraction	29	16	29	16	21	11	13	6	3	1	22	11	16
DALYs	425	206	1144	530	925	482	271	136	23	9	2788	1362	4151

[a] AF was set to zero for ≥80.

experience greater exposure to noise at work than females, owing to differences in occupational categories, economic sectors of employment and working lifetime. In this study, the attributable fraction decreased with age group after 30–44 years, indicating the heavy impact of occupational noise on the burden of hearing loss at younger ages (Table 21.76). The 30–44-year age group accounted for the highest number of DALYs and the 70–79-year age group for the lowest (1 673 000 vs 32 000).

Table 21.77 provides estimates of the number of DALYs (in thousands) produced by occupational noise-induced hearing loss by subregion in the year 2000. Overall, four million DALYs were lost owing to noise-induced hearing loss. SEAR-D and WPR-B accounted for more

Table 21.77 DALYs (000s) due to occupational noise-induced hearing loss, by sex and subregion

Subregion	*Males*	*Females*	*All*
AFR-D	109	49	157
AFR-E	127	60	186
AMR-A	92	31	123
AMR-B	122	43	165
AMR-D	15	6	20
EMR-B	60	21	81
EMR-D	142	88	230
EUR-A	117	47	164
EUR-B	92	50	142
EUR-C	136	92	228
SEAR-B	219	185	404
SEAR-D	799	303	1101
WPR-A	26	22	48
WPR-B	735	365	1100
World	2788	1362	4151

Table 21.78 Attributable fraction and DALYs of low back pain due to occupational ergonomic stressors, by sex and subregion

	AF (%)			*DALYs (000s)*		
Subregion	*Males*	*Females*	*All*	*Males*	*Females*	*All*
AFR-D	36	29	33	21	16	37
AFR-E	36	31	33	25	20	45
AMR-A	35	25	30	17	10	27
AMR-B	41	23	33	32	15	47
AMR-D	34	18	27	4	2	6
EMR-B	31	12	22	9	3	12
EMR-D	36	25	31	25	16	41
EUR-A	34	22	29	21	11	32
EUR-B	43	37	40	18	12	30
EUR-C	45	36	41	21	14	34
SEAR-B	43	34	39	26	19	46
SEAR-D	43	34	38	111	78	189
WPR-A	38	27	33	9	5	14
WPR-B	44	38	41	146	110	256
World	41	32	37	485	333	818

Table 21.79 Summary results describing the global burden of occupational injuries

Measure	*Males*	*Females*	*Total*
Attributable fraction for disease burden (%)	12	2	8
Deaths (000s)	291	19	310
DALYs (000s)	9779	718	10496

than half of the years of healthy life lost (1.1 million each in SEAR-D and WPR-B). Males lost twice the number lost by females (2 788 000 vs 1 362 000).

7.4 Ergonomic factors

The attributable fractions for low back pain ranged from 22% to 41% among the subregions (Table 21.78). Differences by age group were quite small, and the attributable fractions for the total working population (ages 15–65 years) were rather consistent. In most geographical regions, women have a lower attributable burden of low back pain than men, although the difference is most pronounced in the eastern Mediterranean region and the less developed countries in the Americas.

Occupational ergonomic stressors caused 818 000 DALYs due to low back pain in 2000 (Table 21.78). Globally, 37% of low back pain was attributable to occupational causes. The occupational contribution to the burden of low back pain varied relatively little between subregions, with 22% being the lowest (EMR-B) and 41% the highest (EUR-C and WPR-B). The attributable fraction in men (41%) was slightly higher than that in women (32%), which is mainly due to the type of work men perform, involving more vibration, heavy physical loads or handling of materials.

7.5 Risk factors for injuries

Work-related risk factors for unintentional injuries represent 8% of the burden of unintentional injuries. In all regions, the highest attributable fractions were found in males, reflecting the high number of males exposed to hazardous conditions in the workplace. Overall, the attributable fraction for males was 12% and for women 2%. Occupational injuries were responsible for 310 000 deaths (291 000 males and 19 000 females).

In 2000, there were 10 496 000 years of healthy life (DALYs) lost among exposed workers (Table 21.79). Overall, males lost about 90% of healthy life years owing to unintentional injuries at work.

8. Discussion

We have attempted to estimate the burden of disease due to selected occupational risk factors by considering exposure, rather than the

common actuarial approach. In this study, a methodology based on the EAP, economic sectors and subsectors and occupational categories was developed to quantify the exposure. Assignment of exposure (low/high) within these categories allowed us to make estimations about the amount of exposure to a given risk factor or groups of risk factors causing an outcome. The dominant source of uncertainty in this analysis was characterizing exposure, which was solely based on economic subsectors and/or occupations and involved a large number of extrapolations and assumptions. High-quality exposure data are lacking, especially in developing countries, and European and American exposure estimates were thus applied in many instances in developing regions (B, C, D and E subregions). This extrapolation could have substantial impact on the accuracy of analysis for the developing regions if exposures, as usually occur, vary from place to place and over time. Diseases with long latency (e.g. cancers) are those that are more susceptible to the assumptions and extrapolations. In addition to problems produced by the length of the latency period, the magnitude of the excess risk may vary depending on the age of the person when exposure began, the duration and strength of exposure and other concomitant exposures. The turnover of workers is another problem that affects both exposure and risk assessment.

The accuracy of the exposure data is fairly coarse because exposures vary greatly within an occupation. This indirect estimation of exposure may cause misclassification of the true exposure situation. The proportions of exposed workers with high exposure in the A and in the B, C, D and E subregions were less than the published data would indicate. This may be partly because the published literature often focuses on industries and/or occupations with high exposure, but may also indicate an underestimation of true exposure.

Sources of uncertainty in hazard estimates (relative risk and mortality rates) include variations determined from the literature (once again caused by the use of different exposure proxies), extrapolations to regions with different working conditions, the application to females of risk measures from male cohorts, and the application of the same relative risk values to all age groups (e.g. carcinogens).

Restricting the analysis to persons aged ≥15 years excludes the quantification of child labour. The exclusion of children in the estimation was due to the wide variation in the youngest age group for which countries reported EARs. In addition to inconsistent data on EARs for children, there was virtually no data available on their exposure to occupational risk factors or the relative risks of such exposures. Specific, focused research on children is needed to quantify the global burden of disease due to child labour and the resulting implications.

Owing to lack of global data, we could not analyse occupational contributions to the global burden of infectious diseases, cardiovascular dis-

orders, MSDs of the upper extremities, skin disorders and other conditions with recognized occupational etiologies.

8.1 Occupational carcinogens

For each condition, deaths were predominantly in older persons up to 79 years, whereas DALYs tended to be highest in the younger age groups. The estimated overall attributable fractions for lung cancer of 10% for men and 5% for women (9% overall) are similar to those from a recent United States study, based on a review of relevant studies, in which the attributable fraction for lung cancer was estimated to be between 6% and 17% for men, and to be about 2% for women (Steenland et al. 2003). A similar Finnish study used estimates of 29% (men) and 5% (women) but these included a contribution from environmental tobacco smoke, which the study estimated to be about 2% or 3% (Nurminen and Karjalainen 2001). (The United States study did not include any contribution from environmental tobacco smoke, but separately estimated the contribution of workplace environmental tobacco smoke to be 5.7% [Steenland et al. 2003]).

The estimated 2% attributable fraction for leukaemia compares with 0.8–2.8% for the United States (Steenland et al. 2003) and 18% (men) and 2% (women) for Finland (Nurminen and Karjalainen 2001). The higher Finnish estimate seems to arise from the inclusion of occupational exposure to electromagnetic fields, from the reliance on different studies for relative risk estimates, and from the exposure patterns in the Finnish population.

8.2 Nonmalignant respiratory diseases

Many of the issues relevant to a discussion of the results for particulates are also relevant to carcinogens, and were discussed in detail under that rubric. The estimated attributable fractions for asthma mortality of 21% for men and 13% for women (17% overall) are similar to those from two recent reviews, both of which found an occupational attributable fraction of 15% (Balmes et al. 2003; Blanc and Toren 1999). The Finnish study on which most of the occupational relative risk estimates used in this study were based had higher estimates for men (29%) and women (17%) (Karjalainen et al. 2002), but these estimates are based on Finnish workforce patterns, which are likely to differ from those in most other countries.

Estimates of the attributable fraction for COPD mortality varied between subregions from 9% to 16%. The overall value of 12% is very close to the few published estimates of occupational attributable fraction for COPD of 14% in the United States (Steenland et al. 2003, based on Korn et al. 1987), 14% for men and 5% for women in Finland (Nurminen and Karjalainen 2001) and 15% in a recent review by the American Thoracic Society (Balmes et al. 2003).

8.3 Noise

Occupational noise-induced hearing loss accounted for more than four million DALYs, all of them produced by the disability associated with hearing loss (YLD). Worldwide, the burden attributed to occupational noise is 16%, ranging between 7% in WPR-A and 21% in WPR-B. By sex, the effects of exposure to occupational noise are larger for males than for females in all subregions. The attributable fraction decreased with age group after 30–44 years, indicating the heavy impact of occupational noise on the burden of hearing loss at younger ages.

In addition to causing irreversible hearing loss, high noise levels in the workplace cause elevated blood pressure, sleeping difficulties, annoyance and stress. Our findings indicate that occupational noise has multiple consequences, both for the individual and for society, and particularly for those suffering hearing loss at young ages. Most occupational noise exposure can be minimized by the use of engineering controls to reduce the generation of noise at its source, within complete hearing loss prevention programmes that include noise assessment, audiometric monitoring of workers' hearing, appropriate use of hearing protectors and worker education.

8.4 Ergonomic factors

Human capacity for work depends on many functions and attributes: body size, muscle strength, aerobic fitness, sensory perception and cognitive capacity. Features of the work environment that do not accommodate these needs may produce physical or psychosocial stressors on the human system. Work features that have received attention because of their adverse health effects include heavy manual handling and other types of strenuous work, and awkward body postures.

For this analysis, the exposure variable was work in an occupational category with its assigned level of risk (low, medium or high rate of low back pain). This exposure variable is the "proxy" for the combination of occupational exposures found in the specified occupation that are implicated in the etiology of low back pain.

Occupational ergonomic stressors caused 818 000 DALYs from low back pain in 2000. The attributable fractions of low back pain ranged from 22% to 41% among subregions, with the global fraction amounting to 37%. The African countries had the highest attributable fraction of low back pain for all age groups analysed. Fractions of 40% or above were reached in EUR-B and EUR-C and in SEAR-B. The attributable fraction in men (41%) was slightly higher than that in women (32%), which is mainly due to the type of work men perform, involving more vibration, heavy physical load or material handling. Subregional variations reflect differences in occupational types and exposure. Over half of the working population in AFR-D and AFR-E was employed in agriculture. In contrast, about one third of the working populations in the AMR

and EUR subregions were employed in production occupations ("operators") and another large fraction (40% or more) in professional, sales and clerical jobs. In general, males are more exposed than females because they constitute a higher proportion of the labour force. In the less developed subregions, males are generally more exposed because of the higher proportions of workers in formal agriculture than in the developed subregions. The proportion of females in the labour force was particularly low in EMR-B and EMR-D.

The available literature demonstrates the feasibility and benefits of workplace ergonomic interventions (training and engineering controls) that have been implemented by employers in numerous economic sectors. Effective abatement measures include redesigning workstations to eliminate the need for bending and twisting; installing material or patient hoists and other lifting devices; a greater variety of work tasks, to avoid repetitively loading the same body tissues; and improving the mechanical isolation of seating to reduce transmission of whole-body vibration. Training programmes are most effective when they address job design, target supervisory and management personnel along with the manual labour force, and take place in a setting where workers are empowered to utilize the knowledge imparted. In general, the coordination of multiple activities—workstation improvements, training, enhanced medical surveillance and management—within an intervention programme appears to be the most effective. This is consistent with the conclusions of Shannon et al. (1996, 1997) that lower injury rates are associated with workplace characteristics such as general workforce empowerment and top management's active leadership, together with delegation of decision-making authority regarding occupational safety.

8.5 Risk factors for injuries

To our knowledge, this is the first study to estimate attributable fractions of work-related risk factors for unintentional injuries within the overall burden of DALYs. Lack of data on exposure did not allow a risk based approach and the estimates were based on occupational injury registries. This will limit the applicability of these estimates to preventive purposes which are based on exposure. Our findings show that the overall attributable fraction of 9% reported in this study is above the upper range of values reported by Chen et al. (2001) in the United States. Chen et al. reported an overall attributable fraction of 3.8%, varying between 1.5% in Arizona and 9.8% in Alaska. The difference in the findings between the two studies is explained by the heavy burden of mortality in the DALY estimation in developing countries, especially when deaths occur in younger populations.

Our findings understate the importance of the impact of occupational risk factors leading to injuries in the overall burden of disease due to injuries. A major factor in the underestimation was our use of data from an insured population from one country. There is some evidence that

mortality can be greater in uninsured populations, but in the absence of consistent evidence, a similar mortality in the insured and uninsured populations was assumed (Dror 2001; Forastieri 1999; Loewenson 1998). Lerer and Myers (1994) found that 28% of occupational fatalities in Cape Town, South Africa, were not reported despite a statutory requirement to do so. Using this fraction, we may have missed about 100 000 occupational injury deaths due to underreporting. Also, we did not estimate the injury mortality due to intentional injuries such as homicides in the workplace, owing to the lack of data from developing countries. However, current evidence shows that intentional injuries must be present in such countries; thus the lack of an estimation of deaths due to this cause increases the degree of underestimation of the number of deaths due to injuries (e.g. by approximately 4% in Australia and New Zealand).

Analysis of the full contribution of injuries at work within the overall burden of injuries requires indicators that measure not only mortality but also morbidity. In some countries and regions, with constant or slightly decreasing mortality patterns, it has been observed that the decline in mortality is balanced by an increase in the severity of injuries and morbidity, especially long-lasting or permanent disabilities (CDC 2001; Guerrero et al. 1999). In these cases, evaluation of the effectiveness of preventive measures is also hampered.

Injuries are largely preventable by improvements to make work safer and healthier. Engineering controls, administrative policies, health and safety information and education to promote safety-conscious attitudes and behaviour are needed. Surveillance data must be developed to provide the basis for targeting preventive measures towards high-risk groups of workers. The distribution of burden by type of external cause of mortality has allowed the developed countries to focus on preventive actions at work, resulting in a reduction in injury rates over time. Similar analysis and preventive actions in other countries could greatly reduce injuries at the workplace.

8.6 Conclusion

The aim of this study was to estimate the attributable fractions of selected occupational exposures. The risk factors were selected according to the availability of data, the strength of evidence linking the occupational exposure and the outcome, and the amount of risk arising from the exposure. An important feature of these risk factors and the resulting disease burden is their concentration among the working population, especially those in high-risk occupations and sectors. Hazards at workplaces and the resulting illness and injury are understood most accurately in the formal sector, and even there much undercounting occurs. The burden in the informal sector in developing countries, where large proportions of the population work, is high and largely lacks description. Neither household and family agricultural work by women nor child

Table 21.80 Expected rate of growth of the economically active population between 2000 and 2010, by sex and subregion

	Growth rate	
Subregion	*Males*	*Females*
AFR-D	0.33	0.38
AFR-E	0.27	0.25
AMR-A	0.08	0.12
AMR-B	0.17	0.27
AMR-D	0.27	0.45
EMR-B	0.32	0.66
EMR-D	0.31	0.53
EUR-A	–0.35	0.03
EUR-B	–0.26	0.16
EUR-C	0.00	0.00
SEAR-B	0.17	0.24
SEAR-D	0.21	0.27
WPR-A	–0.03	0.04
WPR-B	0.11	0.11

Source: ILO (2002a).

labour were addressed in our study. Due primarily to lack of data in developing countries, we were unable to include important occupational risks for infectious diseases, dermatitis, reproductive disorders, some cancers, ischaemic heart disease, musculoskeletal disorders of the upper extremities, and other conditions such as workplace stress.

The estimated burden of occupational risk factors can be diminished by improving working conditions, as many examples from different countries have shown. Work-related diseases are largely preventable. For example, many dusty activities can be made safer by using wet methods, thus reducing workers' exposures to silica. Work surfaces can be adjusted to a worker's height, thereby reducing suffering from low back pain. Substituting safe chemicals for known carcinogens can prevent many cancers. A change of process can reduce noise levels, thus protecting workers' hearing. Attention to electrical safety or machine guarding can eliminate tragic injuries at the workplace.

9. Projections of future exposure

In the next 50 years, the population of the developing regions will steadily rise, whereas that of more developed regions is expected to change little because fertility levels will remain below replacement level (UN 2001). There will also be differences in growth rates between the

sexes. A negative growth rate among economically active males is expected to occur between 2000 and 2010 in developed regions such as Europe, while comparable female rates will continue increasing in most of the regions, including the developed ones (ILO 2002a) (Table 21.80).

The expected changes in the world population will affect the EAP as well as the median age of workers (Fullerton and Toosi 2001). These changes in the characteristics of the working population will be accompanied by a different distribution of employment in the economic sectors (agriculture, industry and services). Currently, the service sector of many economies is growing at a fast rate, while the agricultural sector is rapidly declining in developing countries and remains at a stable low level in developed countries. It is expected that these different patterns of growth within the economic sectors will continue in the coming years. Moreover, the expected changes will affect the distribution of occupations within an economic sector. In developed countries in which a change in the structure of the economy has been observed, there has been a shift in the proportion of workers from the "production" category in favour of professional, managerial, clerical and sales occupations.

9.1 Exposure estimation for the years 2010, 2020 and 2030

As mentioned above, the EAP by economic sector was used to estimate the working population exposed to some risk factors, including car-

Table 21.81 Projected EARs for the year 2010 by sex and subregion

Subregion	*Males*	*Females*
AFR-D	0.84	0.55
AFR-E	0.85	0.64
AMR-A	0.7	0.59
AMR-B	0.8	0.45
AMR-D	0.81	0.44
EMR-B	0.78	0.39
EMR-D	0.81	0.42
EUR-A	0.57	0.47
EUR-B	0.69	0.56
EUR-C	0.75	0.59
SEAR-B	0.82	0.62
SEAR-D	0.84	0.48
WPR-A	0.71	0.52
WPR-B	0.81	0.68

Source: ILO (2002a).

Table 21.82 Projected distribution of EAP by subregion in 2010, 2020 and 2030

	2010		2020		2030	
Subregion	EAP	% total	EAP	% total	EAP	% total
AFR-D	151 300 284	4.6	199 904 691	5.4	260 126 023	6.3
AFR-E	181 011 306	5.5	232 504 879	6.3	300 771 980	7.3
AMR-A	176 129 373	5.4	191 817 362	5.2	201 632 376	4.9
AMR-B	218 574 298	6.7	251 138 963	6.8	278 348 992	6.8
AMR-D	35 233 802	1.1	43 360 108	1.2	51 164 255	1.2
EMR-B	67 730 185	2.1	82 896 189	2.2	98 826 530	2.4
EMR-D	165 776 470	5.1	214 302 617	5.8	269 408 795	6.5
EUR-A	172 528 633	5.3	171 619 225	4.6	165 811 266	4.0
EUR-B	110 565 142	3.4	118 115 432	3.2	123 452 320	3.0
EUR-C	125 923 283	3.8	118 343 637	3.2	112 001 074	2.7
SEAR-B	173 799 078	5.3	196 214 683	5.3	213 861 114	5.2
SEAR-D	663 743 911	20.3	784 531 784	21.1	885 891 293	21.5
WPR-A	77 452 109	2.4	76 367 545	2.1	72 515 258	1.8
WPR-B	952 086 321	29.1	1 030 847 264	27.8	1 086 544 112	26.4
Total	3 271 854 196	100.0	3 711 964 378	100.0	4 120 355 388	100.0

cinogens, while occupational category within a sector was used for others, including noise and ergonomic stressors. Therefore, to project the exposed population for the years 2010, 2020 and 2030 a three-step procedure was followed: (i) the EAP was estimated; (ii) the EAP was distributed among economic sectors; and, where needed, (iii) occupational categories were distributed within the economic sectors.

EAP estimation

To obtain the EAP for the year 2010, we multiplied the overall population (2010) by the EARs by subregion for the year 2010 as estimated by ILO (See Table 21.81). Then, in the absence of other data, the same EAR by subregion was used for the years 2020 and 2030 to generate the EAP (see Equation 4). Calculations were restricted to persons aged ≥15 years by sex and subregion, thus allowing regional patterns to be preserved.

$$EAP_{15+j} = \sum [EAR_{2010}(\text{for each age group} \geq 15 \times \text{Population}_i(\text{for each age group} \geq 15))] \quad (4)$$

where

EAP_{15+j} = economically active population ≥15 years, φ=year (2010, 2020, 2030)

$Population_i$ = population year 2010, 2020, 2030

EAR = economic activity rate, year 2010

The EAP will increase steadily towards 2010, 2020 and 2030, but the amount of the increase and the patterns are somewhat different between developed and developing countries, as well as among countries having a similar degree of development. The percentage distribution of the EAP by subregion reflects the growth of the overall population, with greater growth in developing countries. WPR-B and SEAR-D will contribute 49.4% of global EAP in the year 2010, whereas developed subregions will contribute only 13.1% (Table 21.82).

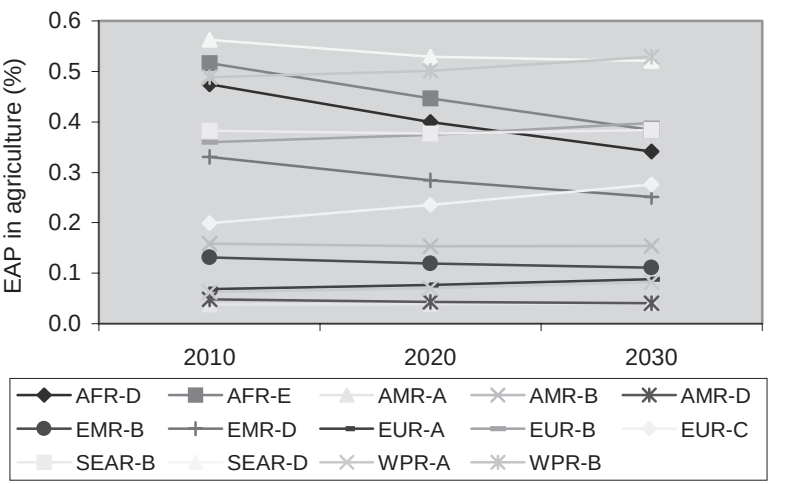
EAP DISTRIBUTION AMONG ECONOMIC SECTORS FOR 2010, 2020 AND 2030

The basic approach to estimating the EAP among economic sectors was to use regression analysis to identify the relationship between the distribution of the economic sectors and the projected years of interest. The dependent variables (proportion of EAP employed in agriculture, industry or services) were separately compared to the independent variable time,[10] using the following model:

$$PEAPA = \ln(a\ Y_T) + \ln b \qquad (5)$$

where

Figure 21.3 Projected distribution of the agricultural sector by year and subregion, 2010, 2020 and 2030



PEAPA = proportion of EAP in agriculture (similarly, PEAPI and PEAPS for industry, or service) in Year T

Y_T = Time (Year T)

The slope factors and intercepts obtained by regression analysis, using the EAP proportion by economic subsector for the years 1990–2000, were then used to estimate the proportion of the EAP for the years 2010, 2020 and 2030, separately for each economic subsector. We did not include economic development (e.g. measured as GDP per capita) as an additional variable in the analysis, assuming that previous trends capture the effects of trends in GDP. Given the economic and social factors that determine occupational distributions, the changes in the EAP in the future are subject to behavioural decisions by individuals, policy decisions in home countries and abroad, and developments in education. The project distribution of EAP among economic sectors showed different patterns among different subregions. As an example, Figure 21.3 presents the distribution of EAP in agriculture.

Occupational categories adjustment

No data were available to develop trends for employment in occupational categories in 2010, 2020 and 2030. Therefore, proportions of exposed workers within occupational categories were adjusted according to the distribution pattern of the year 2000, adjusted only for the new proportions employed within economic sectors in the year of interest.

Notes

1 See preface for an explanation of this term.

2 Dusts are technically defined as dry particle aerosols produced by mechanical processes such as breaking, grinding and pulverizing (Johnson and Swift 1997). Particle sizes range from less than 1 μm to over 100 μm. The smaller particles present a greater hazard, as they remain airborne longer and are more likely to enter the respiratory tract. Dusts may be organic (e.g. grain dust) or inorganic (e.g. silica, asbestos and coal dust).

3 Economic activities comprise agriculture, mining, manufacturing, utilities, construction, trade, transport, finance and services.

4 dBA is the unit of sound pressure level in decibels that has been A-weighted, i.e. measured with an A-weighted sound level meter. Sound levels measured in dBA have been widely used to evaluate occupational and environmental exposures because of the good correlations between the "A" scale and human hearing ability at different frequencies, hearing damage and environmental annoyance.

5 The average of the hearing threshold levels for both ears that exceeds 25 dB at 1000, 2000, 3000 and 4000 Hz.

6 Tinnitus is noise originating in the ear rather than in the environment. The noise may be a buzzing, ringing, roaring, whistling, humming or hissing in the ears. Ringing in the ears is an extremely common phenomenon experienced by up to a third of the adult population at one time or another.

7 A temporary increase in the threshold of hearing for an ear caused by exposure to high-intensity noise.

8 The percentage of workers with a hearing impairment in an occupationally noise-exposed population, after subtracting the percentage in an unexposed population who would normally incur such impairment owing to ageing.

9 Year was the predictor of the data.

References

Ahmed HO, Dennis JH, Badran O et al. (2001) Occupational noise exposure and hearing loss of workers in two plants in eastern Saudi Arabia. *Annals of Occupational Hygiene*, **45**:371–380.

Alidrisi M, Jamil ATM, Jiffry MSA, Jefri MA, Erturk F (1990) Evaluation of noise stresses in Jeddah Industrial State. *Journal of Environment Science and Health*, **A25**:873–896.

Amweelo M (2000) Accident prevention in Namibia. *African Newsletter on Occupational Health and Safety* **10**(1).

Armstrong TJ, Buckle P, Fine LJ et al. (1993) A conceptual model for work-related neck and upper-limb musculoskeletal disorders. *Scandinavian Journal of Work, Environment and Health*, **19**:73–84.

Arndt V, Rothenbacher D, Brenner H et al. (1996) Older workers in the construction industry: results of a routine health examination and a five-year follow-up. *Occupational and Environmental Medicine*, **53**:686–691.

Astrand NE (1987) Medical, psychological, and social factors associated with back abnormalities and self reported back pain: a cross sectional study of male employees in a Swedish pulp and paper industry. *British Journal of Industrial Medicine*, **44**:327–336.

Atallah ZR, Jamous LW, AbuDhaise BA, Alwash RH (1998) Fatal occupational injuries in Jordan during the period 1980 through 1993. *Safety Science*, **28**:177–187.

Attfield M, Wagner G (1998) Respiratory disease in coal miners. In: *Environmental and occupational medicine*, 3rd edn. Rom W, ed. Little, Brown and Co., Boston, MA.

Badley EM, Rasooly I, Webster GK (1994) Relative importance of musculoskeletal disorders as a cause of chronic health problems, disability, and health care utilization: findings from the 1990 Ontario Health Survey. *The Journal of Rheumatology*, **21**:505–514.

Badley EM, Webster GK, Rasooly I (1995) The impact of musculoskeletal disorders in the population: are they just aches and pain? Findings from the 1990 Ontario Health Survey. *The Journal of Rheumatology*, **22**:733–739.

Balmes J, Becklake M, Blanc P et al. (2003) American Thoracic Society statement on occupational contribution to the burden of airway disease. *American Journal of Respiratory and Critical Care Medicine*, **167**:787–797.

Bao S, Winkel J, Mathiassen SE, Shahnavan H (1997) Interactive effect of ergonomics and production engineering on shoulder-neck exposure: a case study of assembly work in China and Sweden. *International Journal of Industrial Ergonomics*, **20**:75–85.

Battie MC, Bigos SJ (1991) Industrial back pain complaints: a broader perspective. *Orthopedics Clinics of North America*, **22**:273–282.

Becklake M (1989) Occupational exposures: evidence for a causal association with chronic obstructive pulmonary disease. *American Review of Respiratory Disease*, **140**:S85–91.

Becklake MR (1994) Symptoms and pulmonary functions as measures of morbidity. *Annals Occupational Hygiene*, **38**:569–580, 418.

Behrens V, Seligman P, Cameron L et al. (1994) The prevalence of back pain, hand discomfort, and dermatitis in the U.S. working population. *American Journal of Public Health*, **84**:1780–1785.

BEIR V (1990) *Health effects of exposure to low levels of ionizing radiation: BEIR V*. Committee on the biological effects of ionizing radiation. National Academy Press, Washington, DC.

Bergenudd H, Nilsson B (1988) Back pain in middle age: occupational work load and psychologic factors: an epidemiologic study. *Spine*, **13**:58–60.

Bernard BP, ed. (1997) *Musculoskeletal disorders and workplace factors: a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back*. CDC/NIOSH, Cincinnati, OH.

Bernard BP, Sauter SL, Petersen M et al. (1993) Health hazard evaluation report: Los Angeles Times. (NIOSH Report No. 90-013-2277). Cincinnati, OH.

Bhatia R, Lopipero P, Smith A (1998) Diesel exhaust exposure and lung cancer. *Epidemiology*, **9**:84–91.

Bhattacharya SK, Saiyed HN, Roy A, Chatterjee SK (1981) Hearing acuity in weavers of a textile mill. *Indian Journal of Medical Research*, **74**:779–785.

Bhattacharya SK, Tripathi SR, Kashyap S (1990) Heat and noise problems in a firm in a drug and pharmaceutical firm in India. *Industrial Health*, **28**:203–207.

Blanc P, Toren K (1999) How much adult asthma can be attributed to occupational factors? *American Journal of Medicine*, **107**:580–587.

Bongers PM, de Winter CR, Kompier MAJ et al. (1993) Psychosocial factors at work and musculoskeletal disease. *Scandinavian Journal of Work, Environment and Health*, **19**:297–312.

Bongers PM, Hulshof CTJ, Dijkstra L, Boshuizen HC, Groenhout HJM, Valken E (1990) Back pain and exposure to whole body vibration in helicopter pilots. *Ergonomics*, **33**:1007–1026.

Bovenzi M, Betta A (1994) Low-back disorders in agricultural tractor drivers exposed to whole-body vibration and postural stress. *Applied Ergonomics*, **25**:231–241.

Broersen JPJ, de Zwart BCH, Dijk F et al. (1996) Health complaints and working conditions experienced in relation to work and age. *Occupational and Environmental Medicine*, **53**:51–57.

Buckle PW, Devereaux J (1999) *Work-related neck and upper limb musculoskeletal disorders*. European Agency for Safety and Health at Work, Luxembourg.

Bundesverband der Betriebskrankenkassen (2001) *Krankheitsartenstatistik 1997*. (Internet communication of 20 September 2001 at: http://de.osha.eu.int/statistics/bkk/index.stm).

Burchfiel CM, Boice JA, Stafford BA, Bond GG (1992) Prevalence of back pain and joint problems in a manufacturing company. *Journal of Occupational Medicine*, **34**:129–134.

Burdorf A (1992) Sources of variance in exposure to postural load on the back in occupational groups. *Scandinavian Journal of Work, Environment and Health*, **18**:361–367.

Burdorf A, Naaktgeboren B, deGroot HC (1993) Occupational risk factors for low backpain among sedentary workers. *Journal Occupational Medicine*, **35**:1213–1220.

Burdorf A, Sorock G (1997) Positive and negative evidence of risk factors for back disorders. *Scandinavian Journal of Work, Environment and Health*, **23**:243–256.

Bureau of Labor Statistics (2001) *Safety and health statistics*. Office of Safety, Health and Working Conditions. U.S. Department of Labor. (Internet Communication of 10 September 2001 at: http://stats.bls.gov/oshhome.htm).

Burton, DJ (1997) Chapter 31. General methods for the control of airborne hazards. In: *The industrial environment—its evaluation and control*. Di Nardi S, ed. AIHA, Fairfax, VA.

Butler S, Twaddle S (1979) Turnover in the restaurant industry: a structural explanation. Mid-South Sociological Association. *Sociological Abstracts*, **27**(4).

CCOHS (1997) *OSH answers*. Canadian Centre for Occupational Health and Safety. Available at http://www.ccohn.ca/oshanswers/diseases.

CDC (2001) Non-fatal occupational injuries and illnesses treated in hospital emergency departments—Unites States 1998. Centers for Disease Control and Prevention. *Journal of the American Medical Association*, **285**:2443–2444.

Chan-Yeung M, Malo J (1994) Aetiological agents in occupational asthma. *European Respiratory Journal*, **7**:346–371.

Checkoway (2000) Epidemiological evidence on the carcinogenicity of silica: factors in scientific judgement: letter. *Annals of Occupational Hygiene*, **44**:483–487.

Chellini E, Baldasseroni A, Giovannetti L, Zoppi O (2002) A survey on fatal work accidents based on mortality registry data: results of the Tuscany study

on INAIL and RMR cases in the period 1992–1996. *Epidemiologia e Prevenzione*, **26**:11–17.

Chen W, Zhuang Z, Attfield MD et al. (2001) Exposure to silica and silicosis among tin miners in China: exposure-response analyses and risk assessment. *Occupational and Environmental Medicine*, **58**:31–37.

Chien V, Chai S, Hai D et al. (2002) Pneumoconiosis among workers in a Vietnamese refractory brick facility. *American Journal of Industrial Medicine*, **42**:397–402.

Chiou, WK, Wong MK (1992) Epidemiology of low back pain in the nurses of Chang Gung Memorial Hospital. *Chang Gung Medical Journal*, **15**: 64–71.

Chiou WK, Wong MK, Lee YH (1994) Epidemiology of low back pain in Chinese nurses. *International Journal of Nursing Studies*, **31**:361–368.

Choi BCK, Tennassee LM, Eijkemans GJM (2001) Developing regional workplace health and hazard surveillance in the Americas. *Pan American Journal of Public Health*, **10**: 376–381.

Davis AC (1989) The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. *International Journal of Epidemiology*, **18**:911–917.

de Zwart BCH, Broersen JP, Frings-Dresen MHW, van Dijk FJH (1997) Repeated survey on changes in musculoskeletal complaints relative to age and work demands. *Occupational and Environmental Medicine*, **54**:793–799.

Derriennic F, Iwatsubo Y, Monfort C et al. (1993) Evolution of osteoarticular disorders as a function of past heavy physical work factors: longitudinal analysis of 627 retired subjects living in the Paris area. *British Journal of Industrial Medicine*, **50**:851–860.

Deyo RA, Rainville J, Kent D (1992) What can the history and physical examination tell us about low back pain? *Journal of the American Medical Association*, **268**:760–765.

Dosemeci M, McLaughlin JF, Chen J-Q et al. (1995) Historical total and respirable silica dust exposure levels in mines and pottery factories in China. *Scandinavian Journal of Work, Environment and Health*, **2**:S39–43.

Dror DM (2001) Reinsurance of health insurance for the informal sector. *Bulletin of the World Health Organization*, **79**:672–678.

Dupre D (2001) *Accidents at work in the EU 1998–1999*. EUROSTAT (Theme 3-16/2001). Statistics in Focus, Population and Social Conditions, Luxembourg.

EASHW (2000) *Monitoring the state of occupational safety and health in the European Union—pilot study*. European Agency for Safety and Health at Work. Luxembourg.

EIA (2001) Energy Information Administration, U.S. Department of Energy, International Energy Database, January 2001, available at www.eia.doe.gov/emeu/iea/coal.html.

EIU (1995) Business operations report, India, 4th quarter, 1995. Shortage of managers. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1996a) China hand. Ch. 13. Human resources: knowing why staff leave, 1 Sept. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1996b) China labour: unemployed reluctant to trade welfare for work. Country view, 27 Nov. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1997a) 28 February. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1997b) Brazil regulations: rules eased on loans for exporters. Country alert, 6 May. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1997c) Business operations report, Laos. Importance of training: training breeds staff loyalty, 26 Jun. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1997d) Caribbean industry: region seeks to lure informatics firms. Country alert, 2 Jul. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1997e) Country commerce South Korea. Human resources. Overview, 1 Jul. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1997f) USA industry: Silicon Valley survey—of mice & men. Country view, 1 Apr. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (1998) Business operations report Eastern Europe. Business services: background checks on security personnel essential, 16 March.

EIU (2001a) Country commerce China. Human resources. Overview, 28 Feb. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (2001b) Country forecast Africa. AIDS/Africa: labour costs are set to increase substantially, 14 May. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (2001c) Country profile Hong Kong. Resources and infrastructure: education, 1 Jul. Economist Intelligence Unit. Available at http://www.eiu.com.

EIU (2002) Country commerce China. Human resources. Overview, 28 Feb. Economist Intelligence Unit. Available at http://www.eiu.com.

Emmett EA (2002) Occupational contact dermatitis I: incidence and return to work pressures. *American Journal of Contact Dermatitis*, **13**:30–34.

Enterline P, Marsh G, Esmen N, Henderson V, Callahan C, Paik M (1987) Some effects of cigarette smoking arsenic, and SO_2 on mortality among US copper smelter workers. *Journal of Occupational Medicine*, **29**:831–838.

Evans WA, Ming HY (1982) Industrial noise-induced hearing loss in Hong Kong: a comparative study. *Annals of Occupational Hygiene*, **25**:63–80.

Ezenwa A (2001) A study of fatal injuries in Nigerian factories. *Occupational Medicine*, **51**:485–489.

Ferris BG Jr, Speizer FE, Spengler JD et al. (1979) Effects of sulfur and respirable particles on human health: methodology and demography of populations in study. *American Review of Respiratory Disease*, **120**:767–779.

Feyer AM, Williamson AM, Stout N, Driscoll T, Usher H, Langley JD (2001) Comparison of work related fatal injuries in the United States, Australia, and New Zealand: method and overall findings. *Injury Prevention*, **7**:22–28.

Fine LJ, Silverstein BA, Armstrong TJ, Anderson CA, Sugano DS (1986) Detection of cumulative trauma disorders of upper extremities in the workplace. *Journal of Occupational Medicine*, **28**:674–678.

FIOH (1999) *CAREX database*. Finnish Institute of Occupational Health. Available at http://www.ocuphealth.fi/internet/english.

FIOH (2001) CAREX database. Finnish Institute of Occupational Health. Available at http://www.occuphealth.fi/list/data/CAREX.

Forastieri V (1999) *Improvement of working conditions and environment in the informal sector through safety and health measures*. International Labour Organization, Geneva.

Frank JW, Kerr MS, Brooker A-S et al. (1996) Disability resulting from occupational low back pain, Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. *Spine*, **21**:2908–2917.

Frank JW, Pulcins IR, Kerr MS et al. (1995) Occupational back pain—an unhelpful polemic. *Scandinavian Journal of Work, Environment and Health*, **21**:3–14.

Fullerton H, Toosi M (2001) Labor force projections to 2010:steady growth and changing composition. *Monthly Labor Review*, **124**:21–38.

Furuya S, Natori Y, Ikeda R (2003) Asbestos in Japan. *International Journal Occupational Environmental Health*, **9**:260–265.

Garg A (1992) Occupational biomechanics and low-back pain. *Occupational Medicine: State of the Art Review*, **7**:609–628.

Gomes J, Lloyd O, Norman N, Pahwa P (2001) Dust exposure and impairment of lung function at a small iron foundry in a rapidly developing country. *Occupational and Environmental Medicine*, **58**:656–662.

Guerrero JL, Sniezek JE, Sehgal (1999) The prevalence of disability from chronic conditions due to injury among adults ages 18–69 years: United States, 1994. *Disability and Rehabilitation*, **21**:187–192.

Guo H-R, Tanaka S, Cameron LL et al (1995) Back pain among workers in the United States: national estimates and workers at high risk. *American Journal of Industrial Medicine,* **28**:591–602.

Guo H-R, Tanaka S, Halperin WE, Cameron LL (1999) Back pain prevalence in U.S. industry and estimates of lost workdays. *American Journal of Public Health*, **89**:1029–1035.

Hagberg M, Silverstein BA, Wells RP et al., eds (1995) *Work-related musculoskeletal disorders (WMSD): a handbook for prevention*. Taylor and Francis. London.

Hales TR, Bernard BP (1996) Epidemiology of work-related musculoskeletal disorders. *Orthopedic Clinics of North America*, **27**:679–709.

Harrison JE, Frommer MS, Ruck EA, Blyth FM (1989) Deaths as a result of work-related injury in Australia, 1982–1984. *Medical Journal of Australia*, **50**:118–125.

Hayes R, Yin S, Dosemeci M et al. (1997) Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine—National Cancer Institute Benzene Study Group. *Journal of the National Cancer Institute*, **89**:1065–1071.

Heistaro S, Vartiainen E, Heliövaara M et al. (1998) Trends of back pain in eastern Finland, 1972–1992, in relation to socioeconomic status and behavioral risk factors. *American Journal of Epidemiology*, **148**:671–682.

Hendrick DJ (1996) Occupational and chronic obstructive pulmonary disease (COPD). *Thorax* **51**:947–955.

Hernandez-Gaytan SI, Santos-Burgo AC, Becker-Meyer JP, Macias-Carrillo C, Lopez-Cervantes M (2000) Prevalence of hearing loss and correlated factors in a cement plant. *Salud Publica de Mexico*, **42**:106–111.

Hessel P (2000) Hearing loss among construction workers in Edmonton, Alberta, Canada. *Journal of Occupational and Environmental Medicine*, **42**:57.

Hessel PA, Sluis-Cremer GK (1987) Hearing loss in white South African gold-miners. *South African Medical Journal*, **71**:364–367.

Hewett P (1996) Interpretation and use of occupational exposure limits for chronic disease agents. *Occupational Medicine: State of the Art Reviews*, **11**:561–590.

Hildebrandt VH (1995) Back pain in the working population: prevalence rates in Dutch trades and professions. *Ergonomics*, **38**:1283–1298.

Hnizdo E, Sluis-Cremer G (1991) Silica exposure, silicosis, and lung cancer: a mortality study of South African gold miners. *British Journal Industrial Medicine*, **48**:53–60.

Hollman S, Klimmer F, Schmidt K-H et al. (1999) Validation of a questionnaire for assessing physical work load. *Scandinavian Journal of Work, Environment and Health*, **25**:105–114.

Holte HH, Tambs K, Bjerkedal T (2000) Manual work as predictor for disability pensioning with osteoarthritis among the employed in Norway 1971–1990. *International Journal of Epidemiology*, **29**:487–494.

Hoogendoorn WE, van Poppel MNM, Bongers PM, Koes BW, Bouter LM (1999) Physical load during work and leisure time as risk factors for back pain. *Scandinavian Journal of Work, Environment and Health*, **25**:387–403.

Hulshof CTJ, Veldhuijzen van Zanten B (1987) Whole-body vibration and low-back pain: a review of epidemiologic studies. *International Archives of Occupational and Environmental Health*, **59**:205–220.

Human Resources Development Canada (2000) *Surveys until 2000*. Available at http://www.hrdc-drhc.gc.ca/hrib/hrp-prh/ssd-des/english/glossary/survey.shtml.

IARC (1977) *Asbestos*. (Monograph 14.) International Association of Cancer Registries, Lyon.

IARC (1980) *Some metals and metallic compounds*. (Monograph 23.) International Association of Cancer Registries, Lyon.

IARC (1987a) *Overall evaluations of carcinogenicity: an updating of IARC.* (Monographs 1–42, Supplement 7.) International Association of Cancer Registries, Lyon.

IARC (1987b) *Silica and some silicates.* (Monograph 42.) International Association of Cancer Registries, Lyon.

IARC (1989) *Diesel and gasoline engine exhausts and some nitroarenes.* (Monograph 46.) International Association of Cancer Registries, Lyon.

IARC (1990a) *Chromium, nickel and welding.* (Monograph 49.) International Association of Cancer Registries, Lyon.

IARC (1990b) *Cancer.* IARC Scientific Publications. International Association of Cancer Registries, Lyon.

IARC (1993) *Beryllium, cadmium, mercury, and exposures in the glass manufacturing industry.* (Monograph 58.) International Association of Cancer Registries, Lyon.

IARC (1997) *Silica, some silicates, coal dust and para-aramid fibrils.* (Monograph 68.) International Association of Cancer Registries, Lyon.

IARC (2000) *Ionizing radiation, part 1: X- and gamma (g)-radiation, and neutrons.* (Monograph 75.) International Association of Cancer Registries, Lyon.

IARC (2001) *Overall evaluation of carcinogenicity to humans.* International Association of Cancer Registries, Lyon. Available at http://193.51.164.11/monoeval/crthall.html.

IARC (2002) *IARC Monographs Programme on the Evaluation of Carcinogenic Risks to Humans.* Available at http://193.51.164.11.

ICNCM (1990) Report of the International Committee on Nickel Carcinogenesis in Man. *Scandinavian Journal of Work, Environment and Health*, **16**:1–84.

ICRP (1991) *Recommendations of the International Commission on Radiological Protection.* (ICRP Publication No. 60. Vol 21:1–3.) International Commission on Radiological Protection, Oxford.

ILO (1987) *Resolutions adopted by the International Conference of Labour Statisticians.* Annex 1. Major, sub-major, minor and unit group titles. International Labour Organization. Available at http://www.ilo.org/public/english/bureau/stat/download/res/isco.pdf.

ILO (1995a) *Yearbook of labour statistics 1995, 54th issue.* International Labour Organization, Geneva.

ILO (1995b) *Sectoral activities programme, recent developments in the coal-mining industry, coal mines committee, 13th session, Report I.* International Labour Organization, Geneva.

ILO (1996) ILO *Yearbook of labour statistics 1996, 55th issue.* International Labour Organization, Geneva.

ILO (2000) *Yearbook of labour statistics, 59th issue.* International Labour Organization, Geneva.

ILO (2001) Laborsta, the labour statistics database. International Labour Organization. Available at http://laborsta.ilo.org/.

ILO (2002a) *Economically active population 1950–2010, 4th edn, rev. 2*. International Labour Organization, Geneva.

ILO (2002b) *The International Labor Conference, Report VI. Decent work and the informal economy*. International Labour Organization. Geneva.

IPCS (1998) *Chrysotile asbestos. Environmental health criteria 203*. International Programme on Chemical Safety, Geneva.

Jarup L, Nordberg G (1998) Cancer. In: *Health effects of cadmium exposure: a review of the literature and a risk estimate*. Jarup L, Berglund M, Elinder CG, Nordberg G, Vahter M, eds. *Scandinavian Journal Work, Environment and Health*, **24**:34–36.

Jensen RC (1988) Epidemiology of work-related back pain. *Topics Acute Care Trauma Rehabilitation*, **23**:1–15.

Jin K, Sorock GS, Courtney T et al (2000) Risk factors for work-related low back pain in the People's Republic of China. *International Journal of Occupational and Environmental Health*, **6**:26–33.

Johanning E (1991) Back disorders and health problems among subway train operators exposed to whole-body vibration. *Scandinavian Journal of Work Environment and Health*, **17**:414–419.

Johanning E, Wilder D, Landrigan P (1991) Whole-body vibration exposure in subway cars and review of adverse health effects. *Journal of Occupational Medicine*, **33**:605–612.

Johnson AG, Whyte WF (1977) The Mondragon system of worker production cooperatives. *Industrial and Labor Relations Review*, **31**:18–30.

Johnson D, Swift D (1997) Sampling and sizing particles. In: *The occupational environment—its evaluation and control*. Di Nardi SR, ed. The American Industrial Hygiene Association, Fairfax, VA.

Joshi TK, Menon KK, Kishore J (2001) Musculoskeletal disorders in industrial workers of Delhi. *International Journal of Occupational and Environmental Health*, **7**:217–222.

Juengprasert W (1997) Towards elimination of silicosis in Thailand. Mineral dusts and prevention of silicosis. *Asian-Pacific Regional Network on Occupational Safety and Health Information (ASIA-OSH)*, **4**(2).

Kam Lam TF (2000) *Occupational safety and health in China*. Asian labour update. Available at http://www.amrc.org.hk/alu/Alu39/013906.htlm.

Kamal AA, Eldamati SE, Faris F (1989) Hearing threshold of Cairo traffic policemen. *International Archives of Occupational and Environmental Health*, **61**:543–545.

Karjalainen A, Kurppa K, Martikanen R, Karjalainen J, Klaukka T (2002) Exploration of asthma risk by occupation -extended analysis of an incidence study of the Finnish population. *Scandinavian Journal of Work, Environment and Health*, **28**:49–57.

Karjalainen A, Kurppa K, Martikanen R, Klaukka T, Karjalainen J (2001) Work is related to a substantial portion of adult-onset asthma incidence in the Finnish population. *American Journal of Respiratory Critical Care Medicine*, **164:**565–568.

Kauppinen T, Toikkanen J, Pedersen D et al. (2000) Occupational exposure to carcinogens in the European Union. *Occupational and Environmental Medicine*, **57**:10–18

Khogali M (1970) Industrial health problems of developing countries. 1.1. Types of risk and their prevention. *Journal of Tropical Medicine and Hygiene*, **73**:269–274.

Klein BP, Jensen RC, Sanderson LM (1984). Assessment of workers' compensation claims for back strains/sprains. *Journal of Occupational Medicine*, **26**:443–448.

Klockhoff I, Drettner B, Svedberg A (1974) Computerized classification of the results of screening audiometry in groups of persons exposed to noise. *Audiology*, **13**:326–334.

Koch JL, Rhodes, SR (1981) Predictors of turnover of female factory workers. *Journal of Vocational Behavior*, **18**:145–161.

Kogevinas M, Antó JM, Soriano JB, Tobias A (1996) The risk of asthma attributable to occupational exposures: a population-based study in Spain. *American Journal of Respiratory Critical Care Medicine*, **154**:137–143.

Kogevinas M, Anto JM, Sunyer J, Tobias A, Kromhout H, Burney P (1999) Occupational asthma in Europe. *The Lancet*, **353**:1750–1754.

Korn RJ, Dockery DW, Speizer FE, Ware JH, Ferris BG Jr (1987) Occupational exposures and chronic respiratory symptoms: a population based study. *American Review of Respiratory Disease*, **136**:298–304.

Krzyzanowski M, Jedrychowski W, Wysocki M (1986) Factors associated with the change in ventilatory function and the development of chronic obstructive pulmonary disease in a 13-years follow-up of the Cracow study. *American Review of Respiratory Disease*, **134**:1011–1019.

Kumar A, Varghese M, Mohan D, Mahajan P, Gulati P, Kale S (1999) Effect of whole-body vibration on the low back. *Spine*, **24**:2506–2515.

Kuwashima A, Aizawa Y, Nakamura K, Taniguchi S, Watanabe M (1997) National survey on accidental low back pain in workplace. *Industrial Health*, **35**:187–193.

Lagerström M, Hansson T, Hagberg M (1998) Work-related low back problems in nursing. *Scandinavian Journal of Work, Environment and Health*, **24**: 449–464.

Lahiri S (2002) *Analysis of cost-effectiveness of interventions to reduce low back pain as a consequence of occupational ergonomic stressors*. World Health Organization, Geneva.

Lawrence RC, Helmick CG, Arnett FC et al. (1998) Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis and Rheumatism*, **41**:778–799.

Lee P (2001) Relation between exposure to asbestos and smoking jointly and the risk of lung cancer. *Occupational and Environmental Medicine*, **58**:145–153.

Leigh J, Driscoll T (2003) Malignant mesothelioma in Australia, 1945–2002. *International Journal Occupational Environmental Health*, **9**:206–217.

Leigh JP, Sheetz RM (1989) Prevalence of back pain among full-time United States workers. *British Journal of Industrial Medicine*, **46**:651–657.

Leino PI, Hänninen V (1995) Psychosocial factors at work in relation to back and limb disorders. *Scandinavian Journal of Work, Environment and Health*, **21**:134–142.

Leino-Arjas P, Hanninen K, Puska P (1998) Socioeconomic variation in back and joint pain in Finland. *European Journal of Epidemiology*, **14**:79–87.

Lerer LB, Myers JE (1994) Application of two secondary documentary sources to identify the underreporting of fatal occupational injuries in Cape Town, South Africa. *American Journal of Industrial Medicine*, **26**:521–527.

Lipsett M, Campleman S (1999) Occupational exposure to diesel exhaust and lung cancer: a meta-analysis. *American Journal of Public Health*, **89**:1009–1017.

Loewenson R (1998) Health impact of occupational risks in the informal sector in Zimbabwe. *International Journal of Occupational and Environmental Health*, **4**:264–274.

Lombardo LJ, Balmes JR (2000) Occupational asthma: a review. *Environmental Health Perspectives*, **8**:S697–704.

Loney P, Stratford P (1999) The prevalence of low back pain in adults: a methodological review of the literature. *Physical Therapy*, **79**:384–396.

Lucifora C (1998) The impact of unions on labour turnover in Italy: evidence from establishment level data. *International Journal of Industrial Organization*, **16**:353–376.

Lynge E, Anttila A, Hemminki K (1997) Organic solvents and cancer. *Cancer Causes and Control*, **8**:406–419.

MacDonald LA, Karasek RA, Punnett L, Scharf T (2001) Covariation between workplace physical and psychosocial stressors: evidence and implications for occupational health research and prevention. *Ergonomics*, **44**:696–718.

Magnusson ML, Pope MH, Wilder DG,Areskoug B (1996) Are occupational drivers at an increased risk for developing musculoskeletal disorders? *Spine*, **6**:710–717.

Malchaire J (2000) Strategy for prevention and control of the risks due to noise. *Occupational and Environmental Medicine*, **57**:361–369.

Malo J, Chan-Yeung (2001) Occupational asthma. *Journal of Allergy and Clinical Immunology*, **108**:317–328.

Marras WS, Allread DL, Burr DL, Fathallah FA (2000). Prospective validation of low-back disorder risk model and assessment of ergonomic interventions associated with manual materials handling tasks. *Ergonomics*, **43**: 1866–1886.

Marsh SM, Layne LA (2001) *Fatal injuries to civilian workers in the United States, 1980–1995 (National profile)*. Department of Health and Human Services. National Institute for Occupational Safety and Health (DHHS/NIOSH) 2001-129. Available at http://www.cdc.gov/niosh/NTOF2000/2001129.htlm.

Martonik JF, Nash E, Grossman E (2001) The history of OSHA's asbestos rule-makings and some distinctive approaches that they introduced for regulating occupational exposure to toxic substances. *American Industrial Hygiene Association Journal*, **62**:208–217.

Miedema HS, Chorus AMJ, Wevers CWJ, van der Linden S (1998) Chronicity of back problems during working life. *Spine*, **23**:2021–2029.

Milton DK, Solomon DM, Roseillo RA, Herrick RF (1998) Risk and incidence of asthma attributable to occupational exposure among HMO members. *American Journal of Industrial Medicine*, **33**:1–10.

Minino AM, Smith BL (2001) Deaths: preliminary data for 2000. *National Vital Statistics Reports*, **49**:1–40.

Ministerio de Trabajo y Asuntos Sociales (2002) [Ministry of Labour and Social Affairs (2002)] Gabinete de Comunicación. Información Estadística. [Communications Bureau. Statistical Information.] Available at http://www.mtas.es/. Informe de Resultados España [Report of results for Spain].

Moffett M (2002) *New Mexico labor market dynamics. Labor market review, special article*. New Mexico Department of Labor, February. Available at http://www3.state.nm.us/dol/dol_lmrsa6.htm.

Monso E, Munoz-Rino F, Izquierdo J et al. (1998) Occupational asthma in the community: risk factors in a western Mediterranean population. *Archives of Environmental Health*, **53**:93–98.

Morata TC, Fiorini AC, Fischer FM et al. (1997) Toluene-induced hearing loss among rotogravure printing workers. *Scandinavian Journal of Work, Environment and Health*, **23**:289–298.

Morken T, Moen B, Riise T et al (2000) Prevalence of musculoskeletal symptoms among aluminum workers. *Occupational Medicine*, **50**:141–421.

Moselhi M, El-Sadik YM, El-Dakhakhny A (1979) A six-year follow up study for evaluation of the 85 dBA safe criterion for noise exposure. *American Industrial Hygiene Association Journal*, **40**:424–426.

Mukherjee A K, Nag DP, Kakde Y, Prakash MN, Rao SR (1995) Noise level monitoring in a watch factory in Bangalore. *Indian Journal of Industrial Medicine*, **41**:42–44.

Muruka AO (1998) Age, height and duration in of service in relation to back pain among tea pickers in Kenya. *African Newsletter for Occupational Health and Safety*, **8**(3). Accessed at http://www.occuphealth.fi/e/info/anl/398/ in May 2002.

Myers JE, Lewis P, Hofmeyr W (1989) Respiratory health of brickworkers in Cape Town, South Africa: background, aims, and dust exposure determinations. *Scandinavian Journal of Work, Environment and Health*, **15**:180–187.

Nachemson A (1999) Back pain: delimiting the problem in the next millennium. *International Journal of Law and Psychiatry*, **22**:473–490.

Nachemson A, Jonsson E, eds (2000) *Neck and back pain: the scientific evidence of causes, diagnosis, and treatment*. Lippincott, Williams and Wilkins, Philadelphia, PA.

National Research Council (2001) *Musculoskeletal disorders and the workplace: low back and upper extremities*. Panel on musculoskeletal disorders and the workplace. Commission on Behavioral and Social Sciences and Education, National Research Council and Institute of Medicine. National Academy Press, Washington, DC. (also accessed at http://www.nap.edu/openbook).

National Statistics Office Philippines (2000*) Proceedings of the seminar on the ILO project on occupational injuries, 6 December*. National Statistics Office, Manila.

Nelson DI (1997) Risk assessment in the workplace. In: *The occupational environment—its evaluation and control*. Di Nardi SR, ed. The American Industrial Hygiene Association, Fairfax, VA.

Ng TP, Hong CY, Goh LG, Wong ML, Koh KT, Ling SL (1994) Risks of asthma associated with occupations in a community-based case-control study. *American Journal of Industrial Medicine*, **25**:709–718.

NIEHS (1999) United States. A world-class program. National Institute on Environmental Health Sciences. *Environmental Health Perspectives*, **107**(7). Available at http://ehpnet1.niehs.nih.gov/docs/1999/107-7/niehsnews.html.

NIOSH (1991) *NIOSH publications on noise and hearing*. Perspectives in disease prevention and health promotion, leading work-related diseases and injuries—United States. *Noise-induced Loss of Hearing*, July. National Institute for Occupational Safety and Health, Washington, DC.

NIOSH (1996) *Violence in the workplace: risk factors and prevention strategies*. National Institute for Occupational Safety and Health (NIOSH-CDC), Cincinnati, OH. Available at http://www.cdc.gov/niosh/violpr.htlm.

NIOSH (1998) *Criteria for a recommended standard: occupational noise exposure. Revised criteria 1998*. National Institute for Occupational Safety and Health, Cincinnati, OH. Available at http://www.cdc.gov/niosh/98-126.html.

NIOSH (1999) *Work-related lung disease surveillance report 1999*. U.S. Department of Health and Human Services, National Institute for Occupational Safety and Health, Washington, DC.

NIOSH (2000a) *Worker health chartbook, 2000*. NIOSH-CDC. US Department of Health and Human Services, National Institute for Occupational Safety and Health, Washington, DC.

NIOSH (2000b) *Injuries, illnesses, and hazardous exposures in the mining industry, 1986–1995: a surveillance report*. (No. 2000-117.) National Institute for Occupational Safety and Health, Washington, DC.

NOHSC (1993) *National code of practice for management and protection of hearing at work* (NOHSC:2009.) National Occupational Health and Safety Commission, Australia. Available at http://www.nohsc.gov.au/OHSInformation/NOHSCPublications/fulltext/docs/h3/19.htm.

NOHSC (2001) National Occupational Health and Safety Commission, Australia. Available at http://www.nohsc.gov.au.

NOHSC (2002) *Statistics*. National Occupational Health and Safety Commission, Australia. Available at http://www.nohsc.gov.au/statistics/default.htm.

Noweir MH, El-Dakhakhny AA, Valic F (1968) Exposure to noise in the textile industry of the U.A.R. *American Industrial Hygiene Association Journal*, **29**:541–546.

Nurminen M, Karjalainen A (2001) Epidemiologic estimate of the proportion of fatalities related to occupational factors in Finland. *Scandinavian Journal Work, Environment and Health*, **27**:161–213.

Obiako MN (1979) Deafness and the mining industry in Zambia. *East African Medical Journal*, **56**:445–449.

Oleru UG (1980) Comparison of the hearing levels of Nigerian textile workers and a control group. *American Industrial Hygiene Association Journal*, **41**:283–287.

Oleru UG, Ijaduola GTA, Sowho EE (1990) Hearing thresholds in an auto assembly plant: prospects for hearing conservation in an [sic] Nigerian factory. *International Archives of Occupational and Environmental Health*, **62**:199–202.

Omokhodion FO, Umar US, Ogunnowo BE (2000) Prevalence of low back pain among staff in a rural hospital in Nigeria. *Occupational Medicine*, **50**:107–110.

Ory FG, Rahman FU, Katagade V, Shukla A, Burdorf A (1997) Respiratory disorders, skin complaints, and low-back trouble among tannery workers in Kanpur, India. *American Industrial Hygiene Association Journal*, **58**: 740–746.

Osibogun A, Igweze IA, Adeniran LO (2000). Noise-induced hearing loss among textile workers in Lagos Metropolis. *The Nigerian Postgraduate Medical Journal*, **7**:104–111.

Oxman A, Muir D, Shannon H, Stock S, Hnizdo E, Lange H (1993) Occupational dust exposure and chronic obstructive pulmonary disease. A systematic overview of the evidence. *American Review of Respiratory Disease*, **148**:38–48.

Ozguler A, Leclerc A, Landre M, Pietri-Taleb F, Niedhammer I (2000) Individual and occupational determinants of low back pain according to various definitions of low back pain. *Journal of Epidemiology and Community Health*, **54**:215–220.

PAHO/WHO (1998) *Informe del proyecto sistematización de datos básicos sobre salud de los trabajadores en países de las Américas—Agosto [Report of the project: systematisation of basic data of health of workers in LAC countries—August]*. División del Ambiente y Salud [Division for Environment and Health]. Pan American Health Organization/World Health Organization.

Palmer K, Pannett B, Griffin M (2001) *Occupational exposure to noise and hearing difficulties in Great Britain*. (Contract Report 361/2001.) University of Southampton for the Health and Safety Executive, Southhampton.

Paquet VL, Punnett L, Buchholz BO (1999) An evaluation of manual materials handling in highway construction work. *International Journal of Industrial Ergonomics*, **24**:431–444.

Partanen T, Jaakkola J, Tossavainen A (1995) Silica, silicosis and cancer in Finland. *Scandinavian Journal of Work, Environment and Health*, **2**:S84–86.

Partridge REH, Duthie JJR (1968) Rheumatism in dockers and civil servants: a comparison of heavy manual and sedentary workers. *Annals of Rheumatic Diseases*, **27**:559–568.

Pearce N, Matos E, Vainio H, Boffetta P, Kogevinas M (1994) *Occupational cancer in developing countries*. (IARC Scientific Publications No. 129) Lyon.

Perkio-Makela M, Riihimäki H (1997) Intervention on seat adjustment among drivers of forest tractors. *International Journal of Industrial Ergonomics*, **19**:231–237.

Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R (2000). Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *British Medical Journal*, **321**: 323–329.

Peto J, Decarli A, La Vecchia C, Levi F, Negri E (1999) The European mesothelioma epidemic. *British Journal of Cancer*, **79**:666–672.

Phan Hong Son, Tran Ngoc Lan, Le Van Trung et al. (1999) *Use geographical information system to identify risk for silicosis and planning strategies for prevention, July, 1999*. Funded by the Fogarty International Center, NIH. Washington, DC.

Pope MH, Andersson GBJ, Frymoyer JW, Chaffin DB, eds (1991) *Occupational low back pain: assessment, treatment and prevention*. Mosby-Year Book, Inc., St. Louis, MO.

Price B, Ware A (2004) Mesothelioma trends in the United States: an update based on surveillance, epidemiology, and End Results program. Data for 1973 through 2003. *American Journal of Epidemiology*, **159**:107–112.

Prince MM, Stayner LT, Smith RJ, Gilbert S J (1997) A re-examination of risk estimates from the NIOSH Occupational Noise and Hearing Survey (ONHS) . *Journal of the Acoustic Society of America*, **101**:950–963.

Punnett L (1998) Ergonomic stressors and upper extremity disorders in vehicle manufacturing: cross-sectional exposure-response trends. *Occupational and Environmental Medicine*, **55**:414–420.

Punnett L (1999) The costs of work-related musculoskeletal disorders in automotive manufacturing. *New Solutions: A Journal of Environmental and Occupational Health Policy*, **9**:403–426.

Punnett L (2002) *Ergonomics and public health. Oxford textbook of public health*. 4th edn. Oxford University Press, Oxford.

Punnett L, Fine LJ, Keyserling WM, Herrin GD, Chaffin DB (1991) Back disorders and non-neutral trunk postures of automobile assembly workers. *Scandinavian Journal of Work, Environment and Health*, **17**:337–346.

Punnett L, Herbert R (2000) Work-related musculoskeletal disorders: is there a gender differential, and if so, what does it mean? In: *Women and health*. Goldman MB, Hatch MC, eds. Academic Press. San Diego, CA.

Quinlan M (2002) *Workplace health and safety effects of precarious employment*. In the Global Occupational Health Network Newsletter, Issue No. 2. World Health Organization, Geneva.

Raja S, Ganguly G (1983) Impact of exposure to noise on the hearing acuity of employees in a heavy engineering industry. *Indian Journal of Medical Research*, **78**:100–113.

Rees D, Cronje R, du Toit RSJ (1992) Dust exposure and pneumoconiosis in a South African pottery. 1. Study objectives and dust exposure. *British Journal of Industrial Medicine*, **49**:459–464.

Riihimäki H (1991) Low-back pain, its origin and risk indicators. *Scandinavian Journal of Work, Environment and Health*, **17**:81–90.

Riihimäki H (1995a) Back and limb disorders. In: *Epidemiology of work related diseases*. McDonald C, ed. BMJ Publishing Group, London.

Riihimäki H (1995b) Hands up or back to work- future challenges in epidemiologic research on musculoskeletal diseases. *Scandinavian Journal of Work, Environment and Health*, **21**:401–403.

Riihimäki H, Mattsson T, Zitting AJ et al. (1990) Radiographically detectable degenerative changes of the lumbar spine among concrete reinforcement workers and house painters. *Spine*, **15**:114–119.

Riihimäki H, Tola S, Videman T, Hänninen K (1989) Low-back pain and occupation: a cross-sectional questionnaire study of men in machine operating, dynamic physical work, and sedentary work. *Spine*, **14**:204–209.

Riihimäki H, Viikari-Juntura E, Moneta G, Kuha J, Videman T, Tola S (1994) Incidence of sciatic pain among men in machine operating, dynamic physical work and sedentary work. *Spine*, **19**:138–142.

Roach S (1992) *Health risks from hazardous substances at work—assessment, evaluation and control*. Pergamon Press, New York.

Schierhout GH, Midgley A, Myers JE (1997) Occupational fatality underreporting in rural areas of the Western Cape Province, South Africa. *Safety Science*, **25**:113–122.

Sears MR (1997) Descriptive epidemiology of asthma. *The Lancet*, **350**:S1–4.

Shaikh GH (1996) Noise problem in a polyester fiber plant in Pakistan. *Industrial Health*, **34**:427–431.

Shannon HS, Mayr J, Haines T (1997) Overview of the relationship between organizational and workplace factors and injury rates. *Safety Science*, **26**:201–17.

Shannon HS, Walters V, Lewchuk W et al. (1996) Workplace organizational correlates of lost-time accident rates in manufacturing. *American Journal of Industrial Medicine*, **29**:258–268.

Silverstein BA, Fine LJ, Armstrong TJ (1986) Hand wrist cumulative trauma disorders in industry. *British Journal of Industrial Medicine*, **43**:779–784.

Silverstein BA, Fine LJ, Stetson DS (1987) Hand-wrist disorders among investment casting plant workers. *The Journal of Hand Surgery*, **12A**:838–844.

Silverstein BA, Stetson DS, Keyserling WM, Fine LJ (1997) Work-related musculoskeletal disorders: comparison of data sources for surveillance. *American Journal of Industrial Medicine*, **31**:600–608.

Silverstein BA, Viikari-Juntura E, Kalat J (2002) Use of a prevention index to identify industries at high risk for work-related musculoskeletal disorders of the neck, back, and upper extremity in Washington state, 1990–1998. *American Journal of Industrial Medicine*, **41**:149–169.

Simpson M, Bruce R (1981) *Noise in America: the extent of the noise problem*. (EPA Report No. 550/9-81-101). Environmental Protection Agency, Washington, DC.

Singapore Government (2000) Statistics Singapore—statistical resources—statistical classifications—Singapore standard classification (SSOC), online portal available at http://www.gov.sg.

Singh N, Davis GS (2002) Review: occupational and environmental lung disease. *Current Opinions on Pulmonary Medicine*, **8**:117–125.

Skovron ML, Szpalski M, Nordin M, Melot C, Cukier D (1994) Sociocultural factors and back pain: A population-based study in Belgian adults. *Spine*, **19**:129–137.

Sluiter JK, Rest KM, Frings-Dresen MHW (2000) *Criteria document for evaluation of the work-relatedness of upper extremity musculoskeletal disorders*. SALTSA Joint Programme for Working Life Research in Europe and Academic Medical Center, University of Amsterdam, Amsterdam.

Smedley J, Egger P, Cooper C, Coggon D (1995) Manual handling activities and risk of low back pain in nurses. *Occupational and Environmental Medicine*, **52**:160–3

Sobti A, Cooper C, Inskip H et al. (1997) Occupational physical activity and long-term risk of musculoskeletal symptoms: a national survey of post office pensioners. *American Journal of Industrial Medicine*, **32**:76–83.

Soutar C, Robertson A, Miller B, Searl A, Bignon J (2000) Epidemiological evidence on the carcinogenicity of silica: factors in scientific judgement. *Annals of Occupational Hygiene*, **44**:3–14.

Sriwattanatamma P, Breysse P (2000) Comparison of NIOSH noise criteria and OSHA hearing conservation criteria. *American Journal of Industrial Medicine*, **37**:334–338.

Stayner L, Smith R, Thun M, Schnorr T, Lemen R (1992) A dose response analysis and quantitative assessment of lung cancer risk and occupational cadmium exposure. *Annals of Epidemiology*, **2**:177–194.

Steenland K, Beaumont J, Elliot L (1991a) Lung cancer in mild steel welders. *American Journal of Epidemiology*, **133**:220–229.

Steenland K, Burnett C, Ward E, Lalich N, Hurrell J (2003) Dying for work: the magnitude of US mortality from selected causes of death associated with occupation. *American Journal of Industrial Medicine*, **43**:461–482.

Steenland K, Loomis D, Shy C, Simonsen N (1996) Review of occupational lung carcinogens. *American Journal of Industrial Medicine*, **29**:474–490.

Steenland K, Sanderson W (2001) Lung cancer among industrial sand workers exposed to crystalline silica. *American Journal of Epidemiology*, **53**:695–703.

Steenland K, Sanderson W, Calvert GM (2001b) Kidney disease and arthritis in a cohort study of workers exposed to silica. *Epidemiology*, **12**:405–412.

Steenland K, Stayner L, Greife A et al. (1991b) Mortality among workers exposed to ethylene oxide. *New England Journal of Medicine*, **324**: 1402–1407.

Steenland K, Ward E (1991) Lung cancer incidence among patients with beryllium disease: A cohort mortality study. *Journal of the National Cancer Institute*, **83**:1380–1385.

Suter A (2000) Standards and regulations. In: *The noise manual, 5th edn*, Berger EH, Royster LH, Rozster JD, Driscoll DP, Layne M, eds. American Industrial Hygiene Association, Fairfax, VA.

Tay P (1996) Severe noise-induced deafness: a 10-year review of cases. *Singapore Medical Journal*, **37**:362–364.

Toren K (1996) Self-reported rate of occupational asthma in Sweden 1990–1992. *Occupational and Environmental Medicine*, **53**:757–761.

Toren K, Balder B, Brisman J et al. (1999) This risk of asthma. *European Respiratory Journal*, **13**:496–501.

Toroptsova NV, Benevolenskaya LI, Karyakin AN, Sergeev IL, Erdesz S (1995) "Cross/sectional" study of low backpain among workers at an industrial enterprise in Russia. *Spine*, **20**:328–332.

Tsai SP, Gilstrap EL, Cowles SR, Waddell LC, Ross CE (1992) Personal and job characteristics of musculoskeletal injuries in an industrial population. *Journal of Occupational Medicine*, **34**:606–612.

UN (2000) *International standard industrial classification of all economic activities (ISIC). Third revision*. United Nations publication (St/ESA/STAT/SER.M/4/Rev.3). United Nations, New York.

UN (2001) *World population prospects. The 2000 revision highlights*. Population Division. Department of Economic and Social Affairs, United Nations, New York.

USDHHS (1986) Perspectives in disease prevention and health promotion, leading work-related diseases and injuries—United States. *Morbidity and Mortality Weekly Report*, **35**(12). (Noise-induced Loss of Hearing. Reprinted in USDHHS 1991, *NIOSH Publications on Noise and Hearing*.)

USDOL OSHA (2000) *Docket for the Federal Register.* (Vol. 65, No. 220.) U.S. Department of Labor, Occupational Safety and Health Administration, Washington, DC.

USDOL OSHA (2002a) *Permissible exposure limits, codified at 29 CFR 1910.1000*. U.S. Department of Labor, Occupational Safety and Health Administration. Available at http://www.osha.gov/SLTC/pel/index.html.

USDOL OSHA (2002b) *Noise and hearing conservation*. U.S. Department of Labor, Occupational Safety and Health Administration. Available at http://www.osha-slc.gov/SLTC/noisehearingconservation/index.html

USEIA (2001) U.S. Energy Information Administration, U.S. Department of Energy, International Energy Database. Accessed January 2001 at www.eia.doe.gov/emeu/iea/coal.html.

Vaaranen V, Vasama M, Toikkanen J et al. (1994) *Occupational diseases in Finland l993*. Institute of Occupational Health, Helsinki.

Venables KM, Chang-Yeung M (1997) Occupational asthma. *The Lancet*, **349**:1465–1469.

Videman T, Nurminen M, Troup JDG (1990) Lumbar spine pathology in cadaveric material in relation to history of back pain, occupation, and physical loading. *Spine*, **15**:728–740.

Viikari-Juntura E (1997) The scientific basis for making guidelines and standards to prevent work-related musculoskeletal disorders. *Ergonomics*, **40**:1097–1117.

Viikari-Juntura E, Riihimäki H (1999) New avenues in research on musculoskeletal disorders. *Scandinavian Journal of Work, Environment and Health*, **25**:564–568.

Volinn E (1997) The epidemiology of low back pain in the rest of the world. A review of surveys in low-and middle-income countries. *Spine*, **22**:1747–1754.

Volinn E, Punnett L (2001) Point of view. *Spine*, **26**:1902–1903.

Volinn E, Spratt KF, Magnusson M, Pope MH (2001) The Boeing prospective study and beyond. *Spine*, **26**:1613–1622.

Waddell G (1991) Low back disability: a syndrome of Western civilization. *Neurosurgery Clinics of North America*, **2**:719–738.

Wagner GR, Wegman DH (1998) Occupational asthma: prevention by definition. *American Journal of Industrial Medicine*, **33**:427–429.

Waitzman N, Smith K (1999) *Unsound conditions: work-related hearing loss in construction, 1960–75*. The Center to Protect Worker's Rights. CPWR Publications, Washington, DC.

Ward E, Okan A, Ruder A, Fingerhut M, Steenland K (1992) A mortality study of workers at seven beryllium processing plants. *American Journal of Industrial Medicine*, **22**:885–904.

Westgaard RH, Jansen T (1992) Individual and work related factors associated with symptoms of musculoskeletal complaints: I. A quantitative registration system. *British Journal of Industrial Medicine*, **49**:147–153.

Westgaard RH, Winkel J (1997) Ergonomic intervention research for improved musculoskeletal health: a critical review. *International Journal of Industrial Ergonomics*, **20**:463–500.

WHO (1999) *International statistical classification of diseases and related health problems (ICD-10) in occupational health*. Karjalainen A, ed. Protection of the Human Environment, Occupational and Environmental Health Series, (WHO/SDE/OEH/99.11). World Health Organization, Geneva.

WHO (2001) *Electromagnetic fields and public health: extremely low frequency fields and cancer*. (Fact Sheet No. 263). World Health Organization, Geneva. Available at http://www.who.int/inf-fs/en/fact263.html.

WHO/FIOSH (2001) *Occupational exposure to noise: evaluation, prevention and control*. Goelzer B, Hansen CH, Sehrndt GA, eds. On behalf of WHO by the Federal Institute for Occupational Safety and Health (FIOSH), Dortmund.

Wikstrom B-O, Kjellberg A, Landstrom U (1994) Health effects of long-term occupational exposure to whole-body vibration: a review. *International Journal of Industrial Ergonomics*, **14**:273–292.

World Bank (2001) *World development indicators 2001*. Available at http://worldbank.com.

Xia Z, Courtney TK, Sorock GS, Zhu J, Fu H, Liang Y, Christiani DC (2000) Fatal occupational injuries in a new development area in the People's Republic of China. *Occupational and Environmental Medicine*, **42**: 917–922.

Xu X, Christiani DC, Dockery DW et al. (1992) Exposure-response relationships between occupational exposures and chronic respiratory illness: a community-based study. *American Review of Respiratory Disease*, **146**:413–418.

Yin SN, Li Q, Liu Y, Tian F, Du C, Jin C (1987) Occupational exposure to benzene in China. *British Journal of Industrial Medicine*, **44**:192–195.

Yip YB (2001) A study of work stress, patient handling activities and the risk of low back pain among nurses in Hong Kong. *Journal of Advanced Nursing*, **36**:794–804.

Zou Changqi, Gao Yun, Ma Qingyan (1997) Pneumoconiosis in China: current situation and countermeasures. *Mineral dusts and prevention of silicosis*, **4**(2) September 1997. Asian-Pacific Regional Network on Occupational Safety and Health Information (ASIA-OSH). Available at http://www.ilo.org/public/english/region/asro/.

Chapter 22

Contaminated injections in health care settings

Anja M. Hauri, Gregory L. Armstrong and Yvan J.F. Hutin

Summary

Injections given in health care settings with injection equipment reused in the absence of sterilization have been associated with infection with hepatitis B virus (HBV), hepatitis C virus (HCV) and HIV.

Input parameters included the annual number of injections per person, the proportion of injections administered with equipment reused in the absence of sterilization, the probability of transmission following percutaneous exposure, the age-specific prevalence of active infection, the prevalence of immunity (i.e. antibody to the hepatitis B core antigen or HbcAg [anti-HBc], anti-HCV and anti-HIV) and the incidence of HBV, HCV and HIV infections. We used mathematical models to transform diverse sources of data available into the prevalence of contaminated injections and the relative risk associated with these practices.

Four subregions[1] (AMR-A, EMR-B, EUR-A and WPR-A) where reuse of injection equipment in the absence of sterilization was negligible were assumed to have zero risk. In the remaining 10 subregions, the annual number of injections per person ranged from 1.9 to 11.3 and the proportion of injections administered with reused equipment ranged from 1.2% to 75%.

In 10 subregions, in 2000, injections caused an estimated 21 million HBV infections, two million HCV infections and 260 000 HIV infections, accounting for 32%, 40% and 5% of new infections, respectively. Thus, the burden in 2000 due to past and present exposure accounted for 501 000 deaths and 10 461 000 disability-adjusted life years (DALYs).

Injection overuse and unsafe practices are common worldwide and account for a high burden of infections with bloodborne pathogens. There is a need for policies and programmes for the safe and appropriate use of injections in countries where poor practices occur.

1. Introduction

During the twentieth century, injection use increased tremendously and today injections are probably the most common health care procedure (Drucker et al. 2001). Poor injection practices, including injection overuse and unsafe practices, have been reported in many developing and transitional countries (Simonsen et al. 1999). Many injections given for curative purposes in developing and transitional countries are unnecessary as they are prescribed for the treatment of conditions that could be treated with oral drugs or for which medications are not needed (Reeler 1990; Simonsen et al. 1999). In addition to being unnecessary, many injections are unsafe. Of particular concern is the reuse of injection equipment in the absence of sterilization. A common practice consists of rinsing injection equipment between injections in a pot of tepid water (Figure 22.1).

Unsafe injection practices constitute an important route of infection for bloodborne pathogens. Recently, a study suggested that the spread of HCV through unsafe injections in Egypt may represent the largest nosocomial outbreak ever reported (Frank et al. 2000). Epidemiological studies have reported an association between contaminated injections and infection with bloodborne pathogens, including HBV, HCV and HIV

Figure 22.1 Injection equipment soaked in tepid water before reuse in the absence of sterilization, Africa, 2000

Note the disposable syringes rinsed in the tepid water (arrow 1) and the multi-dose medication vials (arrow 2).

(Simonsen et al. 1999). The causal nature of this association is supported by many criteria. First, transmission through unsafe injection practices is biologically plausible because all three viruses are present in blood and body fluids of infected individuals (Choo et al. 1989; Molina et al. 1994; Shikata et al. 1977). They can be transmitted by transfusion (Aach et al. 1991; Busch et al. 1996; Senior et al. 1974) and other percutaneous routes, including needle-stick injuries among health care workers (Cardo et al. 1997; CDC 1997; Seeff et al. 1978). Second, several studies in developing countries have demonstrated an association between receiving injections and infection with bloodborne pathogens. Third, the measures of association (e.g. odds ratios) often exceed 2 (Luby et al. 1997; Narendranathan and Philip 1993; Quigley et al. 2000) and show a dose–response relationship (Khan et al. 2000; Ko et al. 1991a; Quigley et al. 2000). Fourth, studies have also reported an association between recent, incident cases of infection with HBV (Hutin et al. 1999), HCV (El-Sakka 1997) and HIV (Quigley et al. 2000), and exposure to injections during the time period that patients were likely to have been infected, indicating that the exposure preceded the outcome.

The proportion of new infections with HBV, HCV and HIV attributable to unsafe injection practices in specific populations can be estimated from case–control and cohort studies. A total of 12 studies (Table 22.1) were identified to examine the association between HBV infection and injections, with population attributable fractions ranging between 21% and 61% (Anonymous 1998; Hsu et al. 1993; Hussain 2001; Hutin et al. 1999; Ko and Chung 1991; Ko et al. 1991a; Luby et al. 1997; Narendranathan and Philip 1993; Simard et al. 2000; Singh et al. 2000; Thuring et al. 1993; Val Mayans et al. 1990). Of these eight (67%) were based upon recent, incident cases. A total of 10 studies (Table 22.2) were identified to examine the association between HCV infection and injections, with population attributable fractions ranging between 20% and 84% (Chang et al. 1996; Chen et al. 1995; El-Sakka 1997; Ho et al. 1997; Khan et al. 2000; Luby et al. 1997; Mohamed et al. 1996; Sun et al. 1999, 2001; Thuring et al. 1993). Of these, three were based upon recent, incident cases. A total of four studies (Table 22.3) based upon recent, incident cases were identified to examine the association between HIV infection and injections, with population attributable fractions ranging between 8% and 45% (Bultreys et al. 1994; N'Galy et al. 1988; Quigley et al. 2000; Wawer et al. 1994). (Studies based upon prevalent cases of HIV infection are not included in this report as the high frequency of HIV transmission through sexual exposure raises the possibility of reverse causation.)

Two limitations were common among the studies of the association between injections and infections. A first limitation was that studies of persons with prevalent, chronic infections are generally unable to distinguish the direction of the causal relationship between injections and infection. While study subjects could have acquired infections because

Table 22.1 Studies examining the association between health care injections and HBV infection

Country or area	*Author(s)*	*Year of study*	*Study design*	*Types of cases*	*Attributable fraction (%)*
Cambodia	Thuring et al. (1993)	1990–1991	Survey	Prevalent	2–13
China (Province of Taiwan)	Hsu et al. (1993)	1994	Case–control	Prevalent	24.6
China (Province of Taiwan)	Ko and Chung (1991)	1984–1989	Cohort	Incident	43.1
China (Province of Taiwan)	Ko et al. (1991a)	1991[a]	Cohort	Incident	73.9
Gambia	Val Mayans et al. (1990)	1988	Cohort	Incident	*
Egypt	Anonymous (1998)	1994	Case–control	Incident	27.7
India	Narendranathan and Philip (1993)	1993[a]	Case–control	Incident	53.3
India	Singh et al. (2000)	1998	Case–control	Incident	49.7
Pakistan	Hussain (2001)	2000–2001	Case–control	Prevalent	52
Pakistan	Luby et al. (1997)	1994	Case–control	Prevalent	35–41
Romania	Hutin et al. (1999)	1998	Case–control	Incident	40
Republic of Moldova	Hutin et al. (1999)	1994–1995	Case–control	Incident	21, 52[b]

* No association found. However, only immunization injections were considered.

[a] Year of publication.

[b] 21% among children, 52% among adults.

Table 22.2 Studies examining the association between health care injections and HCV infection

Country or area	*Author(s)*	*Year of study*	*Study design*	*Types of cases*	*Attributable fraction (%)*
Cambodia	Thuring et al. (1993)	1990–1991	Survey	Prevalent	90.6
China (Province of Taiwan)	Chang et al. (1996)	1991	Survey	Prevalent	50.4
China (Province of Taiwan)	Chen et al. (1995)	1990–1994	Case–control	Incident	20.1
China (Province of Taiwan)	Ho et al. (1997)	1993	Case–control	Prevalent	51–88
China (Province of Taiwan)	Sun et al. (1999)	1992	Case–control	Prevalent	44
China (Province of Taiwan)	Sun et al. (2001)	1994	Case–control	Incident	36.4
Egypt	El-Sakka (1997)	1996–1997	Case–control	Incident	87.9
Egypt	Mohamed et al. (1996)	1996	Survey	Prevalent	9.9
Pakistan	Khan et al. (2000)	1995	Case–control	Prevalent	24.4–78.5
Pakistan	Luby et al. (1997)	1994	Case–control	Prevalent	1.4, 62.9[a]

[a] 1.4% for injections received during past year; 62.9% for injections received during the past 10 years.

Table 22.3 Studies examining the association between health care injections and HIV infection[a]

Country	*Author(s)*	*Year of study*	*Study design*	*Types of cases*	*Attributable fraction (%)*
Democratic Republic of the Congo	N'Galy et al. (1988)	1984–1986	Cohort	Incident	28
Rwanda	Bultreys et al. (1994)	1989–1993	Cohort	Incident	45
Uganda	Quigley et al. (2000)	1990–1997	Case–control	Incident	16, 41[b]
Uganda	Wawer et al. (1994)	1989–1990	Cohort	Incident	8

[a] Restricted to studies recruiting recent, incident cases of HIV infection.

[b] 16% among women, 41% among men.

they received injections, they could also have received injections as a result of complications of their infection. Studies examining risk factors for HCV and HIV infections are more often affected by this bias because recent, acute cases of infection with these two pathogens are difficult to identify. However, three elements suggest that reverse causation is unlikely. First, most case patients in these studies were asymptomatic and therefore unlikely to seek injections for treatment of their infection. Second, a study that included prevalent cases of infection and examined the association between injections received during different time periods reported that injections received in a distant past were more strongly associated with infection than those received in a recent past, precisely the opposite of what could be expected if the hypothesis of reverse causation were true (Luby et al. 1997). Third, studies that included incident, recent cases of infections have reported similar associations (Chen et al. 1995; El-Sakka 1997; Sun et al. 2001). This includes a prospective cohort study examining the risk factors for HCV infection that validated the results of a cross-sectional survey conducted in the same population (Sun et al. 2001).

A second limitation was that the association between injections and infections with bloodborne pathogens may have been confounded by a number of other exposures, including sexually transmitted infections (STIs). In some cases, the apparent association between injections and infection may be secondary to two hidden associations—between STIs and injections on the one hand, and between STIs and infection on the other. However, confounding is unlikely to explain the associations observed because most studies also examined risk factors other than injections, including STIs, and a number of studies used stratification and multivariate analysis to control for these potential confounders. Nonetheless, there is still a need for research to determine the degree to

which STIs and injections confound each other's relationship with bloodborne pathogens, particularly in the case of HIV infection.

Because information from epidemiological studies was too limited to permit estimation of the global burden of disease attributable to unsafe injections, a global mass action mathematical model was generated in 1995 (Aylward et al. 1995) and further developed to formulate regional estimates in 1999 (Kane et al. 1999). This model included input parameters reflecting injection frequency, injection safety, the percutaneous transmission potential of bloodborne pathogens and the epidemiology of infection with HBV, HCV and HIV. Results of this analysis suggested that each year, in the world, reuse of injection equipment in the absence of sterilization accounts for 8 to 16 million HBV infections, 2.3 to 4.7 million HCV infections and 80 000 to 160 000 HIV infections (Kane et al. 1999).

This mass action model had three main limitations. First, it did not address variations of input parameters (i.e. injection frequency, prevalence of immunity and incidence of HIV infection) across age and sex groups within subregions. Second, no systematic procedure was used to review the literature and generate subregional estimates for injection frequency and injection safety. In this work, we used a new mathematical model to estimate the global burden of disease from unsafe injection practices, which although based on the same general approach as Kane et al. (1999) improves on some of the data limitations. In our analysis, we considered only HBV, HCV and HIV infections because of the substantial information on their association with unsafe injections and because these pathogens probably account for the majority of injection-associated infections. Other complications of unsafe injections not included in this model include abscesses (Fontaine et al. 1984; Soeters and Aus 1989), septicaemia (Archibald et al. 1998), malaria (Abulrahi et al. 1997) and infection with viral haemorrhagic fever viruses (Fisher-Hoch et al. 1995; WHO 1976).

2. Methods

2.1 Definitions

Health care injection

We defined a health care injection as a procedure that introduces a substance into the body through a piercing of the skin or of a mucosal membrane, including intradermal, subcutaneous, intramuscular and intravenous injections, for curative or preventive health care purposes, whether administered in a formal health care setting (e.g. clinic, hospital) or other settings (e.g. homes, pharmacies). Injections of illicit drugs were not considered in this work (see chapter 13).

Reuse of injection equipment in the absence of sterilization

We defined reuse of injection equipment as the administration of an injection to a recipient with a syringe or needle that had been previously used on another person and that was reused in the absence of sterilization. In this chapter, reuse of injection equipment in the absence of sterilization will simply be referred to as "reuse of injection equipment".

Choice of exposure variable contaminated injections

Reuse of injection equipment in itself would not be a risk factor in the absence of source patients infected with bloodborne pathogens. Thus, contaminated injections were the risk factor of interest. An injection contaminated with a bloodborne pathogen was defined as an injection given with a needle or a syringe used on an infected patient and reused on a second patient. The exposure under consideration for this study was defined as receiving at least one injection contaminated with HBV, HCV or HIV in one year. Exposure status would therefore depend on reuse of equipment, injection frequency and prevalence of active infection with HBV, HCV and HIV in the population. Persons receiving no contaminated injection in one year were considered unexposed. Four subregions (AMR-A, EMR-B, EUR-A and WPR-A) where reuse of injection equipment in the absence of sterilization was negligible were assumed to have zero risk.

Theoretical minimum level of exposure

The theoretical minimum level of exposure was zero contaminated injections per person and per year. This theoretical minimum is also an achievable goal as there are no reports of reuse of injection equipment in many industrialized countries.

2.2 Transmission model

Data on the risk associated with contaminated injections are generally not available as relative risks, especially since these can change from one place or time to another due to changes in background prevalence. Instead, information from diverse sources such as case–control studies, cross-sectional studies and observational studies of injection practices were brought together and integrated by means of mathematical models to develop internally consistent estimates of prevalence and hazard. The hazard estimates were based on the mass action principle, which states that

$$I_u = p_s\left[1-(1-p_t p_r p_v)^n\right]$$

where p_s is the proportion of the population susceptible to infection (in most cases, 1 minus prevalence of antibody to the virus), p_t is the probability of transmission after percutaneous exposure to a particular

pathogen, p_r is the probability that injection equipment will have been reused, p_v is the prevalence of active infection and n is the annual number of injections per person. This model implicitly assumes that the whole population is equally likely to be currently infected or receive an injection. For HBV, HCV and HIV, the three pathogens under consideration, this incidence is small enough that the equation can be simplified to

$$I_u = p_s \times p_t \times p_r \times p_v \times n$$

which can be further reduced to

$$I_u = p_s \times p_t \times n_c$$

in which n_c is the average annual number of contaminated injections and

$$n_c = p_r \times p_v \times n$$

All parameters were assumed to be different for each of the three pathogens except the annual number of injections per person (n), which was assumed to be constant within a particular age, sex and subregional stratum and the probability of reuse of injection equipment (p_r), which was assumed to be constant within a particular subregion. The probability of transmission (p_t) was based upon studies estimating the risk of infection with HBV, HCV and HIV following a needle-stick exposure from an infected patient. For HBV, p_t was assumed to vary according to the proportion of the infected population that was negative for hepatitis B e-antigen (HBeAg), ($p_t = 0.06$), or HbeAg positive ($p_t = 0.3$), (Seeff et al. 1978). For HCV, p_t was assumed to be 0.018 (CDC 1997). For HIV, the generally accepted value of p_t of 0.003 (Cardo et al. 1997) for needle-stick injuries was too low, since most injuries on which this estimate was based were superficial and did not involve hollow-bore needles. At the same time, the estimated risk from a deep needle-stick injury that can be estimated from the same study, 0.021, was too high (it is higher than the estimated p_t for HCV) because time can elapse during which HIV can be inactivated between the initial use and the reuse of a syringe on a second patient. As a compromise, the mean of the estimates for superficial and deep injuries, 0.012, was used in the model as p_t for HIV.

2.3 Estimates of the proportion of the population exposed to contaminated injections from the mass action model

If $n_c < 1$ and each person in the population could receive only one injection, then the probability of receiving a contaminated injection, p_c, would equal n_c. However, it is possible for someone to receive two, three or more contaminated injections in any given year. Because contaminated

injections are small probability events (Table 22.6), it can be assumed that the number of contaminated injections per individual follows a Poisson distribution in the population with a mean of n_c per individual, then the probability of receiving zero injections would be exp $(-n_c)$, and the probability of receiving at least one injection would be

$$p_c = 1 - \exp(-n_c)$$

Thus when n_c is very small, p_c is approximately equal to n_c and each exposed person will receive on average only one contaminated injection per year, as noted above. In most other situations, p_c will be slightly smaller than n_c and each exposed person will receive on average $n_c/(1 - \exp(-n_c))$ contaminated injections per year.

2.4 Estimates of the relative risk from the mass action model

For the purposes of the model, we considered the total incidence of infection in the population, I_t, to be composed of two components: the incidence due to contaminated injections, I_u, and the baseline incidence, I_b, which can also be thought of as the incidence in the population if contaminated injections could be eliminated. I_t can be estimated from incidence or prevalence surveys and I_b can be estimated if I_t and I_u are known:

$$I_b = I_t - I_u$$

and the proportion of infections attributable to unsafe injections is

$$AF = I_u / I_t$$

As this proportion of infections would have occurred only among the exposed proportion of the population (p_c), the risk among the exposed relative to the unexposed, by back-calculation from attributable fraction (AF) relationship, would be:

$$RR_c = 1 + AF/(p_c \times (1 - AF))$$

In most situations where the necessary variables are available or can be estimated from existing data, this equation can estimate the relative risk. However, in situations where a substantial proportion of infections are attributable to contaminated injections (i.e. situations where I_u approaches I_t), this equation produces unstable estimates of the relative risk, and other methods were used, as below.

2.5 Estimates of relative risks from analytical epidemiological studies

Cohort and case–control studies that examined the association between injections and infection defined exposure as receiving at least one injection, contaminated or not, and the absence of exposure as receiving no injections. If RR_i is the estimate of relative risk from such a study and n_i is the average number of injections received by the cases, then $[1 + (RR_i - 1)/n_i]$ is the relative risk attributable to one injection. Only a portion $(p_r \times p_v)$ of the injections received are contaminated and persons who do not receive contaminated injections are at no increased risk. Therefore, the relative risk of infection in a person who receives only one contaminated injection is $[1 + (RR_i - 1)/(n_i \times p_r \times p_v)]$. In practice, this often underestimates the relative risk because persons who receive injections are more likely to have been infected in the past and are therefore less likely to be susceptible to infection. Because of this phenomenon, the calculated RR_i will be an underestimate of the true RR_i if nonsusceptible controls are not excluded from the relative risk calculation. To account for this phenomenon, we assumed that the number of injections received in the prior year was approximately proportional to the probability of having been previously infected, such that the relative risk from receiving a single contaminated injection is $[1 + (RR_i - 1)/(n_i \times p_r \times p_v \times p_s)]$. This method was used to estimate hazard in cases where a substantial proportion of infections was attributable to contaminated injections, as described above.

2.6 Data sources

Injection practice parameters

Sources of information available to estimate the annual number of injections per person (n) included, by decreasing order of data quality, population-based injection frequency surveys and other population-based data providing injection frequency estimates. Sources of information for estimating the proportion of reuse (p_{re}) included, by decreasing order of data quality, observational studies of injection practices using the World Health Organization (WHO) standardized injection safety assessment survey tool (WHO 2002), studies of injection practices conducted using other, non-standardized methods, and back-calculations in published analytical epidemiological studies using the mass action equation and the relative risks of infection with bloodborne pathogens associated with receiving injections.

Sources of information were obtained through Medline searches, searches in WHO unpublished documents, including evaluations of the Expanded Programme on Immunization (EPI) and unpublished reports made available through the electronic mail list server of the Safe Injection Global Network (SIGN) (Bass 2000; Hutin and Chen 1999). All studies were reviewed using a standardized study abstraction instrument

and entered in a database. Estimates were generated for each subregion for proportion of reuse (p_{re}) or number of injections per person for each age, sex and subregional stratum (n) using a standardized decision-making algorithm to use the best source of data available.

The frequency distribution of the annual number of injections per person that was available from two studies conducted in EUR-B (CDC 1999) and in EUR-C (WHO 1999) indicated that a small proportion of the population above the 90th percentile received more than 20 injections per year. To avoid overestimating the attributable fraction, we made the conservative assumption that those receiving such a high number of injections had already been infected and were already immune. Thus, for these two subregions (EUR-B and EUR-C) where the injection frequency distribution was available, we excluded those who had received more than 20 injections per year (approximately above the 90th percentile), thereby reducing the annual number of injections per year. For the other subregions, data were available in tabulated form in published reports. This format already eliminated the upper 10% of the frequency distribution and no adjustment was necessary (e.g. persons reporting more than seven injections per year were all considered to have received eight injections per year). When more than one source of information regarding injection frequency or reuse of equipment was available for one age, sex and subregion stratum, all were used to compute an estimate.

Prevalence and incidence of HBV, HCV and HIV infection

We used the prevalence of active infection in the general population to estimate the proportion of patients representing a source of contamination for reused syringes and/or needles (p_v). Therefore, we did not assume the prevalence of active infection to be higher in a health care setting, nor considered different strata according to selected settings (e.g. immunization vs clinic for the management of STIs) (see discussion). Estimates for the proportion of the population chronically infected with HBV, HCV and HIV were obtained from the WHO programmes on HBV (C. Nelson, personal communication, 2000) and HCV (D. Lavanchy, personal communication, 2000), and from the Joint United Nations Programme on HIV/AIDS (UNAIDS 2000). In the case of HBV and HCV, catalytic models in which the annual risk of infection was constant over time and over age groups were generated. These models were fitted so that annual risk of infection led to region-specific estimates of the prevalence of active infection. Once the annual risk of infection was obtained, it was used to estimate the age-specific prevalence of susceptibility and the total incidence of infection among susceptible individuals. In the case of HIV, incidence estimates were obtained from UNAIDS (2000).

2.7 Estimates of the proportion of the population exposed to contaminated injections

Proportion of reuse

Sources of information used to generate the estimates (Table 22.4) included observational studies of injection practices using the WHO standardized injection safety assessment survey tool (AFR-D, AFR-E and EUR-B), observational studies of injection practices conducted using non-standardized methods (SEAR-B, SEAR-D and WPR-D), back-calculations using the mass action equation and the relative risks of infection with bloodborne pathogens associated with receiving injections (EUR-C), and a combination of the second and the third methods (EMR-D). No quantitative data were available for six subregions. For two of them, AMR-B and AMR-D, there were qualitative reports of reuse. For AMR-B, these reports suggested that reuse was uncommon (Flaskerud and Nyamathi 1996; Ugalde and Homedes 1988; Villanueva et al. 1997). Thus, estimates from the other subregion with the lowest frequency of reuse (EUR-B) were extrapolated. For AMR-D, as qualitative reports suggested that reuse was more common than in AMR-B (Janszen and Laning 1993), estimates from EUR-C, with the second lowest frequency of reuse, were extrapolated. For EUR-A, EMR-B, AMR-A and WPR-A, representing mostly countries with high per capita gross national product, the proportion of reuse was considered negligible. Among subregions for which quantified estimates were available, SEAR-D had the highest proportion of reuse (75%), followed by EMR-D (70%) and WPR-B (30%). EUR-B had the lowest proportion of reuse (1.2%). (See Figure 22.2.)

Annual number of injections per person

Sources of information used to generate subregional input parameters (Table 22.5) included population-based injection frequency surveys, and other population-based studies that provided information about injection frequency. No injection frequency estimates were generated for those subregions for which reuse was considered negligible as no risk applied. Among subregions with quantified information available, EUR-C was the subregion with the highest injection frequency (11.3 injections per person and per year), followed by EUR-B (5.2 injections per person and per year) (CDC 1999; WHO 1999). However, when the top 10th percentile of injection frequency was removed, EMR-D was the subregion with the highest injection frequency (4.3 injections per person and per year), followed by SEAR-D (4 injections per person and per year). The subregions with the lowest annual number of injections per person were AMR-B (1.7 injections per person and per year) and AMR-D (1.9 injections per person and per year).

Table 22.4 Subregional estimates of the proportion of injections administered with reused equipment and data sources used, 2000

	AFR-D	*AFR-E*	*AMR-B*	*AMR-D*	*EMR-D*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-B*
Proportion of reuse (p_r) (%)	19	17	1.2	11	70	1.2	11	30	75	30
Methods used (see text)	Standard WHO survey	Standard WHO survey	Extrapolation	Extrapolation	Combination of methods	Standard WHO survey	Back-calculation	Non-standard surveys	Non-standard surveys	Non-standard surveys
Countries from which WHO standardized injection safety surveys were used	Burkina Faso, Chad, Gambia, Mauritania and Niger[a]	Eritrea, Ethiopia, Swaziland, Zambia and Zimbabwe[a]	NA	NA	NA	Kyrgyzstan[d]	NA	NA	NA	NA
Countries from which non-standardized injection safety surveys were used	NA	NA	NA	NA	Pakistan (Khan et al. 2000)	NA	NA	Indonesia (Kosen 1999)	India (Lakshman and Nichter 2000)	China (Schnurr et al. 1999)
Countries from which back-calculated injection safety estimates were used	NA	NA	NA	NA	Egypt (El-Sakka 1997)	NA	Republic of Moldova (Hutin et al. 1999)	NA	NA	NA
Use of other subregional data	NA	NA	EUR-B[b]	EUR-C[c]	NA	NA	NA	NA	NA	NA

NA Not applicable.

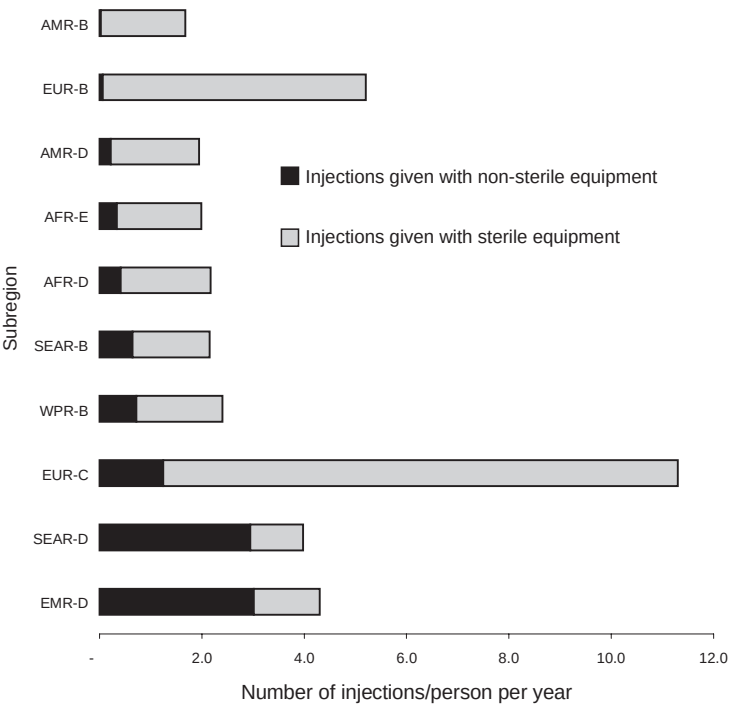
[a] Unpublished WHO reports.

[b] Qualitative information available on injection safety for AMR-B (Flaskerud and Nyamathi 1996; Ugalde and Homedes 1988; Villanueva et al. 1997) suggested occurrence of reuse in the absence of sterilization. To generate a conservative estimate, estimates for the subregion with the lowest proportion were extrapolated.

[c] Qualitative information available on injection safety for AMR-D (Janszen and Laning 1993) suggested occurrence of reuse in the absence of sterilization with a higher frequency than AMR-B. Thus, estimates for the subregion with the second lowest proportion were extrapolated.

[d] J. Fitzner, personal communication, 2002.

Figure 22.2 Number of injections per person and per year, and proportion of these administered with injection equipment reused in the absence of sterilization, by subregion, 2000[a]



[a] Crude injection frequency estimates. For the purpose of the model, estimates for EUR-B and EUR-C were used after subtraction of the 10 top percentiles.

Proportion of the population exposed to contaminated injections

The proportion of the population exposed to contaminated injections reflected the frequency of injections received, the frequency of reuse of injection equipment and the prevalence of active infection with HBV, HCV and HIV (Table 22.6). This estimate varied from 0.03% (AMR-B) to 13.33% (EMR-D) in the case of HBV, from less than 0.03% (AMR-B) to 16.73% (EMR-D) in the case of HCV, and from 0.00% (EUR-B) to 2.05% (AFR-E) in the case of HIV.

2.8 Estimates for the relative risks of infection for receiving contaminated injections

In the case of HIV, contaminated injections did not account for most new infections (i.e. I_u was not close to I_t). Thus, model-based estimates of relative risk were used for all subregions (Table 22.7[c]). In the case

Table 22.5 Subregional injection frequency estimates and data sources used, 2000

		AFR-D	*AFR-E*	*AMR-B*	*AMR-D*	*EMR-D*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-B*
Annual number of injections per person (*n*)[a]	Crude	2.2	2.0	1.7	1.9	4.3	5.2	11.3	2.1	4.0	2.4
	Truncated[b]	NA	NA	NA	NA	NA	2.5	3.5	NA	NA	NA
Countries from which injection frequency surveys were used		Guinea-Bissau (Ferry 1995)	Central African Republic, Côte d'Ivoire, United Republic of Tanzania, Zambia, Burundi (Ferry 1995), Uganda (Priotto et al. 2001)	Brazil (Ferry 1995)	NA	Egypt (Talaat et al. 2001)	Romania (CDC 1999)	Republic of Moldova (WHO 1999)	Thailand (Ferry 1995; Reeler and Hematorn 1994) Indonesia (Ferry 1995; van Staa and Hardon 1996)	India (Anand et al. 2001; Ferry 1995)	NA

Countries from which other population-based data were used	Cameroon (Ferry 1995; Guyer et al. 1980) Nigeria (Otu et al. 2002)	United Republic of Tanzania (Grosskurth 1995; Quigley et al. 1997) Uganda (Quigley et al. 2000)	Latino communities in the USA (Flaskerud and Nyamathi 1996)	Haiti (Pape et al. 1985)	Pakistan (Luby et al. 1997)	NA	NA	NA	India (Deivanayagam et al. 1993)	China (Province of Taiwan) (Chang et al. 1996; Ko and Chung 1991; Ko et al. 1991a, 1991b)
Use of different estimates for males and females	No	Yes	No	No	No	Yes	Yes	No	No	No
Addition of 0.5 injections per year among 1–4 years of age to account for immunization	Yes	Yes	Yes	Yes	Yes	No[c]	No[c]	No[c]	Yes	No[c]

NA Not applicable.

[a] Estimates age-adjusted using age group-specific population sizes to simplify data presentation. The model actually used age group and sex-specific estimates.

[b] Injections received by those receiving more than 20 injections per year excluded for the calculation of the burden of disease.

[c] Not applicable: age-specific injection frequency estimate takes into account immunization injections.

Table 22.6(a) Proportion of the population (%) receiving injections contaminated with HBV annually, 2000[a]

		Age group (years)							
Subregion	*Sex*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	5.51	4.47	4.47	4.47	4.47	4.47	4.47	4.47
	Female	5.51	4.47	4.47	4.47	4.47	4.47	4.47	4.47
AFR-E	Male	4.72	3.31	3.31	3.31	3.31	3.31	3.31	3.31
	Female	4.72	4.20	4.20	4.20	4.20	4.20	4.20	4.20
AMR-B	Male	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	Female	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
AMR-D	Male	0.52	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	Female	0.52	0.41	0.41	0.41	0.41	0.41	0.41	0.41
EMR-D	Male	13.33	12.01	12.01	12.01	12.01	12.01	12.01	12.01
	Female	13.33	12.01	12.01	12.01	12.01	12.01	12.01	12.01
EUR-B	Male	0.28	0.12	0.09	0.15	0.15	0.15	0.15	0.15
	Female	0.28	0.13	0.16	0.19	0.19	0.19	0.19	0.19
EUR-C	Male	1.86	1.09	1.29	1.42	1.42	1.42	1.42	1.42
	Female	1.78	0.92	1.26	1.78	1.78	1.78	1.78	1.78
SEAR-B	Male	9.95	5.15	5.15	5.15	5.15	5.15	5.15	5.15
	Female	9.95	5.15	5.15	5.15	5.15	5.15	5.15	5.15
SEAR-D	Male	11.22	10.02	10.02	10.02	10.02	10.02	10.02	10.02
	Female	11.22	10.02	10.02	10.02	10.02	10.02	10.02	10.02
WPR-D	Male	11.74	7.84	7.84	7.84	7.84	7.84	7.84	7.84
	Female	11.74	7.84	7.84	7.84	7.84	7.84	7.84	7.84

[a] AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

Table 22.6(b) Proportion of the population (%) receiving injections contaminated with HCV annually, 2000[a]

		Age group (years)							
Subregion	*Sex*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	1.29	1.04	1.04	1.04	1.04	1.04	1.04	1.04
	Female	1.29	1.04	1.04	1.04	1.04	1.04	1.04	1.04
AFR-E	Male	1.12	0.78	0.78	0.78	0.78	0.78	0.78	0.78
	Female	1.12	0.99	0.99	0.99	0.99	0.99	0.99	0.99
AMR-B	Male	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	Female	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
AMR-D	Male	0.62	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	Female	0.62	0.49	0.49	0.49	0.49	0.49	0.49	0.49
EMR-D	Male	16.73	15.10	15.10	15.10	15.10	15.10	15.10	15.10
	Female	16.73	15.10	15.10	15.10	15.10	15.10	15.10	15.10
EUR-B	Male	0.09	0.04	0.03	0.05	0.05	0.05	0.05	0.05
	Female	0.10	0.05	0.06	0.07	0.07	0.07	0.07	0.07
EUR-C	Male	1.19	0.70	0.82	0.91	0.91	0.91	0.91	0.91
	Female	1.14	0.59	0.81	1.14	1.14	1.14	1.14	1.14
SEAR-B	Male	3.31	1.68	1.68	1.68	1.68	1.68	1.68	1.68
	Female	3.31	1.68	1.68	1.68	1.68	1.68	1.68	1.68
SEAR-D	Male	5.92	5.27	5.27	5.27	5.27	5.27	5.27	5.27
	Female	5.92	5.27	5.27	5.27	5.27	5.27	5.27	5.27
WPR-D	Male	3.28	2.16	2.16	2.16	2.16	2.16	2.16	2.16
	Female	3.28	2.16	2.16	2.16	2.16	2.16	2.16	2.16

[a] AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

Table 22.6(c) Proportion of the population (%) receiving injections contaminated with HIV annually, 2000[a]

		Age group (years)							
Subregion	*Sex*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	0.64	0.52	0.52	0.52	0.52	0.52	0.52	0.52
	Female	0.64	0.52	0.52	0.52	0.52	0.52	0.52	0.52
AFR-E	Male	2.05	1.43	1.43	1.43	1.43	1.43	1.43	1.43
	Female	2.05	1.82	1.82	1.82	1.82	1.82	1.82	1.82
AMR-B	Male	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	Female	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
AMR-D	Male	0.13	0.11	0.11	0.11	0.11	0.11	0.11	0.11
	Female	0.13	0.11	0.11	0.11	0.11	0.11	0.11	0.11
EMR-D	Male	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.09
	Female	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.09
EUR-B	Male	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Female	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EUR-C	Male	0.08	0.05	0.05	0.06	0.06	0.06	0.06	0.06
	Female	0.07	0.04	0.05	0.07	0.07	0.07	0.07	0.07
SEAR-B	Male	0.33	0.16	0.16	0.16	0.16	0.16	0.16	0.16
	Female	0.33	0.16	0.16	0.16	0.16	0.16	0.16	0.16
SEAR-D	Male	1.19	1.05	1.05	1.05	1.05	1.05	1.05	1.05
	Female	1.19	1.05	1.05	1.05	1.05	1.05	1.05	1.05
WPR-D	Male	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	Female	0.06	0.04	0.04	0.04	0.04	0.04	0.04	0.04

[a] AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

Table 22.7(a) Relative risks associated with receiving injections contaminated with HBV, 2000[a]

		Age group (years)							
Subregion	*Sex*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	3.53	3.45	3.45	3.45	3.45	3.45	3.45	3.45
	Female	3.53	3.45	3.45	3.45	3.45	3.45	3.45	3.45
AFR-E	Male	3.47	3.37	3.37	3.37	3.37	3.37	3.37	3.37
	Female	3.47	3.43	3.43	3.43	3.43	3.43	3.43	3.43
AMR-B	Male	75.21	74.68	74.68	74.68	74.68	74.68	74.68	74.68
	Female	75.21	74.68	74.68	74.68	74.68	74.68	74.68	74.68
AMR-D	Male	25.43	24.78	24.78	24.78	24.78	24.78	24.78	24.78
	Female	25.43	24.78	24.78	24.78	24.78	24.78	24.78	24.78
EMR-D	Male	11.71	11.71	11.71	11.71	11.71	11.71	11.71	11.71
	Female	11.71	11.71	11.71	11.71	11.71	11.71	11.71	11.71
EUR-B	Male	6.49	6.44	6.43	6.45	6.45	6.45	6.45	6.45
	Female	6.49	6.45	6.46	6.47	6.47	6.47	6.47	6.47
EUR-C	Male	7.06	6.77	6.85	6.9	6.9	6.9	6.9	6.9
	Female	7.04	6.71	6.84	7.03	7.03	7.03	7.03	7.03
SEAR-B	Male	7.02	5.56	5.56	5.56	5.56	5.56	5.56	5.56
	Female	7.02	5.56	5.56	5.56	5.56	5.56	5.56	5.56
SEAR-D	Male	11.71	11.71	11.71	11.71	11.71	11.71	11.71	11.71
	Female	11.71	11.71	11.71	11.71	11.71	11.71	11.71	11.71
WPR-D	Male	7.86	6.28	6.28	6.28	6.28	6.28	6.28	6.28
	Female	7.86	6.28	6.28	6.28	6.28	6.28	6.28	6.28

[a] AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

Table 22.7(b) Relative risks associated with receiving injections contaminated with HCV, 2000[a]

		Age group (years)							
Subregion	*Sex*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	19.74	18.88	18.88	18.88	18.88	18.88	18.88	18.88
	Female	19.74	18.88	18.88	18.88	18.88	18.88	18.88	18.88
AFR-E	Male	17.5	16.6	16.6	16.6	16.6	16.6	16.6	16.6
	Female	17.5	17.15	17.15	17.15	17.15	17.15	17.15	17.15
AMR-B	Male	31.36	31.28	31.28	31.28	31.28	31.28	31.28	31.28
	Female	31.36	31.28	31.28	31.28	31.28	31.28	31.28	31.28
AMR-D	Male	21.35	20.81	20.81	20.81	20.81	20.81	20.81	20.81
	Female	21.35	20.81	20.81	20.81	20.81	20.81	20.81	20.81
EMR-D	Male	27.77	27.77	27.77	27.77	27.77	27.77	27.77	27.77
	Female	27.77	27.77	27.77	27.77	27.77	27.77	27.77	27.77
EUR-B	Male	31.9	31.4	31.3	31.49	31.49	31.49	31.49	31.49
	Female	31.9	4.01	31.52	31.62	31.62	31.62	31.62	31.62
EUR-C	Male	31.9	31.4	31.3	31.49	31.49	31.49	31.49	31.49
	Female	31.9	4.01	31.52	31.62	31.62	31.62	31.62	31.62
SEAR-B	Male	38.0	23.86	23.86	23.86	23.86	23.86	23.86	23.86
	Female	38.0	23.86	23.86	23.86	23.86	23.86	23.86	23.86
SEAR-D	Male	37.64	26.72	26.72	26.72	26.72	26.72	26.72	26.72
	Female	37.64	26.72	26.72	26.72	26.72	26.72	26.72	26.72
WPR-D	Male	37.64	26.72	26.72	26.72	26.72	26.72	26.72	26.72
	Female	37.64	26.72	26.72	26.72	26.72	26.72	26.72	26.72

[a] AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

Table 22.7(c) Relative risks associated with receiving injections contaminated with HIV, 2000[a]

		Age group (years)							
Subregion	*Sex*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	14.05	13.84	5.57	5.57	5.57	5.57	5.57	5.57
	Female	14.05	13.84	3.39	3.39	3.39	3.39	3.39	3.39
AFR-E	Male	5.95	5.79	2.3	2.3	2.3	2.3	2.3	2.3
	Female	5.95	5.89	1.72	1.72	1.72	1.72	1.72	1.72
AMR-B	Male	122.12	121.85	21.03	21.03	21.03	21.03	21.03	21.03
	Female	122.12	121.85	41.1	41.1	41.1	41.1	41.1	41.1
AMR-D	Male	66.27	65.09	8.12	8.12	8.12	8.12	8.12	8.12
	Female	66.27	65.09	16.25	16.25	16.25	16.25	16.25	16.25
EMR-D	Male	1	1	64.41	64.41	64.41	64.41	64.41	64.41
	Female	1	1	135.38	135.38	135.38	135.38	135.38	135.38
EUR-B	Male	1	1	1	1	1	1	1	1
	Female	1	1	1	1	1	1	1	1
EUR-C	Male	1	1	6.48	6.48	6.48	6.48	6.48	6.48
	Female	1	1	31.5	31.7	31.7	31.7	31.7	31.7
SEAR-B	Male	198.4	150.67	26.01	26.01	26.01	26.01	26.01	26.01
	Female	198.4	150.67	43.86	43.86	43.86	43.86	43.86	43.86
SEAR-D	Male	213.52	166.26	16.61	16.61	16.61	16.61	16.61	16.61
	Female	213.52	166.26	33.35	33.35	33.35	33.35	33.35	33.35
WPR-D	Male	1	1	41.69	41.69	41.69	41.69	41.69	41.69
	Female	1	1	127.3	127.3	127.3	127.3	127.3	127.3

[a] AMR-A, EMR-B, EUR-A and WPR-A excluded as reuse of injection equipment is negligible in these subregions.

of HCV, contaminated injections did not account for most of new infections in all subregions apart from EUR-C, SEAR-D and EMR-D. For EUR-C and SEAR-D, model-based relative risks for EUR-B and WPR-B, which had similar prevalence patterns, were used. In EMR-D, the available study-based relative risk was used (Table 22.7[b]) (El-Sakka 1997). In the case of HBV, contaminated injections did not account for most of new infections in all subregions apart from SEAR-D and EMR-D. In EMR-D, the available study-based relative risk was used (Anonymous 1998). For SEAR-D, the study-based relative risk for EMR-D was used as this subregion had HBV prevalence patterns and injection practices close to those of EMR-D (Table 22.7[a]).

2.9 Progression of HBV, HCV and HIV infection to disability and death

Time interval between infection and the occurrence of disability and death

The majority of the burden of disease associated with infections with HBV, HCV and HIV is delayed in time. For HBV infection, a small proportion of acute infections lead to death through fulminant liver failure. However, most of the burden of disease is secondary to the long-term consequences of chronic HBV infection, including end-stage liver disease and hepatocellular carcinoma. For HCV infection, the death to case ratio for acute infections is negligible and all the burden of disease is secondary to the long-term consequences of chronic infection, including end-stage liver disease and hepatocellular carcinoma. For HIV infection, the burden of disease is secondary to the progression to AIDS and to death following AIDS.

To take into account the time interval between infection and the progression towards death and disability, two measures need to be distinguished. First, the attributable burden in 2000 includes the current burden in the year 2000 that is secondary to past and present unsafe injection practices. Second, the future burden due to current unsafe injection practices in 2000 includes the future long-term consequences in terms of end-stage liver disease, hepatocellular carcinoma and AIDS of the HBV, HCV and HIV infections acquired in 2000 because of contaminated injections.

Current burden due to past and present unsafe injection practices

In the absence of information regarding injection practices in the past, we were unable to model the fraction of HBV, HCV and HIV infection that was attributable to contaminated injections before the year 2000. Thus, we made the assumption that the fraction of HBV, HCV and HIV infections attributable to contaminated injections in the past was identical to the one modelled for the year 2000. We then applied these attributable fractions to the current burden in 2000 in terms of DALYs and

deaths that were associated with the consequences of HBV, HCV and infection (i.e. acute infections, hepatocellular carcinoma, end-stage liver diseases and HIV infection/AIDS).

Future burden due to current unsafe injection practices

To reflect the delay between infection and disease outcomes, the fraction of new infections with HBV, HCV and HIV attributable to contaminated injections was converted into years of life lost (YLL) using synthetic cohorts of infected individuals followed for mortality associated with HBV, HCV or HIV infection (AIDS or chronic liver disease) and background mortality. Background mortality was taken into account using age, sex and subregion-specific Global Burden of Disease (GBD) life tables.[2] To estimate the years of life lost secondary to HBV infections, the model parameters included:

- a rate of progression to chronic infection of 30% among persons infected under the age of 5 years and of 6% for persons infected at the age of 5 years or older (McMahon et al. 1985);
- an annual rate of clearance of infection (i.e. sero-reversion) of 1% following chronic infection (Alward et al. 1985); and
- an age-dependent yearly mortality rate associated with chronic liver disease among persons chronically infected with HBV (Figure 22.3) that was modelled on the basis of African and Asian studies (Gay et al. 2001).

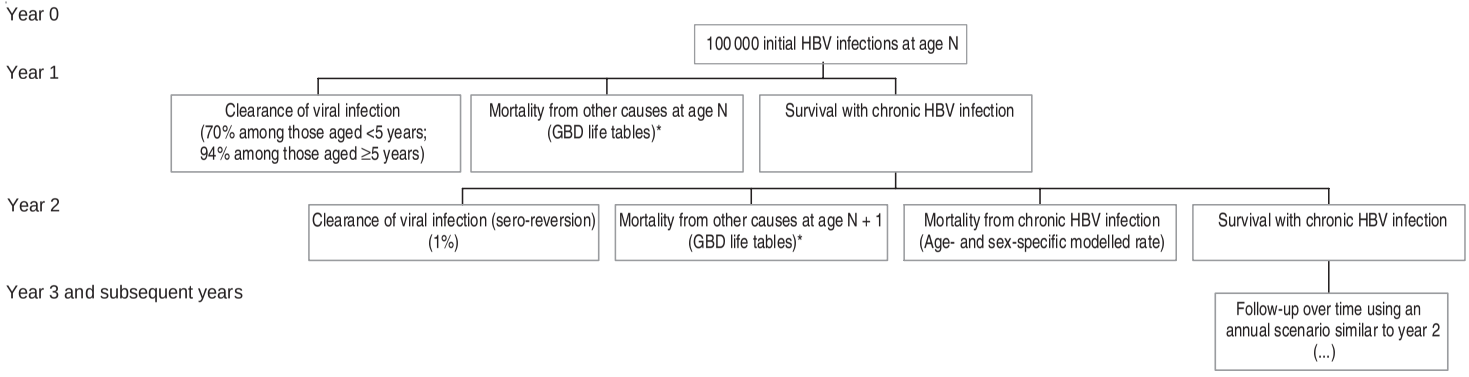
To estimate the years of life lost from HCV infections, two sets of assumptions were used according to the age of the individual at infection (Figure 22.4). For persons infected before the age of 40, the model parameters included:

- a rate of progression to chronic infection of 63%, the average of rates observed in two large studies conducted in this age group (Alter and Seeff 2000);
- a cumulated incidence rate of cirrhosis of 5% at 20 years among patients with chronic infection (Alter and Seeff 2000; Freeman et al. 2001); and
- a yearly mortality rate associated with hepatocellular carcinoma and chronic liver disease of 3.7% after the onset of cirrhosis, the average of two large studies (Alter and Seeff 2000).

For persons infected at the age of 40 years or older, the model parameters included:

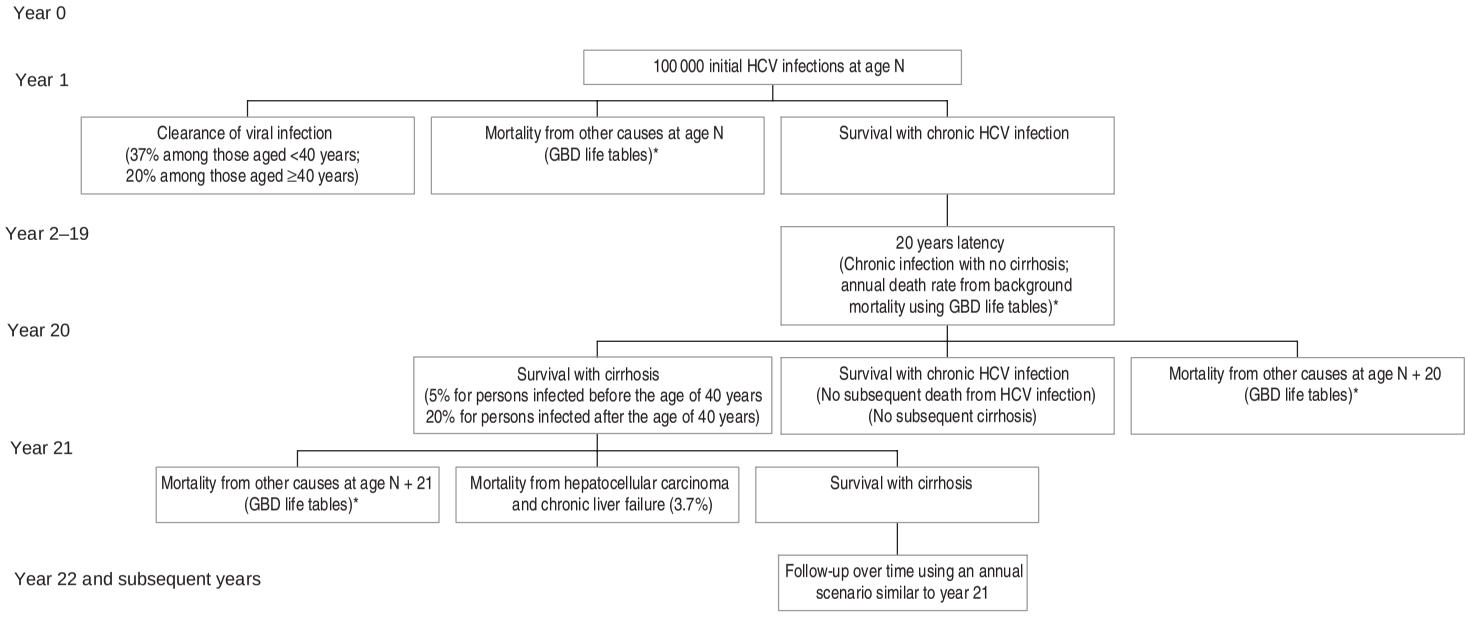
- a rate of progression to chronic infection of 80% (Alter and Seeff 2000);

Figure 22.3 Decision tree for the theoretical cohort used for the calculation of the years of life lost (YLL) secondary to hepatitis B virus infection



* GBD life tables prepared for *World health report 2001* (WHO 2001).

Figure 22.4 Decision tree for the theoretical cohort used for the calculation of the years of life lost (YLL) secondary to hepatitis C virus infection



* GBD life tables prepared for *World health report 2001* (WHO 2001).

- a cumulated incidence rate of cirrhosis of 20% at 20 years (Alter and Seeff 2000; Freeman et al. 2001); and
- a yearly mortality rate associated with chronic liver disease of 3.7% after the onset of cirrhosis, as in the younger age group.

Disability and death attributable to acute viral hepatitis were considered negligible for HBV and HCV in comparison with the disability and death secondary to chronic infection. In the case of HIV, parameters of progression of HIV infection to AIDS and death developed by WHO and UNAIDS were used (N. Walker, personal communication, 2002).

Uncertainty analysis

Standard errors were calculated for selected key input parameters, including the annual number of injections per person and the proportion of reuse. Standard formulae for the calculation of confidence intervals for means and proportions were used. In the specific case of the proportion of reuse estimated on the basis of measures of association, total sample size was assumed to be the total number of study participants included in the study.

For subregions for which good quality data were available on injection frequency (injection frequency surveys) or injection safety (standardized or non-standardized injection safety surveys), a 95% confidence interval was calculated for the input parameter on the basis of the standard error (±2 SE). For subregions for which only lower quality data were available for injection frequency (other population-based injection frequency data) or injection safety (back-calculated estimates), an arbitrary larger interval was used to account for added uncertainty (±4 SE). For subregions for which no data were available and inferences were made using other subregions, an even larger interval was arbitrarily used to account for added uncertainty (±6 SE).

Lower and upper bounds of the 95% confidence intervals for the proportion of the population exposed and for relative risk estimates were obtained by including the values for lower and upper bounds of the input parameters in the model equations. Confidence intervals for the relative risks that were study-based rather than model-based were calculated on the basis of the confidence interval of the relative risk in the original epidemiological studies. See section 4 for further discussion of sources of uncertainty.

3. Results

3.1 Fraction of infections attributable to contaminated injections in 2000

Globally, the fractions of incident HBV, HCV and HIV infections attributable to contaminated injections in the subregions where reuse of injec-

tions was reported were 31.9%, 39.9% and 5.4%, respectively (Table 22.8). For HBV, this proportion was highest in EMR-D (58.3%) and lowest in EUR-B (0.9%). For HCV, this proportion was highest in EMR-D (81.7%) and lowest in EUR-B (0.9%). For HIV, this proportion was highest in SEAR-D (24.3%) and lowest in AMR-B (0.00%). In absolute numbers of infections, our analysis indicated that globally, in 2000, contaminated injections may have caused 20.6 million cases of new HBV infections, 2.0 million cases of HCV infections and 260 000 cases of HIV infections.

3.2 Current burden due to past and present unsafe injection practices

The current burden in 2000 due to past and present unsafe injection practices reached 501 000 deaths, with the majority of deaths occurring in Asia (39% and 31% in WPR-B and SEAR-D, respectively) (Table 22.9) and among persons aged ≥15 years (n = 444 000, 88%). When death and disability were combined, the burden reached 10 461 000 DALYs, with a similar predominance in Asia (27% and 39% in WPR-B and SEAR-D, respectively) and adults (n = 8 419 000, 81% of DALYs among persons aged ≥15 years). Taken together, viral hepatitis B and C and their chronic consequences accounted for 74% and 61% of the deaths and DALYs, respectively, and HIV accounted for the remainder. There were no substantial differences in the distribution of death and disability by sex.

3.3 Future burden due to current unsafe injection practices

Models of natural history and background mortality allowed estimation of the burden of disease attributable to contaminated injections received in 2000. This analysis suggested that the 20.6 million HBV infections in the year 2000 would lead to 26 492 deaths in 2000 from fulminant hepatitis and an additional 49 000 early deaths from the consequences of chronic infection between 2000 and 2030. With respect to HCV infection, we estimated that the two million infections in 2000 would lead to 24 000 early deaths between 2000 and 2030. Finally, 210 000 of the 260 000 persons infected with HIV through contaminated injections in 2000 are expected to die prematurely from AIDS between 2000 and 2030. While our analytic horizon did not go beyond year 2030, it is anticipated that persons infected with HBV and HCV because of contaminated injections in 2000 would continue to develop long-term complications leading to death beyond this date.

4. Discussion

While the consequences of poor injection practices have been recognized for many decades (Anonymous 1945; Wyatt 1984), the safe and

Table 22.8 HBV, HCV and HIV infections attributable to contaminated injections (attributable fraction and absolute numbers, lower and upper estimates), 2000

	HBV		*HCV*		*HIV*	
Subregion	*Attributable fraction (%)*	*Number of infections*	*Attributable fraction (%)*	*Number of infections*	*Attributable fraction (%)*	*Number of infections*
AFR-D	10.9 (8.2–13.9)	639498 (478834–814351)	16.4 (12.3–20.8).	54681 (41078–69402)	2.5 (1.9–3.1)	18317 (13765–23243)
AFR-E	9.2 (6.9–11.5)	630976 (474379–792536)	13.0 (9.8–16.2)	54131 (40819–67794)	2.5 (1.9–3.1)	64412 (48520–80759)
AMR-B	2.3 (0.0–16.3)	14118 (112–98872)	0.9 (0.0–6.4)	2282 (18–15985)	0.2 (0.0–1.5)	305 (2–2132)
AMR-D	9.3 (0.0–26.9)	28570 (16–82490)	9.2 (0.0–26.7)	6304 (4–18215)	1.5 (0.0–4.5)	911 (1–2626)
EMR-D	58.3 (26.2–82.4)	2533443 (1140352–3580611)	81.7 (52.1–95.0)	645486 (412078–750452)	7.1 (5.7–8.5)	2210 (1775–2668)
EUR-B	0.9 (0.0–3.3)	21122 (156–78639)	0.9 (0.0–3.4)	2110 (16–7729)	0.0 (0.0–0.0)	0 (0–0)
EUR-C	7.7 (1.8–15.0)	193636 (46035–378229)	21.2 (6.1–34.7)	35668 (10287–58378)	0.6 (0.2–1.2)	1526 (374–2903)
SEAR-B	22.4 (16.5–28.7)	942038 (694606–1205102)	30.8 (22.8–39.2)	94873 (70235–120979)	7.0 (5.2–8.9)	6260 (4638–7980)
SEAR-D	53.6 (21.6–79.9)	8019210 (3237944–11954579)	59.5 (40.4–93.6)	498166 (338548–784474)	24.3 (18.3–0.1)	156663 (118235–194187)
WPR-B	33.6 (0.0–79.0)	7610161 (2126–17868925)	37.6 (0.0–89.8)	608200 (172–1454478)	2.5 (0.0–5.9)	5549 (2–13378)
World	31.9 (9.4–56.9)	20632772 (6074558–36854335)	39.9 (18.2–66.7)	2001901 (913254–3347885)	5.4 (3.9–7.0)	256152 (187312–329877)

Table 22.9 Current burden in 2000 due to past and present contaminated injections

		AFR-D	*AFR-E*	*AMR-B*	*AMR-D*	*EMR-D*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-B*	*World*
Deaths	HBV and HCV	6	6	1	1	41	1	10	25	88	193	372
(000s)	HIV/AIDS	10	44	0	0	4	0	0	3	66	1	129
	Total	16	51	1	2	44	1	10	28	154	194	501
DALYs	HBV and HCV	106	111	15	22	721	13	162	413	2007	2781	6349
(000s)	HIV/AIDS	325	1436	2	10	106	0	3	99	2094	38	4112
	Total	430	1547	17	31	826	13	165	512	4100	2819	10461

appropriate use of injections remains a target that has not been reached in most developing and transitional countries. Since the early 1990s, epidemiological studies of new HBV, HCV and HIV infections have indicated that unsafe injections are a risk factor for each disease (Simonsen et al. 1999). In most transitional and developing countries where HBV and HCV lead to a high burden of chronic liver disease, unsafe injections account for a high proportion of these infections (Hutin et al. 1999; Khan et al. 2000; Luby et al. 1997; Narendranathan and Philip 1993).

This chapter made use of the best available evidence regarding the rates of injection use, the frequency of reuse and the association between injections and infections. Use of a mass action model was needed for the communicable nature of outcomes and lack of transferability of hazard from across populations. The results indicate that in 2000, four decades after the widespread availability of disposable injection equipment and two decades into the HIV pandemic, contaminated injections accounted for close to a third of new HBV infections, 40% of new HCV infections and 5% of new HIV infections globally. The burden of disease in 2000 due to past and current infections reached 501 000 deaths and 10 461 000 DALYs, with the majority of deaths occurring among adults, mostly in Asia.

Using available studies, we described injection practices worldwide in terms of safety and frequency. Our analysis indicated high rates of injection worldwide, with marked subregional variations, for a total exceeding 16 thousand million injections in the 10 (of 14) subregions that were included in our study. This order of magnitude is validated by the market analysis suggesting that in Japan, the United States of America and western Europe, 17.5 thousand million syringes are sold annually (Kaninda 2001).

Four subregions stood out with particularly high estimates. The crude annual number of injections per person was the highest in the former socialist economies of Europe and central Asia, reaching 11.3 and 5.2 in EUR-B and EUR-C, respectively. Most injections in these countries are administered in public health care facilities by physicians or nurses, with a high number of injections per prescription (CDC 1999; WHO 1999). While patients' demand is commonly quoted by health care providers as a major driver of injection overuse, knowledge, attitude and practice (KAP) surveys find that health care providers have a tendency to overestimate patients' preference for injections and that the population is open to alternatives to injected medications (CDC 1999; Vong et al. 2002). In fact, KAP surveys conducted among physicians indicate that prescribers have false preconceptions about the effectiveness of injectable medications and that these preconceptions are sometimes supported by non-evidence-based official treatment protocols. Thus, prescribers' attitudes also contribute to injection overuse (Stoica et al. 1999).

Injection use was also high in the Middle East and in south Asia where the annual number of injections per person reached 4.3 and 4.0 in EMR-

D and SEAR-D, respectively. In these countries, a high proportion of injections are administered by private providers who, in some cases, are not medically qualified (Khan et al. 2000; Kosen 1999; Talaat et al. 2001). In such settings, health care providers' attitudes also drive injection overuse (Khan et al. 2000; Luby et al. 1997). However, practices are different. The reference to any guideline is uncommon. Injections are frequently used on an ad hoc basis to administer mixtures of antibiotics, analgesics, vitamins or antihistamines in the desire to meet what is believed to be the demand of the user (Khan et al. 2000).

Reducing injection overuse would only be a matter of promoting rational drug use if injections were administered safely. Unfortunately, our analysis indicated that injections are given in a way that may harm the injection recipient. Determinants of these unsafe injection practices include the lack of supplies of new, sterile, single-use, disposable injection equipment (Dicko et al. 2000), the lack of awareness among patients and providers regarding the risks associated with unsafe practices (Anand et al. 2001; Khan et al. 2000), and the absence of an efficient sharps waste management system to prevent recycling of contaminated equipment (Hofmann 2001). Of interest, our analysis suggests that injection practices are safer in sub-Saharan Africa (19% and 17% of reuse in AFR-D and AFR-E, respectively) than in the Middle East and south Asia (70% and 75% reuse in EMR-D and SEAR-D, respectively). The proportion of the population aware of the potential risk of HIV infection through unsafe injections was 24% in Pakistan in 1998 (Luby et al. forthcoming), 19% in India in 1999 (Anand et al. 2001) and 52% in Burkina Faso in 2001 (Logez 2001). In addition, the social and economic consequences of the HIV pandemic have been perceived more acutely on the African continent than in Asia. Thus, a higher awareness regarding the risks of HIV infection associated with unsafe injections in sub-Saharan Africa (Birungi 1998) may partly explain this difference observed in the proportion of reuse.

HBV infection was the most common consequence of unsafe injection practices in the world, with more than 20 million cases of infection annually. Among the three pathogens that we examined, HBV is the most prevalent globally (Maynard et al. 1989) and the one most easily transmitted through unsafe injections (Seeff et al. 1978). The subregion-specific fractions of new HBV infections attributable to contaminated injections were compatible with those reported in analytical studies, including 2% (Thuring et al. 1993) to 73.9% (Ko et al. 1991a) in WPR-B (compared with 33.6% in our model), 49.7% (Singh et al. 2000) to 53.3% (Narendranathan and Philip 1993) in SEAR-D (compared with 53.6% in our model) and 27.7% (Anonymous 1998) to 52% (Hussain 2001) in EMR-D (compared with 58.3% in our model). Because attributable fractions were also high among children aged <5 years, a substantial proportion of unexplained transmission of HBV among preschool children may thus be attributed to unsafe injections

(Davis et al. 1989). Such infections would entail a substantial burden of disease and death in the future since the long-term consequences of HBV infections are most severe among persons infected during childhood (McMahon et al. 1985). The natural history of the infection is well described for HBV and there was relatively little uncertainty around the disease progression parameters that we used.

The burden of disability and death secondary to injection-associated HBV infections was estimated to be low in comparison to the number of infections because of the low proportion of progression to chronic infections and the delay between infection and death. This delay between initial infection and death reduced the burden because of the 30-year horizon of this work and because background mortality would lead infected persons to die from other causes during this time interval. In addition, the burden of disease avoidable through the control of contaminated injections as a risk factor is limited because universal childhood immunization against hepatitis B is being increasingly introduced in resource-poor countries with the support of the Global Alliance for Vaccines and Immunization (GAVI). If high vaccination rates are indeed reached in the future, as in our assumptions for 2030, herd immunity will progressively reduce the incidence of injection-associated infections through a high prevalence of immunity and a low prevalence of active infection (Wittet 2001).

HCV infection was estimated to be the second most common consequence of contaminated injections worldwide, with more than two million infections each year. Injection-associated HCV infection was less common than injection-associated HBV infection because of the lower prevalence of HCV infection (WHO 2000a) and the lower percutaneous transmission potential of HCV (CDC 1997). However, the fraction of new HCV infection attributable to contaminated injections was high among all age groups, including adults, and was higher than for HBV infection. These high attributable fractions are compatible with those reported in analytical studies, including 20.1% (Chen et al. 1995) to 90.6% (Thuring et al. 1993) in WPR-B (compared with 37.6% in our model), and 9.9% (Mohamed et al. 1996) to 87.9% (El-Sakka 1997) in EMR-D (compared with 81.7% in our model).

HCV is primarily transmitted through percutaneous or permucosal exposure to blood (Alter 1995). Transmission among sexual partners occurs but is not efficient (Alter 1995), occurring mostly among individuals engaging in high-risk sexual behaviour that may expose them to blood, or body fluids contaminated with blood (Williams et al. 1999). Sexual transmission may account for a higher proportion of HCV infections in industrialized countries, where contaminated health care injections and other unsafe percutaneous procedures are uncommon (Alter et al. 1999). However, in developing and transitional countries, the exposures most likely to transmit HCV include unsafe injections, transfusion of blood, blood components and blood products, and other unsafe

percutaneous exposures conducted in medically-related and other settings (e.g. dental care, surgery, circumcision, shaving).

The risk of HCV infection following transfusion of contaminated blood—about 92% (Aach et al. 1991)—greatly exceeds the risk of HCV infection following a contaminated injection. However, transfusion of blood, blood components and blood products is an infrequent exposure in comparison with injections. Annually, worldwide, it is estimated that over 75 million blood donations occur (WHO 2000b). Our analysis suggested that for developing and transitional countries alone, more than 16 thousand million injections might occur annually. Thus, despite a lower risk associated with each unsafe event, the higher frequency of injections explains why, globally, our analysis suggested that a high proportion of HCV infection was attributable to contaminated injections. While percutaneous exposures other than injections have been associated with HCV infection in epidemiological studies, they rarely explain a high proportion of infections (Wasley and Alter 2000).

HCV infection is currently not preventable through immunization and, in contrast to HBV infection, its long-term consequences may be more severe among persons infected during adulthood than among persons infected during childhood (Alter and Seeff 2000; Vogt et al. 1999). To estimate the burden of disease secondary to current injection-associated HCV infections, we used conservative estimates for the parameters describing the progression of HCV infection towards chronic liver disease and its consequences. However, there is substantial uncertainty as to whether these estimates obtained from studies conducted in industrialized countries are representative of the natural history of HCV infection in developing and transitional countries as environmental factors could influence the risk of progression to chronic liver diseases and its consequences. In addition, because the parameters that we selected assume that infected patients only die after 20 years, the 30-year analytic horizon of this work only captured a small proportion of the future early deaths. Nevertheless, if the parameters used in our model were indeed representative, our analysis would suggest that injection-associated HCV infections would not constitute a major avoidable burden of disease between 2000 and 2030. If we underestimated the severity of the natural history of HCV infection, then the burden of chronic liver disease and death secondary to injection-associated HCV infection that could be avoided in the future, particularly in countries highly endemic for HCV, would be estimated as substantial.

Historically, health care injections have not been viewed as a major vehicle of HIV infection and the promotion of safe injection practices has not held a high priority in HIV prevention programmes. However, most nosocomial outbreaks of HIV infection have been reported from countries with low prevalence of HIV infection, including Colombia (Shaldon 1995), the Libyan Arab Jamahiriya (Yerly et al. 2001), Romania (Hersh et al. 1993) and Ukraine (Simonsen et al. 1999). In

other countries where HIV infection and poor injection practices are more common and where sexual transmission accounts for the majority of infections, injection-associated HIV infections are likely to occur but they have rarely been reported or suspected.

Our analysis suggests that contaminated injections may cause 5.4% of new cases of HIV infection worldwide, representing the largest burden of disease that could be avoided through safe and appropriate use of injection policies. Few epidemiological studies are available with which to validate our estimates, either because transmission through injections was not examined or because these studies were not based upon recent, incident HIV infections. This lack of information represents a substantial source of uncertainty. In AFR-E where several studies were available (Bultreys et al. 1994; N'Galy et al. 1988; Quigley et al. 2000; Wawer et al. 1994), the lowest attributable fraction (8%) calculated on the basis of the data provided by Wawer et al. (1994) largely exceeded our estimate of 2.5%. In EMR-D and SEAR-D, the model suggests that the attributable fraction could reach 7.1% and 24.3%, respectively. These estimates are not validated by epidemiological studies and may be overestimated because the epidemic is still concentrated, violating the assumptions made in the mass action model about the distribution of contaminated injections. Despite large uncertainty in attributable fractions in these two subregions, the high frequency of unsafe injection practices coinciding with emerging HIV epidemics must lead to urgent preventive measures.

In the future, studies of the risk factors for HIV infection should ensure that data are collected in a way that allows examination of the association between HIV infection and injections. In the meantime, HIV prevention programmes should communicate the risk of HIV infection associated with health care injections since safe and appropriate use of injection policies constitute effective interventions against HIV infection (CDC 2001; Logez 2001; Prawitasari Hadiyono et al. 1996).

While much emphasis was put on gathering the best available data from published and unpublished sources, this analysis has a number of limitations due to data scarcity.

- Our model was constrained by the limited number of studies with adequately described injection practices. Moreover, some of these studies employed non-standardized methodologies, which could not be used. The high frequency of injections reported in developing and transitional countries contrasts with the paucity of data available to describe practices. Until recently, few standardized tools for assessment or evaluation were available to routinely collect information on injection frequency or safety. However, WHO has recently developed new assessment tools that utilize standardized methods, which will prospectively generate information of appropriate quality. This should

allow future revisions of these burden of disease estimates to be based upon data of better quality.

- The transmission potential of HBV, HCV and HIV through percutaneous exposure was obtained on the basis of epidemiological studies that estimated the risk of infection with bloodborne pathogens among health care workers following a needle-stick injury. Although these figures are based upon many well-conducted studies that included a large number of study participants, they estimated a different risk: infection associated with a needle-stick injury. Factors that could cause the risk from contaminated injections to be higher than the risk from needle-stick injuries include the liquid flow rinsing the needle that occurs during an injection and the potential survival of HBV and HCV in the pots of tepid water often used to rinse injection equipment between injections. Factors that could cause the risk from contaminated injections to be lower than the risk from needle-stick injuries include a longer time interval between injections that could cause inactivation of some virus particles and a dilution effect in the pots of tepid water when injection equipment is reused.
- Our model only estimated the incidence of infections with HBV, HCV and HIV secondary to the reuse of injection equipment on one patient. It did not take into account the transmission secondary to the reuse of equipment on multiple patients, the transmission associated with unhygienic use of multi-dose medication vials and the transmission that may occur through cross-contamination while preparing injections. Failure to address these specific unsafe practices may have led to an underestimation in our results.
- Our analysis did not take into account any transmission networks by which injection frequencies, background prevalence of infection or probability of exposure to unsafe practices, were assessed. As a result, exposure to contaminated injections would not be distributed independently, thus creating various population subgroups with different bloodborne pathogen transmission dynamics. This would be particularly important in settings with concentrated groups of infected persons (e.g. persons with HIV in SEAR-D). However, we excluded persons presenting with very high injection frequencies to calculate the subregional injection frequency input parameters and adjusted the model for the possibility that persons receiving a high number of injections could already have immunity against infection with bloodborne pathogens. Further a potential network effect could involve percutaneous and sexual transmission (e.g. use of injected antibiotics among commercial sex workers), thereby transforming a dendritic transmission network into a more effective cyclic one (Potterat et al. 1999; Rothenberg et al. 1998).

Safe and appropriate use of injection policies aims to eliminate unnecessary injections and achieve safe injection practices. Such initiatives should not constitute separate programmes but should be integrated with other activities (WHO 2000c) to provide more effective interventions including:

- communication of risks associated with unsafe injections to patients and health care workers through disease prevention programmes such as HIV prevention;
- ensuring access to sufficient quantities of single-use, disposable injection equipment in health care facilities; and
- management of sharps waste to prevent reuse of dirty equipment and needle-stick injuries.

This study generated initial estimates of the burden of disease that could be avoided through the implementation of such policies. Further studies will address sources of uncertainty that remain in the natural history of HCV infection and in the proportion of HIV infections attributable to contaminated injections.

5. Projections of future levels of exposure

The future prevalence of exposure to contaminated injections and the future relative risk of HBV, HCV or HIV infection associated with contaminated injections were calculated by including different input parameters into the same model. Assumptions regarding the injection practice parameters in 2030, including the annual number of injections per person and proportion of reuse, were generated through a survey of nine experts from four of the six WHO regions. These projections took into account the high effectiveness of planned or implemented interventions aimed at improving the safety of injections, the moderate effectiveness of interventions aimed at decreasing injection overuse, the prospective for future increased access to health care in sub-Saharan Africa (which could lead to an increase in injection use), and the potential for health care reform in the former socialist countries of eastern Europe and central Asia (which could lead to a decrease in injection overuse). Assumptions regarding the expected number of injections per person and the proportion of reuse were compatible with a slight decrease of injection frequency and a marked improvement of injection safety, although subregions were expected to remain heterogeneous (Table 22.10).

Epidemiological parameters of HBV infection were modified to account for the expected increased use of hepatitis B vaccine secondary to the accelerated introduction of this vaccine into immunization programmes supported by GAVI: the three-dose vaccine coverage was assumed to be 90% among persons aged <15 years and 50% among persons aged 15–29 years. The prevalence of active infection in the pop-

Table 22.10 Assumptions regarding projected injection practices in 2030

	AFR-D	*AFR-E*	*AMR-B*	*AMR-D*	*EMR-D*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-B*
Proportion of reuse (%)	8	7	0	4	25	0	4	8	17	9
Annual no. of injections per person	2.3	2.2	1.2	1.9	3.0	1.6	2.2	1.7	2.4	1.8

ulation was assumed to be 2% in all subregions, except for AMR-B and AMR-D, where it was assumed to be 0.5%. The annual incidence among susceptible individuals was assumed to be 0.1% in all subregions, except for AMR-B and AMR-D where it was assumed to be 0.01%. In the absence of available epidemiological projections, the incidence and the prevalence of HCV and HIV were assumed to remain constant. Changes in parameters were assumed to be linear between 2000 and 2030.

Our projections into the future of the risk of infection with bloodborne pathogens associated with exposure to contaminated injections did not take into account the dynamic effect of new injection-associated infections on the prevalence of infection with bloodborne pathogens. Our model, a Bernoulli risk projection model, is more adapted to the estimation of the current and past incidence of injection-associated infections. In the case of HIV, where contaminated injections account for only a small proportion of new infections and the pandemic continues to be largely driven by sexual transmission, this limitation is unlikely to substantially affect our results. In the case of HBV, contaminated injections account for a substantial proportion of new infections. The impact of prevention policies on the future burden of disease is likely to be underestimated because prevented cases of HBV infection will reduce the pool of chronically infected persons who constitute sources of infection. However, in countries of intermediate or high HBV endemicity, age-specific prevalence of infection and historical data suggest that the endemicity level has not substantially changed over the past decades, and there is no evidence of injection practices playing a major role in the introduction of HBV in a community.

In contrast, in the case of HCV, where the proportion of infections attributable to injections is high, the effect of this limitation is likely to be considerable. In addition, there is evidence that in some countries, including China (Province of Taiwan) (Sun et al. 1999), Egypt (Frank et al. 2000) and Pakistan (Luby et al. 1997), HCV was recently introduced,

largely through contaminated injections, and rapidly reached high prevalence levels. In fact, in some of these countries, the prevalence is heterogeneous and areas persist where the virus has not been widely introduced (Mujeeb et al. 2000; Sun et al. 1999). In subregions where reuse of injection equipment is common but the prevalence of HCV infection is not yet high (e.g. SEAR-D), there is an opportunity at present to prevent future community-wide outbreaks of HCV infections. Our model does not reflect this opportunity.

Acknowledgements

We are grateful to Steve Luby for critical comments on the methodology and on the manuscript; to David Gisselquist for assistance with the review of epidemiological studies that examined the association between injections and HIV infection; and to John Potterat for assistance in addressing the network effect issue.

Notes

1 See preface for an explanation of this term.

2 Prepared for *World health report 2001* (WHO 2001).

References

Aach RD, Stevens CE, Hollinger FB et al. (1991) Hepatitis C virus infection in post-transfusion hepatitis. *New England Journal of Medicine*, **325**:1325–1329.

Abulrahi HA, Bohlega EA, Fontaine RE, Al-Seghayer SM, Al-Ruweis AA (1997) Plasmodium falciparum malaria transmitted in hospital through heparin locks. *The Lancet*, **349**:23–25.

Alter MJ (1995) Epidemiology of hepatitis C in the West. *Seminars in Liver Disease*, **15**:5–14.

Alter MJ, Kruszon-Moran D, Nainan OV et al. (1999) The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *New England Journal of Medicine*, **341**:556–562.

Alter HJ, Seeff LB (2000) Recovery, persistence and sequelae in hepatitis C virus infection: a perspective on long-term outcome. *Seminars in Liver Disease*, **20**:17–35.

Alward WL, McMahon BJ, Hall DB et al. (1985) The long-term serological course of asymptomatic hepatitis B virus carriers and the development of primary hepatocellular carcinoma. *Journal of Infectious Diseases*, **151**: 604–609.

Anand K, Pandav CS, Kapoor SK, Undergraduate study team (2001) Injection use in a village in north India. *National Medical Journal of India*, **14**: 143–144.

Anonymous (1945) Role of syringes in the transmission of jaundice. A memorandum by medical officers of the Ministry of Health. *The Lancet*, **2**:116–119.

Anonymous (1998) Possible risk factor for acute hepatitis B infection, case control study, Qena and Luxor city fever hospitals, 1994. *Epidemiology Bulletin—Egypt*, **1**:6–8.

Archibald LK, Ramos M, Arduino MJ et al. (1998) Enterobacter cloacae and Pseudomonas aeruginosa polymicrobial bloodstream infections traced to extrinsic contamination of a dextrose multidose vial. *Journal of Pediatrics*, **133**:640–644.

Aylward B, Kane M, McNair-Scott R, Hu DJ (1995) Model-based estimates of the risk of human immunodeficiency virus and hepatitis B virus transmission through unsafe injections. *International Journal of Epidemiology*, **24**: 446–452.

Bass A (2000) SIGNPOST: the safe injection global network internet forum listserve. *Bulletin of the World Health Organization*, **18**:277.

Birungi H (1998) Injections and self-help: risk and trust in Ugandan health care. *Social Science and Medicine*, **47**:1455–1462.

Bultreys M, Chao A, Habimana P et al. (1994) Incident HIV-1 infection in a cohort of young women in Butare, Rwanda. *AIDS*, **8**:1585–1591.

Busch MP, Operskalski EA, Mosley JW et al. (1996) Factors influencing human immunodeficiency virus type 1 transmission by blood transfusion. *Journal of Infectious Diseases*, **174**:26–33.

Cardo DM, Culver DH, Ciesielski CA et al. (1997) A case-control study of HIV seroconversion in health care workers after percutaneous exposure. *New England Journal of Medicine*, **337**:1542–1543.

CDC (Centers for Disease Control and Prevention) (2001) Injection practices among nurses—Valcea, Romania, 1998. *Morbidity and Mortality Weekly Report*, **50**:59–61.

CDC (1997) Recommendations for follow-up of health care workers after occupational exposure to hepatitis C virus. *Morbidity and Mortality Weekly Report*, **46**:603–606.

CDC (1999) Frequency of vaccine-related and therapeutic injection—Romania 1998. *Morbidity and Mortality Weekly Report*, **48**:271–274.

Chang SJ, Chen HC, Ying J, Lu CF, Ko YC (1996) Risk factors of hepatitis C virus infection in a Taiwanese aboriginal community. *Kao Hsiung I Hsueh Ko Hsueh Tsa Chih*, **12**:241–247.

Chen TZ, Wu JC, Yen FS et al. (1995) Injection with nondisposable needles as an important route for transmission for acute community-acquired hepatitis C virus infection in Taiwan. *Journal of Medical Virology*, **46**: 247–251.

Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M (1989) Isolation of a cDna clone derived from a bloodborne non-A, non-B viral hepatitis genome. *Science*, **244**:359–362.

Davis LG, Weber DJ, Lemon SM (1989) Horizontal transmission of hepatitis B virus. *The Lancet*, **1**:889–893.

Deivanayagam N, Nedunchelian K, Ahamed SS, Ashok TP, Mala N, Ratnam SR (1993) Intramuscular injection as a provoking factor for paralysis in acute poliomyelitis. A case control study. *Indian Pediatrics*, **30**:335–340.

Dicko M, Oni AQ, Ganivet S, Kone S, Pierre L, Jacquet B (2000) Safety of immunization injections in Africa: not simply a problem of logistics. *Bulletin of the World Health Organization*, **78**:163–169.

Drucker E, Alcabes PG, Marx PA (2001) The injection century: massive unsterile injections and the emergence of human pathogenes. *The Lancet*, **358**: 1989–1992.

El-Sakka H (1997) *Risk factors for HCV infection. Gharbia Governorate, Egypt, 1996–1997*. (Presentation given at the EIS conference, April 1997.) Atlanta, GA.

Ferry B (1995) Risk factors related to HIV transmission. In: *Sexual behaviour and AIDS in the developing world*. Cleland J, Ferry B, eds. World Health Organization, Geneva.

Fisher-Hoch SP, Tomori O, Nasidi A et al. (1995) Review of cases of nosocomial Lassa fever in Nigeria: the high price of poor medical practice. *British Medical Journal*, **311**:857–859.

Flaskerud JH, Nyamathi AM (1996) Home medication among Latina women in Los Angeles: implications for health education and prevention. *AIDS Care*, **8**:95–102.

Fontaine O, Diop B, Beau JP (1984) La diarrhee infantile au Senegal [Childhood diarrhoea in Senegal]. *Medicina Tropical*, **44**:27–31.

Frank C, Mohamed MK, Strickland GT (2000) The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *The Lancet*, **355**:887–891.

Freeman AJ, Dore GJ, Law MG et al. (2001) *The natural history of chronic hepatitis C: a systematic review*. (Unpublished manuscript communicated by Dr Gregory Dore, April 2001.) University of New South Wales, Sydney.

Gay NJ, Edmunds WJ, Bah E (2001) *Estimating the global burden of hepatitis B*. (Unpublished manuscript prepared for the Global Burden of Disease study, April 2001.) World Health Organization, Department of Vaccines and Biologicals, Geneva.

Grosskurth H, Mosha F, Todd J et al. (1995) A community trial of the impact of improved sexually transmitted disease treatment on the HIV epidemic in rural Tanzania. 2. Baseline survey results. *AIDS*, **9**:927–934.

Guyer B, Bisong AAE, Gould J, Brigoud M, Aymard M (1980) Injections and paralytic poliomyelitis in tropical Africa. *Bulletin of the World Health Organization*, **58**:285–291.

Hersh BS, Popovici F, Jezek Z et al. (1993) Risk factors for HIV infection among abandoned Romanian children. *AIDS*, **7**:1617–1624.

Ho M-S, Hsu C-P, Yuh Y et al. (1997) High rate of hepatitis C virus infection in an isolated community: persistent hyperendemicity or period-related phenomena? *Journal of Medical Virology*, **52**:370–376.

Hofmann CA (2001) *Presentation made at the annual meeting of the Safe Injection Global Network*. World Health Organization, Geneva. (Unpublished document WHO/BCT/DCT/ 01.04; available on request from the Department of Essential Health and Technologies, World Health Organization, 1211 Geneva 27, Switzerland.)

Hsu SC, Chang MH, Ni YH, Hsu HY, Lee CY (1993) Horizontal transmission of hepatitis B virus in children. *Journal of Pediatric Gastroenterology and Nutrition*, **16**:66–69.

Hussain R (2001) *A case-control study of risk factors for acute hepatitis B virus infection in Karachi, Pakistan* [Dissertation]. Aga Khan University, Karachi.

Hutin YJF, Chen RT (1999) Injection safety: a global challenge. *Bulletin of the World Health Organization*, **77**:787–788.

Hutin YJF, Harpaz R, Drobeniuc J et al. (1999) Injections given in healthcare settings as a major source of acute hepatitis B in Moldova. *International Journal of Epidemiology*, **28**:782–786.

Janszen E, Laning W (1993) Injections: more expensive, more effective and faster. A report from Ecuador. In: *The impact of injections on daily medical practice*. Bloem M, Wolfers I, eds. VU University Press, Amsterdam.

Kane A, Lloyd J, Zaffran M, Simonsen L, Kane M (1999) Transmission of hepatitis B, hepatitis C and human immunodeficiency viruses through unsafe injections in the developing world: model-based regional estimates. *Bulletin of the World Health Organization*, **77**:801–807.

Kaninda AV (2001) *Frequency of injections worldwide: market data analysis and injections given for selected purposes*. (Unpublished WHO report, Department of Blood Safety and Clinical Technology.) World Health Organization, Geneva.

Khan AJ, Luby SP, Fikree FF et al. (2000) Unsafe injections and the transmission of hepatitis B and C in a periurban community in Pakistan. *Bulletin of the World Health Organization*, **78**:956–963.

Ko YC, Chung DC (1991) Transmission of hepatitis B virus infection by iatrogenic intramuscular injections in an endemic area. *Kao Hsiung I Hsueh Ko Hsueh Tsa Chih*, **7**:313–317.

Ko YC, Chung DC, Pai HH (1991b) Intramuscular-injection-associated gluteal fibriotic contracture and hepatitis B virus infection among school children. *Kao Hsiung I Hsueh Ko Hsueh Tsa Chih*, **7** :358–362.

Ko YC, Li SC, Yen YY, Yeh SM, Hsieh CC (1991a) Horizontal transmission of hepatitis B virus from siblings and intramuscular injection among preschool children in a familial cohort. *American Journal of Epidemiology*, **133**:1015–1023.

Kosen S (1999) *Assessment on the safety of EPI injection practices in four provinces of Indonesia*. (Unpublished document.) Health Services and Development Center, Ministry of Health, Indonesia.

Lakshman M, Nichter M (2000) Contamination of medicine injection paraphernalia used by registered medical practitioners in south India: an ethnographic study. *Social Science and Medicine*, **51**:11–28.

Logez S (2001) *Accessibility to injection equipment: impact on injection safety*. (WHO Mission Report.) Pharmaciens Sans Frontières Comité International, Clermont-Ferrand.

Luby SB, Hoodboy F, Aziz M, Shah A (forthcoming) *Long term improvement of unsafe injection practices following community intervention*.

Luby SP, Qamruddin K, Shah AA et al. (1997) The relationship between therapeutic injections and high prevalence of hepatitis C infection in Hafizabad, Pakistan. *Epidemiology and Infection*, **119**:349–356.

Maynard JE, Kane MA, Hadler SC (1989) Global control of hepatitis B through vaccination: role of hepatitis B vaccine in the Expanded Programme on Immunization. *Reviews of Infectious Diseases*, **11**:S574–578.

McMahon BJ, Alward WL, Hall DB, Heyward WL, Bender TR, Francis DP (1985) Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. *Journal of Infectious Diseases*, **151**:599–603.

Mohamed MK, Hussein MH, Massoud AA et al. (1996) Study of the risk factors for viral hepatitis C infection among Egyptians applying for work abroad. *Journal of the Egyptian Public Health Association*, **71**:113–142.

Molina JM, Ferchal F, Chevret S et al. (1994) Quantification of HIV-1 virus load under ziovudine therapy in patients with symptomatic HIV infection: relation to disease progression. *AIDS*, **8**:27–33.

Mujeeb SA, Shahad S, Hyder AA (2000) Geographical display of health information: study of hepatitis C infection in Karachi, Pakistan. *Public Health*, **114**:413–415.

N'Galy B, Ryder RW, Bila K et al. (1988) Human immunodeficiency virus infection among employees in an African hospital. *New England Journal of Medicine*, **319**:1123–1127.

Narendranathan M, Philip M (1993) Reusable needles—a major risk factor for acute virus B hepatitis. *Tropical Doctor*, **23**:64–66.

Otu AA (1987) Hepatocellular carcinoma, hepatic cirrhosis and hepatitis B virus infection in Nigeria. *Cancer*, **60**:2581–2585.

Pape JW, Liautaud B, Thomas F et al. (1985) The acquired immunodeficiency syndrome in Haiti. *Annals of Internal Medicine*, **103**:674–678.

Potterat JJ, Rothenberg RB, Muth SQ (1999) Network structural dynamics and infectious disease propagation. *International Journal of STD and AIDS*, **10**:182–185.

Prawitasari Hadiyono JE, Suryawati S, Danu SS, Sunartono, Santoso B (1996) Interactional group discussion: results of a controlled trial using a behavioural intervention to reduce the use of injections in public health facilities. *Social Science and Medicine*, **42**:1177–1183.

Priotto G, Ruiz A, Kyobutungi C (2001) *Injection use in the population of Mbarara district, Uganda, cross sectional population survey.* Epicentre, Paris.

Quigley MA, Morgan D, Malamba SS (2000) Case-control study of risk factors for incident HIV infections in rural Uganda. *Journal of Acquired Immune Deficiency Syndromes*, **23**:418–425.

Quigley MA, Munguti K, Grosskurth H et al. (1997) Sexual behaviour patterns and other risk factors for HIV infection in rural Tanzania: a case-control study. *AIDS*, **11**:237–248.

Reeler AV (1990) Injections: a fatal attraction. *Social Science and Medicine*, **31**:1119–1125.

Reeler AV, Hematorn C (1994) *Injection practices in the third world*. World Health Organization, Geneva. (Unpublished document WHO/DAP/94.8; available on request from the Drug Action Programme, World Health Organization, 1211 Geneva, Switzerland.)

Rothenberg RB, Sterk C, Toomey KE et al. (1998) Using social network and ethnographic tools to evaluate syphilis transmission. *Sexually Transmitted Diseases*, **25**:154–160.

Schnurr A, Xiaojun W, Jingjin Y (1999) *Sterilizable syringes and needles use in China*. (Presentation made at the TechNet meeting, December 1999 in Harare, Zimbabwe.)

Seeff LB, Wright EC, Zimmerman HJ, Alter HJ, Dietz AA, Felsher BF (1978) Type B hepatitis after needle-stick exposure: prevention with hepatitis B immune globulin. *Annals of Internal Medicine*, **88**:285–293.

Senior JR, Sutnick AI, Goeser E, London WT, Dahlke MB, Blumberg BS (1974) Reduction of post-transfusion hepatitis by exclusion of Australia antigen from donor blood in an urban public hospital. *American Journal of the Medical Sciences*, **267**:171–177.

Shaldon S (1995) Transmission of HIV in dialysis centres. *The Lancet*, **346**:451.

Shikata T, Karasaw T, Abe K et al. (1977) Hepatitis B antigens and infectivity of hepatitis B virus. *Journal of Infectious Diseases*, **136**:571.

Simard E, Wasley A, Pistol A, Popa M, Hutin Y, Mast E (2000) *Surveillance for acute hepatitis B among children, Romania, 1997–1999*. (Presentation given at the American Public Health Association meeting, December 2000.) Boston, MA.

Simonsen L, Kane A, Lloyd J, Zaffran M (1999) Unsafe injections in the developing world and transmission of bloodborne pathogens. *Bulletin of the World Health Organization*, **77**:789–800.

Singh J, Bhatia R, Patnaik SK et al. (2000) Community studies on hepatitis B in Rajahmundry town of Andra Pradesh, India, 1997–1998: unnecessary therapeutic injections are a major risk factor. *Epidemiology and Infection*, **125**:367–375.

Soeters R, Aus C (1989) Hazards of injectable therapy. *Tropical Doctor*, **19**: 124–126.

Stoica A, Hutin YJF, Paun M, Mast EE, Margolis HS (1999) *Attitudes of physicians regarding the use of therapeutic injections, Arges district, Romania*. (Abstract presented at the annual meeting of the Society for Healthcare Epidemiology of America, April 1999.) San Francisco, CA.

Sun CA, Chen HC, Lu CF (1999) Transmission of hepatitis C virus in Taiwan: prevalence and risk factors based on a nationwide survey. *Journal of Medical Virology*, **59**:290–296.

Sun CA, Chen HC, Lu SN et al. (2001) Persistent hyperendemicity of hepatitis C virus infection in Taiwan: the important role of iatrogenic risk factors. *Journal of Medical Virology*, **65**:30–34.

Talaat M, Oun S, El-Halawani A, Bodenschatz C, Mahoney F (2001) *A population-based survey of injection practices among Egyptians*. (Proceeding of the regional FETP meeting), Amman.

Thuring EG, Joller-Jemelka HI, Sareth H, Sokhan U, Reth C, Grob P (1993) Prevalence of markers of hepatitis viruses A, B, C and of HIV in healthy individuals and patients of a Cambodian province. *Southeast Asian Journal of Tropical Medicine and Public Health*, **24**:239–249.

Ugalde A, Homedes N (1988) Medicines and rural health services: an experiment in the Dominican Republic. In: *The context of medicines in developing countries*. van der Geest S, Whyte SR, eds. Kulwer Academic Publishers, Dordrecht.

UNAIDS (2000) *Global HIV epidemic update: December 2000*. UNAIDS, Geneva.

Val Mayans M, Hall AJ, Inskip HM et al. (1990) Risk factors for transmission of hepatitis B virus to Gambian children. *The Lancet*, **336**:1107–1109.

van Staa A, Hardon A (1996) *Injection practices in the developing world*. World Health Organization, Geneva. (Unpublished document WHO/DAP/96.4; available on request from the Drug Action Programme, World Health Organization, 1211 Geneva, Switzerland.)

Villanueva A, Calderon RV, Vargas BA et al. (1997) Report on an outbreak of postinjection abscesses due to Mycobacterium abscessus, including management with surgery and clarithromycin therapy and comparison of strains by random amplified polymorphic DNA polymerase chain reaction. *Clinical Infectious Diseases*, **24**:1147–1153.

Vogt M, Lang T, Frosner G et al. (1999) Prevalence and clinical outcome of hepatitis C infection in children who underwent cardiac surgery before the implementation of blood-donor screening. *New England Journal of Medicine*, **341**:866–870.

Vong S, Perz J, Hutin Y, Drobeniuc J, Bell BP, 1998 International Field Epidemiology Course Participants (2002) *Determinants of high frequency of therapeutic injections, Chisinau, Republic of Moldova*. (Abstract presented at the 51st Epidemic Intelligence Service conference, April 2002) Centers for Disease Control and Prevention, Atlanta, GA.

Wasley A, Alter MJ (2000) Epidemiology of hepatitis C: geographic differences and temporal trends. *Seminars in Liver Disease*, **20**:1–16.

Wawer MJ, Sewankambo NK, Berkley S et al. (1994) Incidence of HIV-1 infection in a rural region of Uganda. *British Medical Journal*, **308**:171–173.

WHO (1976) Ebola haemorrhagic fever in Zaire. *Bulletin of the World Health Organization*, **56**:271–293.

WHO (1999) High frequency of therapeutic injections, Republic of Moldova. *Weekly Epidemiological Record*, **11**:84–86.

WHO (2000c) *"Aide Mémoire" for a national strategy for the safe and appropriate use of injections*. World Health Organization, Department for Blood Safety and Clinical Technology, Geneva.

WHO (2000b) *Blood safety for too few*. (Press release 2000/25, April 7.) World Health Organization, Geneva.

WHO (2000a) Hepatitis C: global prevalence (Update). *Weekly Epidemiological Record*, **75**:18–19.

WHO (2001) *World health report 2001*. World Health Organization, Geneva.

WHO (2002) *Tool for the assessment of injection safety*. World Health Organization, Geneva. (Unpublished document WHO/V&B/01.30; available on request from the Department of Immunization, Vaccines and Biologicals, World Health Organization, 1211 Geneva 27, Switzerland.)

Williams CL, Mast EE, Coleman PJ et al. (1999) *Sexual practices associated with hepatitis C virus infection among non injecting-drug-using female prostitutes in the United States*. (Program and abstracts of the 6th International Symposium on Hepatitis and Related Viruses: Molecular Virology and Pathogenesis, June 6–9.) National Institutes of Health, Bethesda, MD.

Wittet S (2001) *Introducing hepatitis B vaccine*. (Immunization focus, March 6–7.) Available at http://www.vaccinealliance.org.

Wyatt HV (1984) The popularity of injections in the third world: origins and consequences for poliomyelitis. *Social Science and Medicine*, **19**:911–915.

Yerly S, Quadri R, Negro F et al. (2001) Nosocomial outbreak of multiple bloodborne viral infections. *Journal of Infectious Diseases*, **184**:369–372.

Chapter 23

Child sexual abuse

Gavin Andrews, Justine Corry, Tim Slade,
Cathy Issakidis and Heather Swanston

Summary

This chapter summarizes the evidence of a relationship between child sexual abuse and subsequent mental disorder. Child sexual abuse (CSA) typically includes unwanted and inappropriate sexual solicitation of, or exposure to, a child by an older person; genital touching or fondling; or penetration in terms of oral, anal or vaginal intercourse or attempted intercourse. CSA can vary along a number of dimensions including frequency, duration, age at onset, and relationship of victim to perpetrator. However, the most common dimension used to define CSA is type of abuse. Three categories are commonly reported in the literature. Non-contact abuse encompasses a range of acts and includes inappropriate sexual solicitation or indecent exposure. The two other categories are contact abuse, which includes touching or fondling, and intercourse, which includes oral, anal or vaginal intercourse. In this chapter the upper age limit used to define childhood was 18 years. The theoretical minimum exposure was defined as no abuse.

The disease outcomes chosen for the current analysis were depression, panic disorder, post-traumatic stress disorder (PTSD), alcohol abuse/dependence, drug abuse/dependence and suicide attempts. Evidence for causality came from twin studies, prospective studies and representative community studies. In particular, the three twin studies available provided strong evidence of a causal relationship as they inherently controlled for the genetic and family environment factors that are also associated with mental disorders (Dinwiddie et al. 2000; Kendler et al. 2000; Nelson et al. 2002). These studies provide evidence of significant associations between CSA and depression, panic disorder, alcohol abuse/dependence, drug abuse/dependence and suicide attempts. While PTSD was not considered as an outcome in the twin studies, data from a prospective study (Silverman et al. 1996) and three representative community studies (Davidson et al. 1991; Molnar et al. 2001; Saunders

et al. 1999) provided consistent evidence that a strong association exists. There was insufficient evidence to support the relationship with obsessive–compulsive disorder (OCD) and so it was excluded. A number of other mental disorder outcomes that were not considered here have been linked to child sexual abuse. Eating disorders have long been conceptualized as a response to a dysfunctional family environment, of which child sexual abuse can be a part. A number of studies have also looked at the association between child sexual abuse and personality disorders, particularly antisocial and borderline personality disorder. Additionally, child sexual abuse does not only produce an increased risk of mental disorder. There has been anecdotal and experimental evidence suggesting that CSA increases the probability of negative psychological outcomes such as poor self-esteem, lack of a sense of control or agency, difficulties with intimacy and continuing sexual difficulties.

Review articles on the prevalence of child sexual abuse have commonly reported a range of prevalence anywhere from 2% to 62%. Previous studies have demonstrated a significant difference in the prevalence of CSA, depending on a number of methodological factors, including method of data collection, number of questions used to assess CSA, definition of childhood and the type of sample assessed. Therefore, multiple linear regression analyses were employed separately for males and females to identify the methodological characteristics that significantly contribute to the variation in prevalence. Any unwanted variation was removed by adjusting individual prevalence estimates accordingly. These "adjusted" prevalence estimates were then broken down into the three levels of abuse (non-contact, contact and intercourse) and the eight age groups. Meta-analysis was then used to combine the estimates within each country. These country estimates were weighted by the population of the country and combined to provide the final subregional[1] estimates. Prevalence estimates were higher in females than in males and varied across subregions with the highest prevalence estimates found in AFR-E and SEAR-D.

Relative risk estimates were derived from studies examining the relationship between CSA and psychiatric outcomes. Where relative risks were reported only for overall exposure to CSA, these were extrapolated into the three levels of abuse. Few studies adjusted for the confounding effects of family dysfunction. An external method of adjustment was applied to the relative risks from those studies that did not control for family dysfunction. The small number of studies available for analysis meant that separate relative risks for different age, sex and subregional groups could not be obtained. Relative risks were stratified by psychiatric outcome before being combined using meta-analysis. Results showed that the relative risks are not significantly different across types of mental disorder suggesting that CSA is not particularly associated with any one disorder. Additionally, across types of abuse there was a general trend for increased risk to be associated with "increased" exposure to CSA.

That is, as more severe forms of CSA were experienced the risks for developing a mental disorder increased. After external adjustment the contact and intercourse categories of abuse remained significant across most disorders but non-contact abuse became non-significant.

Across the world CSA contributed to between 4% and 5% of the burden of disease in males and between 7% and 8% of the burden of disease in females, for each of the conditions depression, alcohol abuse/dependence and drug abuse/dependence. The attributable fractions were higher for panic disorder (7% for males and 13% for females) and higher still for PTSD (21% for males and 33% for females). For suicide attempts attributable fractions were 6% for males and 11% for females. There were slight regional variations in the amount of burden that could be attributed to CSA, with AFR-E and SEAR-D having higher attributable fractions. Prevalence was estimated to be higher in these subregions. However, data for these subregions came from a few studies that were methodologically poor. The burden of disease attributed to CSA was greater in the younger age groups and declined in the older age groups. Since risk was assumed to be constant across age, this merely reflected the age distribution of the mental disorder disease burden, which impacts largely on the younger age groups due to its early onset and chronic nature. Avoidable burden would be the same as attributable since it was assumed that the prevalence of CSA does not change over time under a "business-as-usual" scenario.

1. Introduction

1.1 Definitions of risk factor

Definitions of exposure to child sexual abuse vary. Patterns of interpersonal behaviour are being described, not the results of measuring a physical attribute like body mass index, hypertension or blood lead levels. In physical inactivity we accept the report of a person's habitual behaviour and calculate the health consequences of that behaviour to that person. In CSA we quantify one person's behaviour with a child, and usually have to rely on the retrospective report of that child when adult. We then estimate the health consequences of that occurrence. It is important to point out that, even in prospective studies, data on CSA are gathered retrospectively. It is unethical, and in many countries illegal, to prospectively identify CSA and not intervene. The problems of measurement have proven to be difficult but not insurmountable.

In its broadest sense CSA includes unwanted and inappropriate sexual solicitation of, or exposure to, a child by an older person (non-contact abuse), genital touching or fondling (contact abuse), and penetration in terms of oral, anal or vaginal intercourse or attempted intercourse (intercourse). Many studies have used a narrow definition of CSA to include contact abuse and intercourse only. Definitions of CSA also differ

depending on the cut-off age used to define childhood. While in most countries 18 years of age is the legal cut-off used to define childhood, in many countries the age of consent, especially for sexual activity, is lower. However, the most widely-reported definition of childhood in large population surveys of CSA is 18 years or less. Very few studies provided estimates of prevalence by different age groups and none provided estimates of mental disorder risk by different age groups. Furthermore, as described below, it has been shown that the first onset of CSA is less likely to occur between the ages of 15 and 18 years than in younger children. For these reasons the cut-off age used to define childhood in this chapter was set at 18 years.

1.2 Choice of exposure variable, reasons and implications

The causes of adult mental disorders have proven difficult to define. There is considerable evidence from longitudinal and twin studies that both genetic and environmental factors are implicated in different proportions in different disorders. Because we are dealing with human behaviour there is also evidence of substantial gene–environment interaction (Kendler et al. 2000; Rutter 1999). Advances in medicine generally have been simplified by the availability of animal models of the condition, which allow the biology to be explicated. Animal models of mental disorders just do not exist, and progress in understanding has to be made by association and inference and does not come from experimental paradigms.

It is simplistic to assume that genetic contributions are immutable. Few would claim that the genetic factors are, except in the case of Huntington chorea and some dementias, a full and sufficient explanation. Genetic factors act to enhance vulnerability. For example, the indicated prevention programmes in anxiety disorders almost certainly work by inhibiting the expression of that genetic vulnerability. Nevertheless, it is inherently more plausible to identify a risk factor that is purely environmental, as CSA is, when looking for a risk factor that might be avertible.

In broad terms, the risk factors for mental disorders can be grouped into vulnerability produced by temperament, by adversity and by deprivation. Trauma and CSA are examples of adversity; family dysfunction and neglect are examples of deprivation (Bryant and Range 1995; Kessler et al. 1997). Temperament, adversity and deprivation co-occur more often than is predicted by chance and, as these impact throughout childhood and adolescence in complex ways, their influence in adult functioning is not likely to be simple or linear (Mullen et al. 1996, 2000; Rutter 1999).

CSA is no exception to this rule. It is more frequent in situations in which the other factors are present. Fortunately, it is possible to estimate the independent contribution of CSA to adult mental disorders from both

twin studies on twin pairs discordant for CSA (where both genetics and family environment like deprivation and other adversity are controlled by virtue of the twin method), and longitudinal studies of young people growing up (where data were prospectively gathered on temperament, non-CSA adversity and deprivation) (Borowsky et al. 1999; Kendler et al. 2000; Molnar et al. 2001; Mullen et al. 1993, 1996; Yama et al. 1995; Zuravin and Fontanella 1999). The link between CSA and adult mental disorders has been established, even after controlling for these other determinants of adult mental disorders.

Two other issues make CSA an appropriate exposure variable to choose. First, it is not rare—most reviews have concluded that close to one child in six experiences an episode of CSA as defined using the broad definition above (Fergusson and Mullen 1999). Second, there is an extensive literature on CSA, much of it recent and much using the definitions presented above, which made a systematic review possible. The literature on deprivation or other adversity is nowhere near as extensive, coherent or accessible.

1.3 Choice of theoretical minimum

Given the nature of CSA and the way it is defined in this chapter, the only acceptable theoretical minimum is zero.

2. Estimating risk factor levels

2.1 Methods

The methods used to identify sources and studies for estimation of both risk factor levels and risk factor–disease relationships are presented in the following section.

2.2 Criteria for considering sources and studies

The following inclusion criteria were used:

- any study which determined the prevalence of childhood sexual abuse in any sample;
- any study which determined both the presence and absence of CSA and the subsequent presence and absence of our chosen outcomes;
- any review chapters or reports published in the last 10 years where the topic was CSA (books where the central topic was CSA were included regardless of year of publication);
- methodological papers to assist with the interpretation of the results; and
- meta-analyses of original research results.

The following exclusion criteria were then applied before collection of the articles:

- prevalence studies with total sample sizes of less than 100 (unless data was from an underrepresented country); and
- studies where the population was sampled on the basis of the presence of one of our chosen outcomes.

The first exclusion criterion was applied on the basis of the recommendation of Fergusson and Mullen (1999) and because samples of less than 100 may not provide accurate prevalence estimates for CSA. The second exclusion criterion was applied because studies that sampled on the basis of the presence or absence of an outcome could not be used to calculate a relative risk for that outcome (Streiner 1998). In epidemiological terms this means we could not determine the number of CSA-exposed individuals who developed the outcome of interest. Instead, we determined the number of people with outcome A who were exposed to CSA, thus answering a different question. Case–control studies were therefore only included where "cases" were individuals exposed to CSA and "controls" were those who were not exposed.

2.3 Search strategy for identification of studies

Several strategies were used to locate studies for this chapter. First, computer searches of 16 databases were conducted. Databases searched were the following: Medline; Embase; Psychinfo; E-psych; Healthstar; Cinahl; Cochrane; Social Work Abstracts; Health & Society; Family & Society; General Science Abstracts; Cambridge Life Sciences; Family Studies; Dissertation Abstracts; Child Abuse; Child Welfare & Adoption; Social Sciences Citation Index.

Subject heading (SH) and key word (KW) searches were carried out in two stages and were defined as follows:

Stage 1

SH: Child Sexual Abuse

OR

KW: "child*" (child, childhood, children) *AND* KW: ["sexual abuse", "sexual assault", "molestation", "incest"]

AND

Stage 2 (Risk factor levels)

SH: Epidemiology *OR* KW: ["epidemiology", "prevalence", "incidence"]

Stage 2 (Risk factor–disease relationships)

SH: Depression, *OR* KW: "depress*" (depression, depressive)

SH: Anxiety Disorders, *OR* SH: [Panic Disorder, Agoraphobia, Obsessive Compulsive Disorder, post-traumatic stress disorder] *OR* KW: ["panic disorder", "agoraphobia", "obsessive compulsive disorder", "OCD", "post-traumatic stress disorder", "PTSD"]

SH: [Substance Related Disorders, Alcoholism] *OR* KW: ["alcohol abuse", "alcohol dependence", "alcohol use", "substance abuse", "substance dependence", "substance use", "drug abuse", "drug dependence", "drug use"]

SH: [Suicide, Suicide Attempted], *OR* KW: ["suicid*" (suicide, suicidal)]

These searches generated a list of over 12 000 studies, of which approximately 4000 were duplicates. The abstracts of the remaining 8000 were examined to isolate potentially appropriate studies. The inclusion criteria were applied and the number of potentially relevant articles was reduced to approximately 1000. Many of the articles excluded at this step were those that either did not measure CSA separately from other types of abuse, or did not include a non-exposed control group.

After the exclusion criteria were applied to the 1000 articles, an initial sample of 460 articles was identified for collection. As the articles were reviewed the reference lists were examined in an attempt to uncover additional studies. The tables of contents of *Child Abuse and Neglect*, the leading journal in the area of child sexual abuse, were searched to locate any articles missed in searches. Experts in the area were contacted for information on unpublished data or for data from countries underrepresented in the usual databases and journals. These experts were located through the membership directory of the International Society for Prevention of Child Abuse and Neglect (ISPCAN), an organization sponsored by UNICEF-NY, Child Protection Division. The Society has over 300 members. All members were contacted via email or fax, with the exception of those from countries where we had adequate published data. A request for data was also put out on a "child maltreatment researchers" list on the Internet which yielded further information. The researchers had access to translation services so articles were not excluded on the basis of language.

The final database consisted of 604 articles. Of these, a further 91 were excluded after collection on the basis of the exclusion criteria defined above (CSA not reported separately from other types of abuse or from abuse as an adult $n = 33$; subjects sampled on the basis of outcome $n = 33$; no non-exposed control group $n = 11$; sample size too small $n = 11$; individual case study data only $n = 3$). In these cases it was not possible to determine from the abstract alone whether the study

should be included. This produced a final set of 513 articles or reports. They were as follows:

- 103 reviews or meta-analyses;
- 55 methodological papers;
- 179 coded studies;
- 48 studies not coded as data had been collected from another paper or report;
- 52 prevalence studies not coded as data on CSA were derived from a secondary source such as official reports or records;
- 13 studies not coded as subjects were gathered from a special sample;
- 48 studies not coded as the outcome data were either not adequately measured or outcome was measured using a continuous as opposed to a categorical measure; and
- 15 studies not coded, as we were unable to obtain copies of the relevant papers or reports.

The 179 coded studies were coded for risk factor levels, risk factor–disease relationships or both.

2.4 Methods for obtaining estimates where more than one data source exists

More than one estimate was available for each subregion. Therefore a meta-analysis was conducted to combine estimates. However, before this was carried out the prevalence estimates were adjusted to remove any differences that can be explained by variations in methodology employed in the studies.

Coding of studies

Prevalence of CSA was obtained from each study included in this analysis. Where possible, an overall prevalence of any CSA was coded as well as a breakdown into the mutually exclusive groups of non-contact, contact and intercourse. A number of methodological characteristics were also coded from each study. These included the type of sample, the representativeness of the sample, age and sex distributions, the method of data collection employed, how childhood was defined, survey response rate, how many questions were used to elicit the presence of CSA and whether restrictions were placed on the definition of abuse.

Correcting for methodological variability within studies

Previous studies have noted that the prevalence of CSA differs depending on a number of methodological factors, including method of data collection, number of questions used to assess CSA, definition of child-

hood and the type of sample assessed (Bolen and Scannapieco 1999; Gorey and Leslie 1997; Haugaard and Emery 1989; Wynkoop et al. 1995). In order to examine the methodological characteristics that influence the variability in prevalence, multiple linear regression analyses were conducted. As it was assumed that the impact of these factors may differ for males and females, separate regression models were fitted for all estimates of prevalence for males ($N = 93$ estimates) and all estimates of prevalence for females ($N = 143$ estimates). Too few data points existed to carry out the regression analyses in each of the levels of exposure. Thus, the independent variable was prevalence of any CSA (coded as either broad or narrow definition).

Model-building steps

1. Each independent variable was screened for adequate cell sizes as well as its relationship with all other independent variables. Cells were collapsed where necessary and choices were made between two variables when they were significantly related (Table 23.1 displays the variables chosen for the regression analysis and their categories).
2. The distribution of the dependent variable was tested for skewness, kurtosis and univariate outliers. Three outliers were removed from the analysis.
3. All variables were tested univariately and then entered together into a single model. A process of backward elimination was employed to remove the least significant variable at each step until all variables left in the final model were statistically significant at the 0.05 level. As shown in Table 23.1 two of the dependent variables entered into the model were "subregion" and "CSA definition". Variability in prevalence across different World Health Organization (WHO) regions and across the broad and narrow definitions of CSA was considered to be important. Thus, these variables were left in the multivariate model even if they were non-significant.

Results

The final model examining the multivariate effects of methodological factors accounted for 27% of the variance in prevalence of CSA in males and 22% of the variance in prevalence of CSA in females (beta weights and significance levels for the statistically significant variables in the multivariate model are shown in Table 23.1). The variables remaining in the "males" model were "subregion" ($F = 0.68$, $df = 3$, $P = 0.5688$), "CSA definition" ($F = 1.98$, $df = 1$, $P = 0.1637$) and "sample type" ($F = 6.13$, $df = 4$, $P = 0.0002$). These results indicate that studies of male college samples and general practitioner (GP) attendees report a significantly higher overall prevalence of CSA than studies of community samples (12% and 12% vs 6%). The variables remaining in the "females" model

Table 23.1 Characteristics of the dependent variables used in the regression analysis

		Variables left in final model for males		*Variables left in final model for females*	
Dependent variable	*Categories (for categorical variables)*	*Beta weights*	*P-value*	*Beta weights*	*P-value*
Subregion	1 = AMR-A	0.00	—	0.00	—
	2 = EUR-A	–1.16	0.4616	–8.18	0.0002*
	3 = WPR-A	0.95	0.6499	1.57	0.5897
	4 = All other subregions	3.63	0.2988	5.75	0.2553
CSA definition	1 = Broad (non-contact, contact or intercourse)	0.0	—	0.00	—
	2 = Narrow (contact or intercourse)	–1.70	0.1637	–4.60	0.0247*
Sample type	1 = Community volunteers	0.00	—	Null	Null
	2 = College students	4.52	0.0126*		
	3 = School students	–1.54	0.1982		
	4 = GP attendees	5.77	0.0145*		
	5 = All other samples	0.69	0.7509		
Number of questions	1 = One	Null	Null	–7.72	0.0003*
	2 = More than one			0.00	—
Definition of childhood	1 = <18 years old	Null	Null	Null	Null
	2 = <16 years old				
	3 = <14 years old				
	4 = Not reported				
Restrictions on CSA definition	1 = None	Null	Null	Null	Null
	2 = Coercion only				
	3 = Age only				
	4 = Age or coercion				
Sample size	Continuous variable	Null	Null	Null	Null
Response rate	Continuous variable	Null	Null	Null	Null
Current age	Continuous variable	Null	Null	Null	Null

Null Non-significant.

— Reference category.

* Significant at the $P < 0.05$ level.

were "subregion" ($F = 7.18$, $df = 3$, $P = 0.0002$), "CSA definition" ($F = 5.17$, $df = 1$, $P = 0.0247$) and "Number of questions" ($F = 13.65$, $df = 1$, $P = 0.0003$). These results indicate reported prevalence is lower among women in some European countries compared to North America (15% vs 22%), among studies that employed a narrow definition of CSA

(19%) than those that employed a broad CSA definition (23%), and those which used only one question to assess the presence of CSA (14%) than those with more than one question (23%).

The finding that CSA definition independently contributed to variance in prevalence for females but not males is unusual. The difference between the broad and narrow prevalence estimates is wholly explained by whether or not non-contact forms of abuse are included in the definition. Therefore, if non-contact abuse is rare as the only form of abuse experienced in males then it is likely that including non-contact abuse in a definition of CSA in males would not substantially alter the overall prevalence. However, the pattern across types of abuse for males does not indicate this to be the case (40% of all CSA in males is non-contact CSA). A more plausible explanation is that the low overall prevalence in males coupled with the smaller number of estimates and their substantial variance contributed to the lack of effect observed for different CSA definitions in males. One might expect that had more estimates been available in males a significant difference may have emerged.

The finding that studies that included more than one question about abuse yielded higher prevalence rates is widely supported in the CSA literature (Fergusson and Mullen 1999; Mullen et al. 2000; Peters et al. 1986; Plunkett and Oates 1990); it is also reported that the higher rates observed in such studies are likely to be more accurate (Bolen and Scannapieco 1999). Within this context, adjusting the prevalence estimates from studies that used only one question to more closely reflect those that used multiple questions (rather than the other way around) is likely to be a more accurate reflection of true prevalence in the population. Once again, the lack of effect in male estimates may be a function of the smaller number of estimates, especially given that the number of questions did significantly predict prevalence in males in the univariate analyses.

More intriguing is the finding that the prevalence in females was lower in EUR-A countries than in North America, from where the majority of the world's prevalence estimates come. Finkelhor (1994) reported on the international epidemiology of CSA and concluded that between-country differences were more likely to be due to methodological differences than reflective of true differences in prevalence. However, the fact that the effect in the present analysis remains after controlling for methodological factors may reflect true differences between these cultural groups. Moreover, in light of the larger number of studies available for the present analysis, it does appear to be true that international differences in prevalence exist. Within this context, it is important to note that there were substantial variations in CSA prevalence within subregions even when controlling for methodological variations, and this is reflected in the modest proportion of explained variance for the final model (22%). Therefore, the lower observed prevalence may, in fact, be a function of a non-examined explanatory variable. Moreover, a substantial number

of subregions were either underrepresented or not represented at all in this analysis, and it is therefore difficult to interpret these findings in a broader international context.

Weighting of prevalence estimates

The results of the above regression analyses were used to weight the overall CSA prevalence estimates for both males and females. Specifically, the unstandardized regression coefficients from the two final models were used to adjust the raw prevalence estimates. This was achieved by subtracting the coefficients from the prevalence for all levels of a variable that were statistically significant (e.g. if the prevalence in males was derived from a college sample the adjusted prevalence would be the raw prevalence minus 4.52 percentage points). The variables of "subregion" and "CSA definition" did not contribute to this adjustment because they do not reflect differences in methodological quality. This produced the desired effect of controlling for any influence these factors had on other variables while at the same time keeping any variance in prevalence due to these factors.

It is worth noting that implicit in this process of adjustment was the assumption that estimates from the group used as the reference group more closely reflected the true population prevalence. Thus, estimates for males from college and GP samples were adjusted to more closely reflect the prevalence observed in community samples. For females, prevalence estimates were adjusted to more closely reflect those observed in studies where more than one question to define CSA was asked.

Extrapolation across levels of exposure

Once weighted prevalence estimates were obtained for each study they were divided into the mutually exclusive groups of non-contact, contact and intercourse using the following method. The relationship between each level of exposure, expressed as a proportion of the overall prevalence, was calculated for all studies that reported prevalence of CSA by type of exposure. When a study reported prevalence based on a broad definition, this was apportioned into non-contact CSA, contact CSA and intercourse (eight studies for males, 12 studies for females). When a study reported prevalence based on a narrow definition this was apportioned into contact CSA and intercourse (12 studies for males, 20 studies for females). These proportions were then applied to those studies only reporting an overall prevalence to obtain an estimate of prevalence for each level of exposure in each study. A small number of studies reported proportions in the three levels of exposure that were opposite in direction to all other studies, and these were excluded from the present calculations. Sensitivity analyses were carried out to assess the impact of these two apportioning fractions (e.g. including all studies, excluding studies where the proportion was opposite to all other studies) on the prevalence in each of the three levels of exposure. The results of these

Table 23.2 The impact of different apportioning fractions on the prevalence of CSA in the three levels of exposure

	Method 1		*Method 2*	
	Apportioning fraction	*Prevalence*	*Apportioning fraction*	*Prevalence*
Males				
Non-contact CSA	0.297	2.6	0.387	3.1
Contact CSA	0.448	4.0	0.378	3.7
Intercourse	0.255	2.0	0.235	1.9
Females				
Non-contact CSA	0.279	6.7	0.291	6.8
Contact CSA	0.500	12.7	0.512	13.2
Intercourse	0.221	5.8	0.197	5.3

Method 1: Proportions derived from all studies (n = 12 for males, n = 20 for females).

Method 2: Proportions derived from all studies (n = 10 for males, n = 17 for females) except those in opposite direction to expected proportions (this method was used in the final calculations).

analyses are shown in Table 23.2. The overall impact of these excluded studies on each level of exposure was minimal. They were therefore not included in the calculation of this apportioning fraction.

The observed relationship between non-contact, contact and intercourse for both males and females indicated that intercourse was the least common form of abuse, a finding that has been widely reported (Fergusson and Mullen 1999; Mullen et al. 2000).

Extrapolation of combined estimates across age

Exposure to CSA, by definition, occurs in childhood. Thus, during childhood the prevalence of CSA represents current exposure, and reflects cumulative exposure to CSA from birth until current age. The prevalence of CSA in persons aged ≥18 years necessarily represents past exposure and reflects cumulative exposure from birth until the age of 18 years. Therefore prevalence of CSA will vary across age groups from 0–17 years dependent on both the age at which exposure to CSA usually begins (age at onset) and the duration of CSA throughout childhood. Because exposure is fixed at the age of 18 years, the prevalence of CSA across age groups after 18 years will vary only if the prevalence of CSA is changing over time. For example, if birth cohorts between 1956 and 1970 (current age 30–44 years) were more likely to be exposed to CSA in childhood than birth cohorts before 1956 (current age ≥45 years) then the prevalence of CSA for these age groups would differ.

The age groups required for reporting by WHO therefore necessitated three separate steps for calculation of exposure across age groups. The first step estimated prevalence in the 0–4-, 5–14-, and 15–17-year age groups. The second step estimated prevalence in the 18–29-, 30–44-, 45–59-, 60–69-, 70–79- and ≥80-year age groups. The third step esti-

mated a combined prevalence in the 15–17- and 18–29-year age groups to obtain prevalence for the age group 15–29 years.

Step 1: Estimating prevalence in persons aged 0–17 years

No data exist on the prevalence of CSA in different childhood age categories. Age is usually examined in terms of the age at first onset of the abuse or the duration of abuse. In order to estimate prevalence of CSA across different childhood age categories, all studies that contained information about onset or duration were examined. The main difficulty arose from the use of disparate age categories that did not always conform to those required for the WHO estimates. The following steps were carried out to examine this issue.

Age at onset of abuse. A total of 22 studies presented data on age at onset of abuse. Although categorization of age varied considerably across studies, onset was consistently more prevalent in the 5–14-year age group. This pattern was the same for both males and females. It is noted that the validity of self-reported abuse with an onset before the age of 5 years should be considered, at best, speculative. Age at onset for different levels of exposure and between different subregions could not be examined with the available data. Therefore, the following procedures were applied equally to males and females, across all levels of exposure and across all subregions.

- Two studies presented data in the 0–4-year age group. These studies indicate that approximately 6.5% of abuse begins in this age group. If the age band is extended to include 0–5 year olds (six studies in total) then the equivalent estimate is 9.7%. The midpoint of these two values indicates that approximately 8.1 % of all cases of CSA have their onset before the age of 5 years.
- Three studies presented data for those aged ≥15 years. These studies show that approximately 19.7% of abuse begins in this age group. If the age band is restricted to those aged ≥16 years (six studies in total) then the equivalent estimate is 18.8%. The midpoint of these two values indicates that approximately 19.3% of all cases of CSA have their onset after the age of 14 years.
- The middle age group was derived from the above calculations to give a value of 72.6% (100% – 8.1% – 19.3%), which is consistent with the overall pattern observed.

Duration of abuse. If prevalence of CSA in different childhood age groups is based only on the age at onset, then abuse that begins in one age group and continues into the next will not be counted. One way to account for this is to adjust for duration, that is, to include a proportion of cases with long duration in more than one age category.

Finkelhor (1979) estimated that 16% of those who experience CSA experience it on more than one occasion and for a duration of more than one week. Bagley and Mallick (2000) estimated this figure at approximately 20%. However, of most interest to the current analysis were the studies that reported the number of CSA cases where duration was at least one year. Two studies with such estimates were found (Collings 1997; Risin and Koss 1987), and they placed the prevalence of CSA cases with a duration of more than one year at 6.3% and 12% of all CSA cases. The midpoint of these values indicated that approximately 9.2% or one in ten cases of CSA would continue for more than one year (Risin and Koss 1987). Finkelhor (1979) further examined duration by level of exposure and indicated that non-contact abuse was the least likely to continue for more than one year while intercourse was the most likely (4.3% for non-contact, 5.8% for contact and 8.8% for intercourse, representing respective proportions of 0.68, 0.92 and 1.40, compared to the overall figure of 6.3%).

Combining onset and duration. Weights derived from age at abuse (0.081, 0.726 and 0.193) were combined with information about duration in the following way: 9.2% of cases in the 0–4-year age group were carried over to the 5–14-year age group using the above proportions for each level of abuse (0.68, 0.92 and 1.40). The same calculations were made in carrying over cases from the 5–14-year age group to the 15–18-year age group.

Step 2: Estimating prevalence in persons aged >18 years

For the reasons outlined earlier, prevalence of CSA in adults will only vary if the prevalence of CSA is changing over time. Several reviews have attempted to examine cohort effects in order to address this issue (Bagley 1990, 1995; Bagley and Ramsay 1985; Bickerton et al. 1991; Feldman et al. 1991; Fergusson et al. 2000). Three of these reviews concluded that the prevalence of CSA could be increasing over time while three reviews also concluded that there is no evidence to support a change in prevalence over time. This pattern is not explained by the publication dates of these reviews and many of the authors also pointed out that it is difficult to interpret these results without reference to a potential reporting phenomenon. That is, women in older age groups may be less willing or less able to report experiences of sexual abuse in childhood.

This issue was also examined empirically in the current data set. Each estimate for males and females was assigned a year of birth, calculated by subtracting the mean age of the sample from the year of publication (or where available the year the survey was conducted). This variable was then examined in a linear regression to determine whether year of birth explained any of the variance in prevalence. The continuous year of birth variable was also converted to a categorical variable with birth cohort defined according to the current age categories provided by

WHO. Birth cohort did not explain any variance in prevalence estimates for either males or females.

Given both these findings indicated no clear trend in prevalence of CSA over time, it was decided that estimates of prevalence would be combined across adult age groups. This had the effect of providing more stable estimates, particularly from underrepresented subregions.

Step 3: Combining estimates for 15–17 and 18–29 year olds

In order to combine estimates of prevalence in these two groups, it was necessary to determine the proportion of the population that fell into these two age groups for each subregion. This was calculated for each subregion represented in the data and was done separately for males and females. The data were obtained from the estimates provided by WHO and were based on population figures for the year 2000. As these data were only available for 15–19 year olds, as opposed to 15–17 year olds, population proportions were calculated using this age group.

Combining estimates within countries

Once estimates of prevalence were apportioned into the three levels of exposure and across the eight separate age groups (yielding 24 separate prevalence estimates for each study), prevalences were combined within countries. The prevalence estimates were combined using meta-analysis with STATA Intercooled 7. For ease of calculation the "meta" macro was used (Sharp and Sterne 1997). Heterogeneity between studies within each country was tested using the chi-squared statistic. When only two or three studies were available for combination, the between-study variance was estimated with poor precision (Cooper and Hedges 1994). Countries with less than five estimates were combined using a fixed-effects model, and countries with five or more estimates, and statistical heterogeneity, were combined using a random-effects model.

Combining estimates within subregions

In order to combine prevalence estimates between countries within each subregion, each country estimate was weighted for the population of that country. This meant that prevalence estimates from countries with large populations were given more weight in the final estimates. A combined estimate was obtained for each subregion by calculating a mean for each level of exposure in each age group for males and females.

2.5 Methods for obtaining estimates where no data existed

Extrapolation across subregions was one of the most difficult issues encountered in the construction of the prevalence estimates and arguably represents one of the greatest threats to the validity of the estimates in subregions where no data were available. For two out of the 14 subregions no prevalence data were found (EMR-B and EMR-D). These sub-

regions represent vastly different cultural, socioeconomic and geographic populations. In the absence of data it is impossible to speculate on how these differences might have impacted on the prevalence of CSA. No estimates were obtained for any countries in the Middle East. Therefore the estimate that was used comes from Turkey (EUR-B), which was considered the most appropriate in the absence of any other data. This extrapolation should be considered conjectural and the resultant estimates of prevalence in these subregions should be quoted with caution. It should also be noted that there was an uneven distribution of estimates in the remaining 12 subregions with a small number of countries making up a large proportion of the total number of estimates.

2.6 Description of studies, including methodological qualities

Prevalence studies included in the analysis are presented in Table 23.3. They are presented in three levels according to the type of sample used and the representativeness of this sample. The levels are defined as follows.

Level A: Representative community samples—samples of adolescents or adults where the article explicitly stated that the sample was representative of the population from which it was drawn. In general this was achieved through the use of complex sampling procedures or weighting.

Level B: Other community samples (representativeness not known)—samples of adolescents or adults where it was not known whether the samples were truly representative of the population from which they were drawn.

Level C: Community subgroups and convenience samples—samples of adolescents or adults drawn from a subgroup within the community based on factors such as ethnicity, educational or socioeconomic status.

2.7 Characteristics of excluded studies

There were 48 studies that were not coded due to their outcome measurement methods. Where the outcome was measured using a continuous measure for which there were established diagnostic cut-off points, authors were contacted and 2×2 tables were requested. The majority of authors responded and supplied data and these studies were included in the 179 coded studies. The 48 remaining were either not coded because authors could not supply data or authors were not contacted because the measure used could not be mapped to diagnostic criteria.

The characteristics of the 15 studies that could not be obtained are presented in Table 23.4. Inter-library loans were requested where items were only available interstate or overseas but several articles had not arrived at the time this chapter was being compiled.

Table 23.3 Characteristics of studies included in prevalence analysis

	Sample				*CSA data collection*		
Study authors	*Type*	*Number*	*% female*	*Mean age (years)*	*Method*	*No. of questions*	*Childhood definition*
AFR-D (Cameroon $n = 1$)							
Level C: Community subgroups and convenience samples							
Menick and Ngoh (1998)	School students	1 688	54	12	Self-report	>1	<15
AFR-E (Ethiopia $n = 1$, South Africa $n = 3$)							
Level B: Other community samples (representativeness not known)							
Mulugeta et al. (1998)	School students	719	100	16	Self-report	>1	<18
Level C: Community subgroups and convenience samples							
Collings (1997)	College students	640	100	20	Self-report	>1	<18
Collings (1991)	College students	284	0	20	Postal SR	>1	<18
Madu and Peltzer (2000)	School students	414	52	19	Self-report	>1	<18
AMR-A (Canada $n = 7$, USA $n = 72$)							
Level A: Representative community samples							
Davidson et al. (1991)	Adults	2 985	54	42	Face-to-face	>1	<16
Vogeltanz et al. (1999)	Adults	733	100	—	Face-to-face	>1	<18
Molnar et al. (2001)	Adults	5 877	50	35	Self-report	>1	<18
Siegel et al. (1987)	Adults	3 132	53	42	Face-to-face	>1	<16
Finkelhor et al. (1990)	Adults	2 626	56	—	Telephone	>1	<18
MacMillan et al. (1997)	Adults	9 953	51	40	Face-to-face	>1	<18
Wyatt (1985)	Adults	248	100	27	Face-to-face	>1	<18
Wyatt et al. (1999)	Adults	338	100	30	Face-to-face	>1	<18
Saunders et al. (1999)	Adults	4 008	100	45	Telephone	>1	<18
Brown et al. (1999)	Adults	639	48	18	Face-to-face	1	<18
Finkelhor and Dziuba-Leatherman (1994)	Adolescents	2 072	48	13	Telephone	>1	<17
Boney-McCoy and Finkelhor (1996)	Adolescents	1 457	47	13	Telephone	>1	<16
Kilpatrick et al. (2000)	Adolescents	4 023	49	15	Telephone	>1	<17
Risin and Koss (1987)	College students	2 972	0	21	Self-report	>1	<14
Fromuth and Burkhart (1989)	College students	253	0	20	Self-report	>1	<17
Harrison et al. (1997)	School students	122 824	51	15	Self-report	>1	<18
Bagley et al. (1995)	School students	2 112	49	15	Self-report	>1	<18
Blum et al. (1988)	School students	36 283	49	15	Self-report	1	<18
American School Health Association (1989)	School students	3 490	50	13	Self-report	1	<18
Nelson et al. (1994)	School students	2 332	51	16	Self-report	>1	<18
Hernandez (1992)	School students	3 179	48.3	14	Self-report	>1	<15
Bensley et al. (1999)	School students	4 790	48	16	Self-report	1	<18

Restriction on CSA definition	Prevalence in males (%)					Prevalence in females (%)				
	Any broad	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*
No	9.6	—	—	—	—	21.3	—	—	—	—
Yes	—	—	—	—	—	—	—	—	—	5.2
Yes	—	—	—	—	—	—	34.8	—	29.0	5.8
No	28.9	9.2	19.7	4.3	4.9	—	—	—	—	—
Yes	—	56.0[a]	—	42.5	13.5	—	53.2	—	35.6	17.6
No	—	—	—	—	1.1[b]	—	—	—	—	1.1[b]
Yes	—	—	—	—	—	24.0	17.7	6.3	—	—
No	—	2.5	—	1.9	0.9	—	13.5	—	8.5	5.0
No	—	3.8	—	—	—	—	6.8	—	—	—
No	16.0	—	—	—	—	27.0	—	—	—	—
Yes	6.7	—	—	—	—	19.5	—	—	—	—
Yes	—	—	—	—	—	62.0[a]	46.0	16.0	—	—
Yes	—	—	—	—	—	—	34.0	—	—	—
No	—	—	—	—	—	—	—	—	—	8.5
Yes	—	3.4[b]	—	—	—	—	3.4[b]	—	—	—
No	5.9	—	—	—	0.0	15.3	—	—	—	1.3
No	3.1	—	—	—	—	9.7	—	—	—	—
Yes	—	8.0[b]	—	—	—	—	8.0[b]	—	—	—
No	7.3	4.7	2.5	2.5	2.2	—	—	—	—	—
Yes	15.0	—	—	—	—	—	—	—	—	—
Yes	—	4.3	—	—	—	—	11.5	—	—	—
No	—	9.8	—	—	—	—	23.6	—	—	—
Yes	—	2.0	—	—	—	—	14.0	—	—	—
Yes	—	—	—	—	6.2	—	—	—	—	18.5
No	8.1	—	—	—	—	33.1	—	—	—	—
Yes	—	10.0[b]	—	—	—	—	10.0[b]	—	—	—
Yes	—	5.0	—	—	—	—	23.8	—	—	—

continued

Table 23.3 Characteristics of studies included in prevalence analysis (*continued*)

	Sample				*CSA data collection*		
Study authors	*Type*	*Number*	*% female*	*Mean age (years)*	*Method*	*No. of questions*	*Childhood definition*
Level B: Other community samples (representativeness not known)							
L. George and I. Winfield-Laird, unpublished document, 1986	Adults	1157	100	41	Face-to-face	1	<16
Keckley Market Research, unpublished document, 1983	Adults	603	—	—	Telephone	1	<18
Murphy (1997)	Adults	818	51	—	Telephone	1	<18
Wolf (1992)	Adults	637	56	—	Telephone	1	<16
Essock-Vitale and McGuire (1985)	Adults	300	100	40	Face-to-face	1	<18
Saunders et al. (1992)	Adults	391	100	42	Face-to-face	1	<18
Bagley and Ramsay (1985)	Adults	377	100	40	Face-to-face	>1	<17
Bagley (1991)	Adults	750	100	23	Face-to-face	>1	<17
Kercher and McShane (1984)	Adults	1054	56	—	Postal SR	1	<18
Russell (1983)	Adults	930	100	—	Face-to-face	>1	<18
Bagley et al. (1994)	Adults	750	0	23	Self-report	>1	<17
Roosa et al. (1998)	Adults	2003	100	20	Self-report	>1	<18
Bagley (1995)	Adults	1833	56	—	Self-report	>1	<17
Watts and Ellis (1993)	School students	670	100	15	Self-report	>1	<18
Erickson and Rapkin (1991)	School students	1197	50	15	Self-report	>1	<18
Lodico et al. (1996)	School students	6224	48	16	Self-report	>1	<18
Level C: Community subgroups and convenience samples							
Kendler et al. (2000)	Adults twins	1411	100	40	Postal SR	>1	<16
Silverman et al. (1996)	Adults (long study)	375	50	21	Face-to-face	>1	<18
White and Strange (1993)	College students	131	100	20	Postal SR	>1	<17
Peters and Range (1995)	College students	266	51	—	Self-report	>1	<12
Thakkar et al. (2000)	College students	707	100	19	Self-report	>1	<15
Schaaf and McCanne (1998)	College students	238	100	19	Self-report	>1	<15
Finkelhor (1979)	College students	796	67	21	Self-report	>1	<17
deLahunta (1996)	College students	787	38	—	Postal SR	—	—
Arroyo (1997)	College students	221	100	25	Face-to-face	>1	<18
Briere and Runtz (1988)	College students	278	100	20	Self-report	>1	<15
Wellman (1993)	College students	824	80	20	Self-report	>1	<18
Edwards and Alexander (1992)	College students	103	100	23	Self-report	>1	<18
Fritz et al. (1981)	College students	952	57	—	Self-report	>1	<14

	Prevalence in males (%)					Prevalence in females (%)				
Restriction on CSA definition	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*
—	—	—	—	—	—	—	2.0	—	—	—
Yes	7.0	—	—	—	—	11.0	—	—	—	—
Yes	—	3.0	—	—	—	—	13.0	—	—	—
No	9.0	—	—	—	—	27.0	—	—	—	—
No	—	—	—	—	—	—	17.3	—	—	—
No	—	—	—	—	—	33.5	24.6	9.0	14.6	10.0
Yes	—	—	—	—	—	—	21.7	—	—	—
No	—	—	—	—	—	—	32.0	—	—	—
Yes	3.0	—	—	—	—	11.6	—	—	—	—
Yes	—	—	—	—	—	54.0	38.0	16.0	—	—
Yes	—	15.5	—	—	—	—	—	—	—	—
No	—	—	—	—	—	—	39.0	—	16.0	23.0
Yes	—	8.2	—	—	—	—	17.6	—	—	—
No	—	—	—	—	—	—	14.5	—	—	—
No	—	15.0[b]	—	—	—	—	15.0[b]	—	—	—
Yes	4.2	—	—	—	—	16.5	—	—	—	—
Yes	—	—	—	—	—	30.3	22.5	7.8	14.1	8.4
Yes	—	1.1[c]	—	—	—	—	12.3[c]	—	—	—
Yes	—	—	—	—	—	33.5	—	—	—	—
Yes	19.1	12.2	6.9	—	—	31.9	19.3	12.6	—	—
Yes	—	—	—	—	—	—	13.5	—	—	—
No	—	—	—	—	—	12.2	—	—	—	—
Yes	8.6	—	—	—	—	19.2	—	—	—	—
No	—	3.5[c]	—	—	—	—	10.8[c]	—	—	—
No	—	—	—	—	—	31.2	—	—	—	—
Yes	—	—	—	—	—	—	14.8	—	—	—
No	23.0	13.4	9.6	—	—	15.0	5.6	9.4	—	—
Yes	—	—	—	—	—	—	43.6	—	—	—
Yes	—	4.8	—	—	—	—	7.7	—	—	—

continued

Table 23.3 Characteristics of studies included in prevalence analysis (*continued*)

	Sample				*CSA data collection*		
Study authors	*Type*	*Number*	*% female*	*Mean age (years)*	*Method*	*No. of questions*	*Childhood definition*
Haugaard and Emery (1989)	College students	1089	61	19	Self-report	>1	<17
Duane et al. (1997)	College students	958	61	—	Self-report	>1	<13
Stepakoff (1998)	College students	393	100	20	Postal SR	>1	<17
Boudewyn and Liem (1995)	College students	438	61	25	Self-report	>1	<14
Bryant and Range (1997)	College students	486	74	24	Self-report	>1	<18
Bolstad and Zinbarg (1997)	College students	117	100	26	Self-report	>1	<15
Fromuth (1986)	College students	482	100	20	Self-report	>1	<16
Sedney and Brooks (1984)	College students	301	100	19	Self-report	—	—
Hibbard et al. (1988)	School students	712	50	14	Self-report	1	<18
Riggs et al. (1990)	School students	600	52	16	Postal SR	1	<18
Greenwood et al. (1990)	GP attendees	100	59	42	Face-to-face	>1	—
Walch and Broadhead (1992)	GP attendees	405	100	29	Self-report	—	<18
Kellogg and Hoffman (1995)	GP attendees	142	60	20	Self-report	1	<18
Gould et al. (1994)	GP attendees	292	71	48	Self-report	>1	<17
Felitti et al. (1998)	GP attendees	9508	54	56	Postal SR	>1	—
Kilpatrick (1986)	Misc. comm. groups	501	100	28	Self-report	>1	<15
DiVasto et al. (1984)	Misc. comm. groups	500	100	27	Self-report	1	<13
Moeller et al. (1993)	Clinic sample	668	100	34	Postal SR	—	<18
Bayatpour et al. (1992)	Clinic sample	352	100	15	Face-to-face	1	<18
Descamps et al. (2000)	Lesbian women	1925	100	35	Postal SR	—	—
Blum et al. (1992)	American Indian and Alaska Native youth	13454	51	15	Self-report	—	<18
Robin et al. (1997)	American Indians	375	58	37	Face-to-face	>1	<16
Greenwald and Leitenberg (1990)	Nurses	1500	100	—	Self-report	—	<16
Hall et al. (1993)	Low income women	203	100	27	Face-to-face	>1	<18
Zuravin and Fontanella (1999)	Low income women	513	100	30	Face-to-face	>1	<14
Wingood and DiClemente (1997)	African-American women	165	100	24	Face-to-face	1	<16
Romero et al. (1999)	Latina women	300	100	32	Face-to-face	>1	<18
AMR-B (Brazil $n = 2$, Costa Rica $n = 1$, Dominican Republic $n = 1$, El Salvador $n = 1$, Mexico $n = 1$)							
Level A: Representative community samples							
Ramos-Lira et al. (1998)	School students	61779	47	14	Self-report	>1	<18

	Prevalence in males (%)					Prevalence in females (%)				
Restriction on CSA definition	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*
Yes	5.0	—	—	—	—	11.9	—	—	—	—
No	—	4.0	—	—	—	—	4.3	—	—	—
Yes	—	—	—	—	—	—	14.9	—	10.1	4.8
Yes	—	16.2	—	—	—	—	23.8	—	—	—
No	28.2[b]	—	—	—	—	28.2[b]	—	—	—	—
Yes	—	—	—	—	—	31.6	—	—	—	—
Yes	—	—	—	—	—	22.0	—	—	—	—
—	—	—	—	—	—	—	16.9[c]	—	—	—
No	—	8.0[b,c]	—	—	—	—	8.0[b,c]	—	—	—
No	8.1[b]	—	—	—	—	8.1[b]	—	—	—	—
No	0.0	—	—	—	—	16.9	—	—	—	—
Yes	—	—	—	—	—	35.6	—	—	—	6.1
No	—	16.0	—	—	—	—	39.0	—	—	—
No	—	12.0	—	—	—	—	30.0	—	—	—
Yes	—	22.0[b]	—	15.1[b]	6.9[b]	—	22.0[b]	—	15.1[b]	6.9[b]
No	—	—	—	—	—	55.0	—	—	—	1.8
No	—	—	—	—	—	—	—	—	—	1.8
Yes	—	—	—	—	—	19.8	—	—	—	—
No	—	—	—	—	—	14.8	—	—	—	—
—	—	—	—	—	—	—	28.7	—	—	—
Yes	—	10.0	—	—	—	—	21.6	—	—	—
Yes	14.0	—	—	—	—	49.0	—	—	—	—
Yes	—	—	—	—	—	3.6	—	—	—	0.7
No	—	—	—	—	—	—	22.0	—	—	—
No	—	—	—	—	—	—	20.7	—	8.0	12.7
No	—	—	—	—	—	—	—	—	—	12.7
No	—	—	—	—	—	33.0	—	—	—	8.6
Yes	—	4.3	—	—	—	—	4.3	—	—	—

continued

Table 23.3 Characteristics of studies included in prevalence analysis (*continued*)

	Sample				*CSA data collection*		
Study authors	*Type*	*Number*	*% female*	*Mean age (years)*	*Method*	*No. of questions*	*Childhood definition*
Level B: Other community samples (representativeness not known)							
WHO (2001)	Adults	1172	100	32	Face-to-face	>1	<15
WHO (2001)	Adults	1473	100	32	Face-to-face	>1	<15
Z.A. Ruiz et al., unpublished document, 1986	College students	893	54	23	Self-report	—	—
Level C: Community subgroups and convenience samples							
Barthauer and Leventhal (1999)	Adults	83	100	35	Face-to-face	>1	<18
Krugman et al. (1992)	College students	497	45	20	Self-report	>1	<19
AMR-D (Nicaragua $n = 1$, Peru $n = 2$)							
Level A: Representative community samples							
Olsson et al. (2000)	Adults	336	60	34	Self-report	4	<19
Level B: Other community samples (representativeness not known)							
WHO (2001)	Adults	1415	100	32	Face-to-face	>1	<15
WHO (2001)	Adults	1847	100	32	Face-to-face	>1	<15
EMR-B (no estimates available)							
EMR-D (no estimates available)							
EUR-A (Austria $n = 1$, Belgium $n = 1$, Czech Republic $n = 1$, Denmark $n = 1$, Finland $n = 1$, France $n = 3$, Germany $n = 3$, Greece $n = 1$, Ireland $n = 1$, Israel $n = 1$, Netherlands $n = 1$, Norway $n = 4$, Spain $n = 2$, Sweden $n = 4$, Switzerland $n = 2$, United Kingdom $n = 9$)							
Level A: Representative community samples							
Vandewege et al. (1988)	Adults	956	100	35	Face-to-face	—	—
Cawson et al. (2000)	Adults	2869	57	21	Computer	—	<13
Baker and Duncan (1985)	Adults	2019	52	40	Face-to-face	>1	<16
Edgardh and Ormstad (2000)	Adults	1943	58	17	Self-report	>1	<17
Ernst et al. (1993)	Adults	421	47	28	Face-to-face	1	<16
Halperin et al. (1996)	Adults	1116	51	15	Self-report	>1	<17
Spak et al. (1998)	Adults	316	100	—	Face-to-face	1	<18
Lopez et al. (1995)	Adults	1821	47	39	Face-to-face	>1	<17
Garnefski and Arends (1998)	Adults	13894	50	15	Self-report	>1	<19
Weiss and Zverina (1997)	Adults	1719	50	38	Postal SR	>1	<15
Bouhet et al. (1992)	Adults	1511	51	39	Postal SR	>1	<18
Hill et al. (2000)	Adults	862	100	31	Postal SR	>1	<16
Bendixen et al. (1994)	College students	996	51	23	Self-report	>1	—
Choquet et al. (1997)	School students	8140	51	16	Self-report	>1	<16
Sariola and Uutela (1994)	School students	6913	52	16	Self-report	>1	<15

	Prevalence in males (%)					Prevalence in females (%)				
Restriction on CSA definition	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*
—	—	—	—	—	—	7.8	—	—	—	—
—	—	—	—	—	—	5.8	—	—	—	—
—	33.0[b]	—	—	—	—	33.0[b]	—	—	—	—
No	—	—	—	—	—	17.0	—	—	—	9.6
No	—	12.8	—	—	—	—	32.2	—	—	—
Yes	20.0	—	—	—	—	26.0	—	—	—	—
—	—	—	—	—	—	19.5	—	—	—	—
—	—	—	—	—	—	7.9	—	—	—	—
—	—	—	—	—	—	19.0	—	—	—	—
Yes	—	11.0	—	—	—	—	21.0	—	—	—
No	8.0	—	—	—	0.7	12.0	—	—	—	0.8
Yes	3.1	—	—	—	—	11.2	—	—	—	—
Yes	1.8	—	—	—	—	4.9	—	—	—	—
No	10.9	3.3	7.7	2.2	1.1	33.8	20.4	13.4	14.8	5.6
No	—	—	—	—	—	9.8	—	—	—	—
—	14.0	12.0	2.0	8.2	3.8	27.0	19.7	7.3	12.7	7.0
Yes	—	2.2	—	—	—	—	8.2	—	—	—
No	—	4.6	—	4.3	0.3	—	8.4	—	7.5	0.9
No	4.6	3.1	1.4	—	—	7.8	5.2	2.6	—	—
No	—	—	—	—	—	—	17.5	—	11.9	5.6
Yes	3.5	—	—	—	—	19.4	—	—	—	—
Yes	—	—	—	—	0.9	—	—	—	—	0.7
Yes	3.3	—	—	—	—	7.6	—	—	—	—

continued

Table 23.3 Characteristics of studies included in prevalence analysis (*continued*)

	Sample				*CSA data collection*		
Study authors	*Type*	*Number*	*% female*	*Mean age (years)*	*Method*	*No. of questions*	*Childhood definition*
Level B: Other community samples (representativeness not known)							
Leth (2001)	Adults	1 235	54	—	Postal SR	—	<18
H. Holter, unpublished document, 1990	Adults	1 017	—	—	Postal SR	1	—
Rönström (1985)	Adults	938	—	—	Self-report	—	—
Schei (1990)	Adults	118	100	33	Face-to-face	1	—
J. Kinzl and W. Biebl, unpublished data	College students	1 125	—	—	Self-report	—	—
Agathonos et al. (1992)	College students	746	—	—	Self-report	>1	—
de Paul et al. (1995)	College students	403	74	21	Postal SR	>1	<13
Lazartigues et al. (1989)	College students	963	58	20	Postal SR	>1	<16
Schoetensack et al. (1992)	College students	1 841	48	21	Self-report	>1	—
Kelly et al. (1991)	College students	1 244	62	19	Self-report	>1	<18
Pederson and Aas (1995)	School students	465	54	19	Postal SR	>1	<13
Schein et al. (2000)	GP attendees	1 005	65	36	Self-report	>1	<18
Level C: Community subgroups and convenience samples							
Market Research Bureau of Ireland (1987)	Adults	500	—	—	Self-report	>1	<18
Richter-Appelt and Tiefensee (1996)	College students	1 068	42	24	Self-report	—	<12
Raupp and Eggers (1993)	College students	1 009	50	—	Self-report	—	<18
Bickerton et al. (1991)	GP attendees	1 232	100	—	Face-to-face	>1	—
Coxell et al. (1999)	GP attendees	2 474	0	46	Computer	>1	<16
Palmer et al. (1994)	GP attendees	115	0	32	Self-report	>1	<16
Risberg et al. (1999)	GP attendees	175	100	41	Postal SR	1	<18
Palmer et al. (1993)	GP attendees	120	100	30	Face-to-face	>1	<16
Brown and Harris (1993)	Low income women	404	100	—	Face-to-face	>1	<17
EUR-B (Turkey $n = 1$)							
Level C: Community subgroups and convenience samples							
Elal et al. (2000)	College students	1 597	62	—	Self-report	>1	<18
EUR-C (Russia $n = 1$)							
Level C: Community subgroups and convenience samples							
N. Lvoff and V. Lvoff, unpublished document, 1998	College students	723	50	—	Self-report	—	<18

	Prevalence in males (%)					Prevalence in females (%)				
Restriction on CSA definition	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*
Yes	7.0	6.5	0.4	2.2	4.3	14.0	12.1	2.0	5.2	6.9
No	—	9.0	—	—	—	—	19.0	—	—	—
—	3.0	—	—	—	—	11.0	—	—	—	—
No	—	—	—	—	—	14.0	—	—	—	—
—	19.0	—	—	—	—	36.0	—	—	—	—
No	6.0	—	—	—	—	16.0	—	—	—	—
No	3.9	—	—	—	—	6.4	—	—	—	—
Yes	3.4	—	—	—	—	10.2	—	—	—	—
Yes	5.8	3.9	1.9	3.0	0.9	16.1	10.8	5.3	8.9	1.9
No	27.0	11.0	16.0	—	—	59.0[a]	27.0	32.0	—	—
Yes	0.5	0.0	0.5	—	—	6.8	6.0	0.8	—	—
Yes	15.7	—	—	—	—	30.7	—	—	—	—
—	5.0	—	—	—	—	7.0	—	—	—	—
No	4.0	—	—	—	—	23.0	—	—	—	—
No	6.2	—	2.4	2.3	1.5	25.2	—	11.2	11.7	2.3
Yes	—	—	—	—	—	—	3.5	—	—	—
Yes	—	13.0	—	—	—	—	—	—	—	—
Yes	13.9	—	—	—	4.3	—	—	—	—	—
No	—	—	—	—	—	6.8	5.7	1.1	—	—
Yes	—	—	—	—	—	33.3	20.8	12.5	15.8	5.0
Yes	—	—	—	—	—	—	6.9	—	—	—
Yes	16.0	—	—	—	—	28.0	—	—	—	—
—	9.0	—	—	—	—	27.0	—	—	—	—

continued

Table 23.3 Characteristics of studies included in prevalence analysis (*continued*)

	Sample				*CSA data collection*		
Study authors	*Type*	*Number*	*% female*	*Mean age (years)*	*Method*	*No. of questions*	*Childhood definition*
SEAR-B (Indonesia $n = 1$, Sri Lanka $n = 1$, Thailand $n = 2$)							
Level B: Other community samples (representativeness not known)							
WHO (2001)	Adults	1536	100	32	Face-to-face	>1	<15
WHO (2001)	Adults	1282	100	32	Face-to-face	>1	<15
WHO (2001)	Adults	765	100	32	Face-to-face	>1	<15
Level C: Community subgroups and convenience samples							
Miles (2000)	School students	145	43	15	—	1	<18
SEAR-D (India $n = 3$)							
Level C: Community subgroups and convenience samples							
Interventions for support (2001)	School students	198	75	15	Self-report	—	<18
Interventions for support (2001)	School students	426	100	15	Self-report	—	<18
Castelino (1985)	College students	133	50	25	Self-report	—	<13
WPR-A (Australia $n = 8$, New Zealand $n = 4$)							
Level A: Representative community samples							
Anderson et al. (1993)	Adults	497	100	42	Face-to-face	>1	<16
Mullen et al. (1988)	Adults	314	100	—	Face-to-face	>1	<13
Goldman and Goldman (1988)	College students	991	61	22	Self-report	>1	<16
Mazza et al. (1996)	GP attendees	2181	100	—	Self-report	>1	<16
Level B: Other community samples (representativeness not known)							
Fleming (1997)	Adults	710	100	40	Postal SR	>1	<16
Martin et al. (1993)	Adults	1376	100	—	Postal SR	>1	<16
Level C: Community subgroups and convenience samples							
Dinwiddie et al. (2000)	Adults	5946	65	43	Telephone	1	<18
Nelson et al. (2002)	Adult twins	3982	58	30	Telephone	>1	<16
Fergusson et al. (1996a)	Adolescents	1019	51	18	Face-to-face	>1	<16
Higgins and McCabe (1994)	College students	253	79	21	Self-report	>1	<18
Goldman and Padayachi (1997)	College students	427	67	21	Self-report	>1	<17
Martin (1996)	School students	352	43	15	Self-report	1	<18
WPR-B (China $n = 1$, Malaysia $n = 1$)							
Level C: Community subgroups and convenience samples							
So-kum Tang (2000)	College students	2038	57	21	Self-report	>1	<17
Singh et al. (1996)	College students	616	77	22	Self-report	>1	<18

Postal SR Postal self-report.

— No data.

[a] Prevalence estimate not included in regression analysis based on outlier analysis (>3 standard deviations away from the mean).

	Prevalence in males (%)					Prevalence in females (%)				
Restriction on CSA definition	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*	*Any broad*	*Any narrow*	*Non-contact only*	*Contact only*	*Intercourse*
—	—	—	—	—	—	7.6	—	—	—	—
—	—	—	—	—	—	4.7	—	—	—	—
—	—	—	—	—	—	6.1	—	—	—	—
No	10.0[b]	—	—	—	—	10.0[b]	—	—	—	—
—	59.0	—	—	—	—	78.0	—	—	—	—
—	—	—	—	—	—	—	65.0	—	—	—
—	26.0	—	—	—	—	26.0	—	—	—	—
Yes	—	—	—	—	—	31.9	24.3	6.8	17.0	7.3
No	—	—	—	—	—	—	9.9	—	—	—
Yes	9.0	—	—	—	—	27.6	—	—	—	—
No	—	—	—	—	—	39.0	—	—	—	6.0
Yes	—	—	—	—	—	32.3	20.3	12.0	18.3	2.0
Yes	—	—	—	—	—	31.9	25.1	6.8	22.0	3.1
No	—	2.5	—	—	—	—	5.9	—	—	—
Yes	—	5.4	—	—	—	—	16.7	—	—	—
Yes	3.4	3.0	0.4	1.6	1.4	17.3	13.0	4.3	7.4	5.6
Yes	22.0	—	—	—	—	24.0	—	—	—	—
Yes	18.6	13.0	5.6	—	—	45.0	39.0	6.0	—	—
No	—	4.5	—	—	—	—	13.2	—	—	—
No	33.3[a]	—	—	—	3.0	28.2	—	—	—	5.8
No	2.1	2.1	0.0	2.1	0.0	8.8	5.5	2.3	5.1	0.4

[b] Estimates of prevalence not provided for males and females separately. Estimate for all persons displayed in the table. Regression modelling used estimates derived from methods outlined in section 3.

[c] Definition of CSA not provided, therefore assumed to be narrow.

Table 23.4 Characteristics of 15 excluded studies

Subregion	*Country*	*Author(s)*	*Type of study*	*Sample size*	*Sample type*
AMR-A	Canada	Berry (1997)	Prevalence	327	College students
AMR-A	Canada	Committee on Sexual Offences Against Children and Youths (1984)	Prevalence	2 008	Community adults
AMR-A	USA	Hernandez et al. (1993)	Prevalence and risk factor	2 973	School students
AMR-A	USA	Hibbard et al. (1990)	Prevalence and risk factor	3 998	School students
AMR-A	USA	Lenihan (1996)	Prevalence	1 687	College students
AMR-A	USA	Locke (1996)	Risk factor	—	College students
AMR-A	USA	Priest (1991)	Prevalence	—	College students
EUR-A	Austria	Friedrich et al. (1997)	Prevalence	—	Adult women
EUR-A	Italy	Meledandri et al. (1996)	Prevalence	—	—
EUR-B	Czech Republic	Pothe et al. (2000)	Prevalence	1 112	Community adults
WPR-A	Australia	Baldini (1996)	—	—	Aboriginal communities
WPR-A	Australia	Barton (1987)	Prevalence	>1 000	College students
WPR-B	China	Wang et al. (1994)	Prevalence	—	—
NA	NA	Garabedian (1994)	Risk factor	—	—
NA	NA	Garabedian (1994)	Risk factor	—	Adult women

NA Not applicable.

— No data.

2.8 Estimates by age, sex and subregion

Table 23.5 presents the final estimates of CSA prevalence by age, sex, level of exposure and subregion. It is presented in the format required for the WHO estimates. Table 23.6 presents the same data with level of abuse and age categories collapsed. This allows an easier comparison across subregions for both males and females. From Table 23.6 a number of interesting findings emerge. First, it can be seen that, on average, the prevalence of CSA is higher in females than in males. This is a commonly reported phenomenon. As demonstrated in the regression analysis, differences in prevalence also exist between subregions. Due to the paucity of prevalence estimates in subregions other than AMR-A, EUR-A and WPR-A it is not possible to look at differences between all subregions. However, the pattern of results does suggest that a high prevalence of CSA is found in AFR-E and SEAR-D. It should be noted that the

Table 23.5 CSA prevalence estimates (%) by subregion, sex, level of exposure and age group

			Age group (years)							
Subregion	*Sex*	*Level*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
AFR-D	Male	1	0.3	2.7	2.5	3.7	3.7	3.7	3.7	3.7
		2	0.3	2.6	2.5	3.6	3.6	3.6	3.6	3.6
		3	0.2	1.6	1.6	2.3	2.3	2.3	2.3	2.3
	Female	1	0.5	4.5	4.3	6.2	6.2	6.2	6.2	6.2
		2	0.9	7.9	7.5	10.9	10.9	10.9	10.9	10.9
		3	0.3	3.0	2.9	4.2	4.2	4.2	4.2	4.2
AFR-E	Male	1	1.3	12.1	11.8	16.6	16.6	16.6	16.6	16.6
		2	0.5	4.8	4.6	7.1	7.1	7.1	7.1	7.1
		3	0.5	4.2	4.2	5.9	5.9	5.9	5.9	5.9
	Female	1	0.5	4.5	4.3	6.2	6.2	6.2	6.2	6.2
		2	2.5	22.2	21.6	30.5	30.5	30.5	30.5	30.5
		3	0.5	4.3	4.1	6.0	6.0	6.0	6.0	6.0
AMR-A	Male	1	0.1	1.9	1.9	2.7	2.7	2.7	2.7	2.7
		2	0.2	1.8	1.8	2.5	2.5	2.5	2.5	2.5
		3	0.1	1.0	1.0	1.4	1.4	1.4	1.4	1.4
	Female	1	0.4	5.4	5.3	7.5	7.5	7.5	7.5	7.5
		2	0.9	9.7	9.7	13.6	13.6	13.6	13.6	13.6
		3	0.4	3.9	4.0	5.5	5.5	5.5	5.5	5.5
AMR-B	Male	1	0.5	4.3	4.2	5.9	5.9	5.9	5.9	5.9
		2	0.3	2.3	2.3	3.2	3.2	3.2	3.2	3.2
		3	0.1	1.2	1.2	1.6	1.6	1.6	1.6	1.6
	Female	1	0.2	1.8	1.8	2.4	2.4	2.4	2.4	2.4
		2	0.3	3.0	3.0	4.2	4.2	4.2	4.2	4.2
		3	0.1	1.3	1.3	1.7	1.7	1.7	1.7	1.7
AMR-D	Male	1	0.6	5.6	5.2	7.7	7.7	7.7	7.7	7.7
		2	0.6	5.5	5.1	7.6	7.6	7.6	7.6	7.6
		3	0.4	3.4	3.2	4.7	4.7	4.7	4.7	4.7
	Female	1	0.3	2.8	2.7	3.9	3.9	3.9	3.9	3.9
		2	0.5	4.9	4.8	6.8	6.8	6.8	6.8	6.8
		3	0.2	1.9	1.8	2.6	2.6	2.6	2.6	2.6
EMR-B	Male	1	0.4	3.2	3.2	4.4	4.4	4.4	4.4	4.4
		2	0.4	3.2	3.2	4.3	4.3	4.3	4.3	4.3
		3	0.2	2.0	2.0	2.7	2.7	2.7	2.7	2.7
	Female	1	0.7	5.9	5.9	8.2	8.2	8.2	8.2	8.2
		2	1.2	10.4	10.4	14.3	14.3	14.3	14.3	14.3
		3	0.4	4.0	4.0	5.5	5.5	5.5	5.5	5.5
EMR-D	Male	1	0.4	3.2	3.2	4.4	4.4	4.4	4.4	4.4
		2	0.4	3.2	3.2	4.3	4.3	4.3	4.3	4.3
		3	0.2	2.0	2.0	2.7	2.7	2.7	2.7	2.7
	Female	1	0.7	5.9	5.9	8.2	8.2	8.2	8.2	8.2
		2	1.2	10.4	10.4	14.3	14.3	14.3	14.3	14.3
		3	0.4	4.0	4.0	5.5	5.5	5.5	5.5	5.5

continued

Table 23.5 CSA prevalence estimates (%) by subregion, sex, level of exposure and age group (*continued*)

			Age group (years)							
Subregion	*Sex*	*Level*	*0–4*	*5–14*	*15–29*	*30–44*	*45–59*	*60–69*	*70–79*	*≥80*
EUR-A	Male	1	0.1	0.9	0.9	1.2	1.2	1.2	1.2	1.2
		2	0.1	1.2	1.2	1.6	1.6	1.6	1.6	1.6
		3	0.1	0.7	0.7	0.9	0.9	0.9	0.9	0.9
	Female	1	0.4	3.5	3.6	4.9	4.9	4.9	4.9	4.9
		2	0.7	6.0	6.2	8.3	8.3	8.3	8.3	8.3
		3	0.2	1.9	2.0	2.7	2.7	2.7	2.7	2.7
EUR-B	Male	1	0.4	3.2	3.2	4.4	4.4	4.4	4.4	4.4
		2	0.4	3.2	3.2	4.3	4.3	4.3	4.3	4.3
		3	0.2	2.0	2.0	2.7	2.7	2.7	2.7	2.7
	Female	1	0.7	5.9	5.9	8.2	8.2	8.2	8.2	8.2
		2	1.2	10.4	10.4	14.3	14.3	14.3	14.3	14.3
		3	0.4	4.0	4.0	5.5	5.5	5.5	5.5	5.5
EUR-C	Male	1	0.3	2.5	2.5	3.5	3.5	3.5	3.5	3.5
		2	0.3	2.5	2.4	3.4	3.4	3.4	3.4	3.4
		3	0.2	1.5	1.5	2.1	2.1	2.1	2.1	2.1
	Female	1	0.6	5.7	5.6	7.9	7.9	7.9	7.9	7.9
		2	1.1	10.0	9.9	13.8	13.8	13.8	13.8	13.8
		3	0.4	3.9	3.8	5.3	5.3	5.3	5.3	5.3
SEAR-B	Male	1	0.2	1.7	1.6	2.3	2.3	2.3	2.3	2.3
		2	0.2	1.6	1.6	2.3	2.3	2.3	2.3	2.3
		3	0.1	1.0	1.0	1.4	1.4	1.4	1.4	1.4
	Female	1	0.2	1.5	1.5	2.1	2.1	2.1	2.1	2.1
		2	0.3	2.6	2.6	3.6	3.6	3.6	3.6	3.6
		3	0.1	1.0	1.0	1.4	1.4	1.4	1.4	1.4
SEAR-D	Male	1	1.1	9.8	9.5	13.6	13.6	13.6	13.6	13.6
		2	1.0	9.5	9.3	13.3	13.3	13.3	13.3	13.3
		3	0.7	5.9	5.7	8.2	8.2	8.2	8.2	8.2
	Female	1	1.1	10.5	10.2	14.7	14.7	14.7	14.7	14.7
		2	3.0	27.7	26.8	39.1	39.1	39.1	39.1	39.1
		3	1.1	10.0	9.7	13.9	13.9	13.9	13.9	13.9
WPR-A	Male	1	0.2	1.4	1.4	1.9	1.9	1.9	1.9	1.9
		2	0.2	1.8	1.8	2.5	2.5	2.5	2.5	2.5
		3	0.1	0.9	0.9	1.4	1.4	1.4	1.4	1.4
	Female	1	0.8	6.3	6.4	8.6	8.6	8.6	8.6	8.6
		2	1.0	11.4	11.6	15.8	15.8	15.8	15.8	15.8
		3	0.3	3.3	3.4	4.6	4.6	4.6	4.6	4.6
WPR-B	Male	1	1.1	9.6	9.9	13.2	13.2	13.2	13.2	13.2
		2	1.0	9.2	9.5	12.7	12.7	12.7	12.7	12.7
		3	0.2	1.9	1.9	2.6	2.6	2.6	2.6	2.6
	Female	1	0.6	5.8	6.0	8.0	8.0	8.0	8.0	8.0
		2	1.1	10.3	10.6	14.1	14.1	14.1	14.1	14.1
		3	0.5	4.1	4.3	5.7	5.7	5.7	5.7	5.7

Table 23.6 CSA prevalence estimates (%) by subregion and sex

	Broad estimate	
Subregion	*Males*	*Females*
AFR-D	9.6	21.3
AFR-E	29.8	42.7
AMR-A	6.7	26.5
AMR-B	10.7	8.4
AMR-D	20.0	13.3
EMR-B[a]	11.5	28.0
EMR-D[a]	11.5	28.0
EUR-A	3.8	15.8
EUR-B	11.5	28.0
EUR-C	9.0	27.0
SEAR-B	6.0	7.1
SEAR-D	35.0	67.7
WPR-A	5.9	29.1
WPR-B	28.6	27.8

[a] Subregions for which no data were available. Estimates were used from EUR-C.

estimates from AFR-E and SEAR-D come from very few studies that were relatively poor methodologically. This makes the estimates for these subregions at best highly uncertain. More studies are obviously needed to confirm whether or not the prevalence is higher in these subregions of the world. Some of these subregional differences appear dependent on sex. For example, in AMR-B, AMR-D and WPR-B the studies reported a higher prevalence in males than females. Only a few studies contributed to the estimates in these subregions and without measures of variability around the estimates it was difficult to draw firm conclusions.

2.9 Quantitative and qualitative sources of uncertainty

Uncertainty in the current analysis came from several sources. Studies of the prevalence of CSA varied in terms of methodological characteristics. Regression analyses demonstrated that several methodological factors contributed to the variability in prevalence estimates. This was the major quantitative source of uncertainty and was reduced by adjusting the prevalence estimates to more closely reflect ideal methodology. Meta-analysis was used as a method of quantifying the uncertainty around the final prevalence estimates, taking into account sample size or variability between studies, whichever is appropriate given the homogeneity of the estimates being combined. Other sources of uncertainty arose from

decisions made regarding inclusion and exclusion criteria, and methods of extrapolation across age, sex and subregion. These decisions and their rationale have been documented in the relevant sections. It should also be noted that the estimates of uncertainty will necessarily be underestimated, as data do not exist for every country in all subregions.

3. Estimating risk factor–disease relationships

3.1 Outcomes to be assessed: including evidence for causality and reasons for exclusion of related outcomes

Outcomes included and excluded

The choice of outcomes to be assessed was guided by two principles. First, choice of outcomes was limited to those diseases or outcomes that were included in the Global Burden of Disease (GBD) study. Second, outcomes were limited to those for which there was sufficient evidence of a causal relationship with CSA, the exposure variable. Section 3.4 presents a detailed assessment of the evidence for causality between CSA and six mental disorders (depression, panic disorder, agoraphobia, PTSD, and alcohol and drug abuse or dependence) and suicide attempts.

Child sexual abuse has also been linked to other mental disorders, including OCD, eating disorders and personality disorders. Three studies examined the relationship between CSA and OCD (Arata 1999; Saunders et al. 1992; Stein et al. 1988), however the evidence was equivocal and further research is required to confirm any association or lack thereof. Eating disorders have long been conceptualized as a response to a dysfunctional family environment. While it was originally thought that CSA played a pivotal role in the development of eating disorders, recent community studies have reported only a modest association between CSA and subsequent development of eating disorders after controlling for confounding influences of the family and social environment (Mullen et al. 1996; Wonderlich et al. 1997). CSA has also been linked to personality disorders, with a couple of studies finding positive associations between CSA and antisocial personality disorder (Scott 1992) and borderline personality disorder (Johnson et al. 1999). These disorders are currently not included in the GBD project and therefore were not examined here.

CSA does not only produce an increased risk of mental disorder. There is anecdotal and epidemiological evidence that CSA increases the probability of negative psychological outcomes such as poor self-esteem (Romans et al. 1997), lack of a sense of control or agency (Mullen and Fleming 1998), difficulties with intimacy and continuing sexual difficul-

ties that demonstrate the far reaching damage that occurs in some individuals (Mullen et al. 1996). These outcomes were outside the scope of this analysis and several meta-analyses examining the relationship between CSA and aspects of psychological adjustment offer a more comprehensive review of these psychological correlates (Jumper 1995; Paolucci et al. 2001; Rind et al. 1998). But this does not infer that they are in any way less harmful or less important to the person than the mental disorders identified in this report. The impact of CSA on adult life has also been studied in terms of non-mental health consequences. These include the increased risk of developing sexually transmitted diseases, teenage pregnancies, multiple sexual partnerships and sexual revictimization (Gorcey et al. 1986; Nagy et al. 1995). It has been suggested that a history of CSA, particularly of the more intrusive types, interrupts the child's development of sexuality and normal sexual relationships (Fleming et al. 1999). However, these were also outside the scope of this analysis.

General assessment of causality

Below, Hill's (1965) criteria for causality were applied to the case of CSA and adult mental disorders. The remainder of this section outlines each of these criteria as they relate to research in the area of child sexual abuse. A more detailed assessment of causality for each outcome is included in section 3.4.

Temporality

To satisfy the requirement of temporality, exposure to child sexual abuse must occur prior to the onset of adult mental disorder. By definition, some mental disorders are neither present nor easy to diagnose in children; therefore the time of onset of these disorders relative to child sexual abuse is a moot point. However, it is generally acknowledged that pre-existing vulnerabilities or predisposition to adult mental disorder exist even in childhood (Caspi et al. 1996). Long-term prospective studies that follow children from a young age and control for these vulnerabilities therefore provide the best evidence regarding temporality. While a number of research projects have been prospective in design, the majority of studies of child sexual abuse have been cross-sectional and retrospective. Traditionally, cross-sectional studies are not considered sources of evidence for temporality. However, by definition, exposure to child sexual abuse occurs during childhood and therefore prior to onset of an adult psychiatric disorder. It is for this reason that cross-sectional studies that indicate a relationship between child sexual abuse and adult mental disorder can still be indicators of temporality.

It is important to point out that even in prospective studies, data on CSA are gathered retrospectively. It is unethical, and in many countries illegal, to prospectively identify CSA and not intervene. However, retrospective studies rely on recall of memories and can therefore suffer from

unreliable data. Moreover, reports of CSA are often ascertained contemporaneously with assessment of disorder, which leaves recall open to bias in which those with disorder are more prone to recall CSA (Mullen et al. 2000). The issue of potentially unreliable recall represents one of the central threats to the validity of the published literature on CSA. Unfortunately, given that very few cases of CSA are reported to other adults, and even fewer to authorities, the validity of retrospective of CSA is very difficult to establish. Fergusson and Mullen (1999) recommended that one way of approaching this issue was to question the same individuals on multiple occasions to examine the consistency of their reports. Although this issue has rarely been examined, evidence suggests moderate-to-good consistency of CSA reports over time (Fergusson and Mullen 1999). Moreover, evidence also indicates that unreliability commonly arises from false negative reports (Fergusson et al. 2000) rather than false positive reports. Presumably greater validity would also be achieved if the presence of disorder were not determined at the same time as reports of CSA were obtained. As almost none of the prospective studies of CSA separated the ascertainment of disorder and reports of CSA, it is difficult to comment on what effect, if any, this may have had on results published to date. In summary, given that retrospective reports of CSA are virtually the only measure of CSA available in the literature, they must be accepted within the context of the caveats stated.

Strength

Several prospective studies (Fergusson et al. 1996b; Silverman et al. 1996) and several large studies with representative samples (Molnar et al. 2001; Saunders et al. 1999; Stein et al. 1988; Wilsnack et al. 1997) have found an association between CSA and mental disorders. These studies have reported odds ratios (ORs) of between 1.1 for depression (Kendler et al. 2000) and 10.2 for PTSD (Molnar et al. 2001). Despite this variability, CSA has been found to have at least a moderate effect on the outcomes studied (see Table 23.7 for a summary of the evidence). The strength of the relationship between CSA and mental disorders or suicide attempts is generally reduced when the effects of mediating variables are taken into account. This is particularly evident for non-contact forms of abuse.

Elimination of other possible causes

Child sexual abuse often co-occurs within the context of other family dysfunction, social deprivation, emotional and physical abuse and other environmental stressors that are also associated with mental disorders (Fergusson and Mullen 1999). The interaction between these additional stressors, CSA and adult mental disorders is not likely to be simple or linear (Mullen et al. 1996, 2000; Rutter 1999). Furthermore, it has been argued that the apparent association between CSA and mental disorders can in fact be attributed to family dysfunction rather than to CSA (Rind

et al. 1998). It is therefore important to establish that the effect of CSA on adult functioning remains after controlling for some of these co-occurring factors. The following section contains a discussion of the mediating factors that are commonly reported in studies of the effects of CSA on adult functioning.

Sociodemographic characteristics. CSA is not evenly distributed across sex and socioeconomic groups (Mullen et al. 2000), factors that have been found to be independently associated with mental disorders in adulthood (Andrews et al. 2001; Kessler et al. 1994). However, after controlling for these factors, several studies have demonstrated an independent association between CSA and mental disorders (see Table 23.7). Additionally, a recent meta-analysis found that sex or socioeconomic status did not mediate the relationship between CSA and depressive symptoms, PTSD or suicide attempts (Paolucci et al. 2001).

Family environment. Family environment is one of the most commonly reported mediating factors in the CSA literature. Aspects of family environment studied are myriad but include parental functioning and relationships, domestic violence, parental separation during childhood, growing up away from parents, poor parental health—both physical and emotional—and parental drug and alcohol use (Conte and Schuerman 1987; Fergusson and Mullen 1999; Fromuth 1986; Jumper 1995; Kendall-Tackett et al. 1993; Martin 1996; Mullen et al. 2000; Neumann et al. 1996; Rind et al. 1998; Wyatt and Newcomb 1990). Although measures of adverse family environment vary substantially, it is generally considered to be one of the most important mediators of the effect of CSA on adult functioning (Chandler and Jackson 1997; Fergusson and Mullen 1999; Rind et al. 1998). Moreover, most studies have found that although the independent effect of CSA on adult functioning is substantially reduced once family environment is controlled for CSA, particularly abuse involving penetration is significantly and independently associated with negative outcome (Mullen et al. 2000).

Other abuse. Children who have experienced CSA are at considerably greater risk of experiencing other types of abuse such as physical abuse and neglect (Bifulco et al. 1991; Briere and Runtz 1990; Fergusson and Mullen 1999; Fergusson et al. 1996a; Hibbard et al. 1990; Mullen et al. 1996, 2000; Paradise et al. 1994). There is also some evidence that psychopathology increases with the number of abuse types experienced. In a retrospective study of adult women in New Zealand, the chances of being assigned a clinical diagnosis increased to 24% for a single type of abuse (sexual, physical or emotional), 41% for two types of abuse and to 60% for three types of abuse (Mullen et al. 1996). As such it is difficult to isolate the independent contribution of each of these types of abuse to adult psychopathology. Nonetheless, of the studies included in

the present report that controlled for other types of abuse the majority supported an independent effect of CSA on outcome (Molnar et al. 2001; Mullen et al. 1993, 1996; Yama et al. 1995; Zuravin and Fontanella 1999).

Temperament. The area of CSA deals specifically with human behaviour and therefore with substantial gene environment interaction (Kendler et al. 2000; Rutter 1999). Genetic factors act to enhance vulnerability to mental disorders in general, and may also act to enhance or reduce the risk of developing mental disorders following CSA. Twin studies provide one of the best ways to examine the interplay between genetic and environmental influences. There have been three studies examining the effects of CSA on mental disorders in twins (Dinwiddie et al. 2000; Kendler et al. 2000; Nelson et al. 2002). All three studies concluded that CSA is independently associated with most mental disorders and suicide attempts; two reported that the effect was found, even in twins discordant for CSA, when genetic vulnerabilities and many family factors were controlled for. Unfortunately small numbers of CSA-discordant twins in these analyses meant that, although these studies were included in the final estimates for this report, we could not include a separate genetic adjustment factor.

Protective factors. Much of the discussion around variables that mediate the effects of CSA on adult mental disorder focuses on negative or destructive environmental influences rather than protective ones. However, a recent prospective study examined external protective factors that can modify a child's psychiatric trajectory. Lynskey and Fergusson (1997) developed a regression model that demonstrated factors that protected against the development of psychiatric disorders. These were higher levels of paternal care during childhood and having fewer affiliations with delinquent or substance abusing peers. Once the model adjusted for both of these factors, severity of the sexual abuse (ranging from none, non-contact, contact through to intercourse) was not a significant predictor of outcome. Similarly, having a warm and supportive relationship with the non-offending parent and lower levels of abuse-related stress have been shown to predict resilience in sexually abused girls (Spaccarelli and Kim 1995). In essence, while the majority of research has shown that other negative factors contribute to the risk of developing adult mental disorder having experienced sexual abuse as a child, the converse is also likely to be true. That is, certain positive mediating variables are likely to reduce the risk of negative outcomes following CSA. This concept is often referred to as resilience (Fergusson and Mullen 1999). Unfortunately very few studies have measured protective factors in a systematic way and as such they could not be quantified for the present report.

Covariation or biologic gradient

For biological risk factors such as blood pressure levels this criterion is typically established by the presence of a dose–response relationship between risk factor and outcome. In the area of child sexual abuse, it is difficult to determine the presence of a dose–response relationship because it is difficult to define a "dose". A dose most closely relates to the severity of abuse to which an individual is subjected. The literature generally defines severity of abuse in five ways: type, frequency, duration, age of onset of abuse, and relationship of victim to offender. Regardless of how it is defined, there is broadly supportive evidence relating the severity of the abuse to the degree of psychiatric or psychological disturbance.

Type of abuse. Using this definition, severity of abuse is generally taken to express the spectrum that ranges from non-contact forms of sexual abuse (e.g. verbal sexual invitations, showing pornography), to contact forms of abuse (touching), through to intercourse. In those studies that have presented risk for disorder according to exposure to different types of abuse, risk for disorder increases as exposure to more severe types of abuse occurs (Baynard 1999; Fergusson et al. 1996b; Kendler et al. 2000; Saunders et al. 1992). In general the literature supports the notion that CSA involving contact or intercourse is associated with a more negative outcome in adulthood than non-contact CSA. It is also the most widely reported definition of the severity of exposure to CSA and is therefore used to obtain estimates for the current report.

Frequency of abuse. It has been demonstrated that experiencing one episode of child sexual abuse is often associated with further sexual victimization. In one study, for example, 16.8% of children were reabused in the 61–72 months prior to follow-up, with the greatest risk period occurring in the two years immediately after the initial abuse (Levy et al. 1995). In another study of 24 507 children with substantiated abuse/neglect who were monitored up to four years after the initial maltreatment incident, 9.3% of the children experienced abuse or neglect in the follow-up period (Fryer and Miyoshi 1994). For these children, the risk of reabuse continued to be greater than the risk of abuse in the general population and was greatest immediately following the first notified abuse/neglect incident. For example, 24% of abused/neglected children were revictimized in the first month following the index event. Bentovim et al. (1987) followed up families who were referred to a treatment programme for sexual abuse and found that 16% of children had experienced revictimization; in 15% of children it was unclear whether children had been reabused or not. In a sample of children and adolescents, Boney-McCoy and Finkelhor (1995) found that 39% of children who had histories of prior sexual victimization had been sexually abused in

the last year. Similar outcomes can occur in adults. Women with a history of CSA were significantly more likely to experience rape as an adult and to be victims of domestic violence (Fleming et al. 1999), which raises the question as to whether child sexual abuse is a vulnerability factor to further sexual abuse in and of itself.

Several studies have reported that not only is reabuse common, it is also associated with poorer outcome. In their meta-analysis, Kendall-Tackett et al. (1993) observed that poor psychological and behavioural outcomes in children were related to a variety of abuse-related variables, including greater abuse frequency. Increases in frequency of abuse have also been shown to be significantly associated with greater severity of psychological disorder, such as making more numerous suicide attempts (Bagley et al. 1995). While it is generally acknowledged that frequency of abuse is associated with more negative outcome, very few studies report outcome for varying abuse frequencies.

Duration of abuse. Duration of CSA has been shown to significantly affect psychological outcome, both in meta-analyses (Kendall-Tackett et al. 1993) and other research. Peters (1988) found that the greater the duration of the abuse, the more mental disorders and suicide attempts in adulthood. This is consistent with the notion that cumulative trauma has a more substantial effect than a single or less frequent abusive event. Once again, outcome of CSA is rarely reported for varying duration of abuse.

Age at onset of the abuse. At first glance, the evidence of the effect of age at abuse onset on mental disorder appears to be conflicting in terms of its direction. Peters (1988) found that women aged 18–36 years who had been older at the time they were sexually abused were diagnosed with more mental disorders or had made more suicide attempts. In contrast, Lynskey and Fergusson (1997) found a significant relationship between being younger at the time of the sexual abuse and increasing rates of mental disorders in a sample of 18 year olds. In both of these analyses, once adjusted for confounders, the relationship between age at abuse onset and mental disorders was no longer significant. Consistent with the conclusion of Browne and Finkelhor (1986) there appears to be no solid evidence for a relationship between age at onset and mental disorders after controlling for other aspects of the abuse and relationship variables.

Relationship of the offender to the child. A meta-analysis of the child sexual abuse research found no significant association between the relationship of the offender to the child, and mental disorders (Paolucci et al. 2001). This relationship is surprising given that intrafamilial abuse may occur over a longer period of time and with greater frequency than extrafamilial abuse (Fergusson and Mullen 1999). Browne and

Finkelhor (1986) suggest two reasons why the relationship of the child to the offender may not be a consistent predictor of negative outcomes. First, that lack of a consistent association may reflect variations in the degree of betrayal, rather than whether the victim and perpetrator are related. Second, while abuse by someone who is trusted may involve betrayal, abuse by a stranger may involve more fear and therefore be more aversive to the victim.

Consistency

The literature has consistently reported that psychiatric disorders are frequently found to be more common among those subjected to CSA compared to their non-abused peers. This finding persists across a range of populations that have included college students, community samples, school students, children, adolescents and adults, and cohorts in a number of different developed countries. Few studies, however, have been conducted in developing countries. Further research is required to confirm that the deleterious consequences associated with CSA reported in the literature so far also applies to the rest of the world and is not mediated by cultural and social factors.

Plausibility

It is acknowledged that psychiatric disorder arises from an interaction between adverse environmental influences and an individual's genetic make-up (Rutter 1999). Genetic influences aside, it is accepted that childhood adversity is a potent influence on psychiatric outcome (Brown and Moran 1994; Kessler and Magee 1993). This milieu of adversity has been described as a matrix of disadvantage (Mullen et al. 2000) and includes a variety of socioeconomic, familial and other environmental factors, as outlined earlier. Child sexual abuse falls at the more severe end of the spectrum of this adversity.

In terms of the effect that child sexual abuse has on the individual, it is logical that a child exposed to a traumatic event such as sexual assault may function less well psychologically and may develop phobic responses and anxiety-related symptoms, including PTSD (Green 1988). It has been proposed that the sexual abuse, regardless of type, involves four traumagenic dynamics (Finkelhor and Browne 1988). These are betrayal, powerlessness, traumatic sexualization and stigmatization. Synthesizing the child sexual abuse literature, Polusny and Follette (1995) placed the various outcomes associated with child sexual abuse in the context of emotional avoidance, suggesting that these outcomes are the result of maladaptive coping behaviour. Within this framework, a spectrum of avoidance, anxiety, despair and attempts to control becomes evident. When that fails, it produces anxiety disorders, alcohol and substance abuse, depression and other psychopathology, and suicide at the extreme. Within this context, despite the lack of a biological link between CSA and mental disorders, a causal relationship would certainly be plausible.

3.2 Description of studies including methodological qualities

Table 23.7 presents the characteristics of the studies that contributed to the risk factor–disease relationship grouped by psychiatric outcome. Also reported in Table 23.7 are the range of odds ratios and significance for each study.

3.3 Overview of methods

The criteria for identifying relevant studies and the characteristics of excluded studies have been reported on previously. All articles that met the inclusion criteria were coded against each outcome measured. The majority of studies measured more than one of our chosen outcomes. Where possible both unadjusted and adjusted measures of risk (RRs and ORs) were coded from articles. The majority of studies did not present estimates of risk adjusted for relevant confounders. Moreover, many presented proportions only. Regardless of which estimates of risk or association were quoted in articles, 2x2 tables were coded from every article included in the analysis. Where possible 2x2 tables were coded for each exposure level and where these data were not available 2x2 tables were coded for overall exposure. Where studies presented risks or proportions for both lifetime and current levels of outcome, both were coded. Many articles presented outcome as a continuous rather than a categorical variable. In these cases and where studies had used a measure of outcome that could be mapped to diagnostic criteria via a validated cut-off point, authors were contacted and asked to supply 2x2 tables. Unadjusted RRs and ORs and corresponding 95% confidence intervals were calculated from each 2x2 table using conventional formulae (Gardner and Altman 1989; Streiner 1998).

Extrapolations across sex, diagnostic time frame and levels of exposure

Given the requirements for data presentation (relative risk by age and sex for each level of exposure for each subregion) the biggest source of error in the data arose from having a small number of studies from which to derive estimates. The following sections detail the decisions that were made for each of these extrapolations. However, several general rules applied.

- Consider any theoretical implications of the extrapolation (e.g. is there any reason to expect relative risk will vary with age or sex or that confounding factors will differ for different mental disorder outcomes? Is there a plausible hypothesis or explanation?).
- Assume that no difference exists between the groups of interest (e.g. across sex, age or diagnostic time frame) unless there is clear evidence of a consistent pattern.

Table 23.7 Characteristics of studies included in the risk factor–disease relationship analysis

	Sample				Adjusted for confounders									
Level of evidence	*Type*	*N*	*% female*	*Age (years)*	*Demo-graphics*	*Family function*	*Other abuse*	*Other disorders*	*Outcome measure*[a]	*Diagnosis time frame*	*Childhood definition*	*CSA definition*[b]	*OR*	*P <0.05*
Depression														
Level 1: Twin studies														
Dinwiddie et al. (2000)	Adult twins	5946	65	43	Yes	No	No	No	SSAGA	Lifetime	<18	Narrow	2.2–3.9	Yes
Kendler et al. (2000) (10-year follow-up)	Adult twins	1411	100	40	No	Yes	No	No	SCID	Lifetime	<16	Broad	1.1–2.8	Yes[c]
Nelson et al. (2002)	Adult twins	33892	58	30	Yes	Yes	No	Yes	SSAGA	Lifetime	<16	Narrow	1.3–1.7	Yes[c]
Level 2: Prospective studies														
Brown and Harris (1993) (8-year follow-up)	Low income women	404	100	—	No	Yes	Yes	No	PSE	12 months	<17	Broad	—	Yes
Brown et al. (1999) (13-year follow-up)	Community	639	48	—	Yes	Yes	No	Yes	DISC	—	<18	Narrow	3.2	Yes
Ernst et al. (1993) (10-year follow-up)	Community	421	47	28	No	No	No	No	SPIKE interview	—	<16	Broad	—	No
Fergusson et al. (1996b) (Follow-up U/K)	Community	1019	51	18	Yes	Yes	No	Yes	CIDI	2 years	<16	Broad	3.0–5.4	Yes
Silverman et al. (1996) (17-year follow-up)	Community	375	50	—	No	No	No	No	CDI & DIS	2 weeks & lifetime	<18	—	2.0	Yes

continued

Table 23.7 Characteristics of studies included in the risk factor–disease relationship analysis (*continued*)

	Sample				Adjusted for confounders									
Level of evidence	*Type*	*N*	*% female*	*Age (years)*	*Demo-graphics*	*Family function*	*Other abuse*	*Other disorders*	*Outcome measure*[a]	*Diagnosis time frame*	*Childhood definition*	*CSA definition*[b]	*OR*	*P <0.05*
Level 3: Cross-sectional studies														
A. Representative community samples														
Molnar et al. (2001)	Adults	5877	50	—	Yes	Yes	Yes	No	CIDI	Lifetime	<18	Broad	1.8	Yes[c]
Saunders et al. (1999)	Adults	4008	100	45	Yes	No	No	No	Clinical interview	Lifetime & 12 months	<18	Intercourse	2.5–2.6	Yes
Stein et al. (1988)	Adults	2683	51	—	Yes	No	No	No	DIS	Lifetime & 6 months	<16	Narrow	2.0–2.64	Yes
Wilsnack et al. (1997)	Community	641	100	—	Yes	No	No	No	DIS	Lifetime	<18	Broad	2.51	Yes
B. Other community samples (representativeness not known)														
Bagley and Ramsay (1985)	Adults	377	100	40	No	No	No	No	CESD	1 month	<17	Narrow	eta = 0.25	Yes
Bagley et al. (1994)	Adults	750	0	23	No	No	No	No	CESD ≥28	1 month	<17	Narrow	—	Yes
Lopez et al. (1995)	Adults	1821	47	—	—	—	—	—	—	—	<17	Broad	—	Yes
Mullen et al. (1996)	Adults	497	100	—	No	Yes	Yes	No	PSE	Lifetime	<16	Broad	1.8	Yes
Peters (1988)	Adults	119	100	—	Yes	No	No	No	SADS	Lifetime	<18	Narrow	—	Yes
Saunders et al. (1992)	Adults	391	100	42	No	No	No	No	DIS	Lifetime & 1 month	<18	Broad	1.65–1.75	Yes[c]
C. Community subgroups or convenience samples														
Arata (1999)	College	92	100	24	No	No	No	No	SCID	Lifetime & 1 month	<14	Broad	—	No

Chandler and Jackson (1997)	College	266	100	19	No	No	No	No	BDI	2 weeks	<18	Broad	—	Yes
Jackson et al. (1990)	College	40	100	23	No	No	No	No	BDI	2 weeks	<18	Narrow	—	No
Zuravin and Fontanella (1999)	Low income women	513	100	30	Yes	Yes	Yes	No	DIS	1 month	<14	Narrow	3.36	Yes
Level 4 studies: Prospective case–control studies														
Swanston et al. (1997) (5-year follow-up)	CPU & controls from community	143	100	15	Yes	Yes	No	No	BDI & CDI	2 weeks	<16	Narrow	48.1	Yes
Level 5: Single wave case–control studies														
Mullen et al. (1993)	Community	492	100	—	No	No	No	No	PSE	Lifetime	<16	Broad	2.6–5.2	Yes
Wise et al. (2001)	Community	732	100	—	Yes	Yes	No	No	SCID	Lifetime	<18	Narrow	2.2	Yes
Panic disorder														
Level 1: Twin studies														
Dinwiddie et al. (2000)	Adult twins	5946	65	43	Yes	No	No	No	SSAGA	Lifetime	<18	Narrow	3.5–5.0	Yes
Kendler et al. (2000) (10-year follow-up)	Adult twins	1411	100	40	No	Yes	No	No	SCID	Lifetime	<16	Broad	1.3–2.6	Yes[c]
Level 2: Prospective studies														
Brown and Harris (1993) (8-year follow-up)	Low income women	404	100	—	No	Yes	Yes	No	PSE	12 months	<17	Broad	—	Yes
Ernst et al. (1993) (10-year follow-up)	Community	421	47	28	No	No	No	No	SPIKE interview	—	<16	Broad	—	No

continued

Table 23.7 Characteristics of studies included in the risk factor–disease relationship analysis (*continued*)

	Sample				Adjusted for confounders									
Level of evidence	*Type*	*N*	*% female*	*Age (years)*	*Demographics*	*Family function*	*Other abuse*	*Other disorders*	*Outcome measure*[a]	*Diagnosis time frame*	*Childhood definition*	*CSA definition*[b]	*OR*	*P <0.05*
Level 3: Cross-sectional studies														
A. Representative community samples														
Molnar et al. (2001)	Adults	5877	50	—	Yes	Yes	Yes	No	CIDI	Lifetime	<18	Broad	0.8–1.4	Yes[c]
Stein et al. (1988)	Adults	2683	51	—	Yes	No	No	No	DIS	Lifetime & 6 months	<16	Narrow	3.4–3.9	Yes[c]
B. Other community samples (representativeness not known)														
Saunders et al. (1992)	Adults	391	100	42	No	No	No	No	DIS	Lifetime & 1 month	<18	Broad	5.0	Yes[c]
C. Community subgroups or convenience samples														
Arata (1999)	College	92	100	24	No	No	No	No	SCID	Lifetime & 1 month	<14	Broad	—	No
Obsessive–compulsive disorder														
Level 3: Cross-sectional studies														
A. Representative community samples														
Stein et al. (1988)	Adults	2683	51	—	Yes	No	No	No	DIS	Lifetime & 6 months	<16	Narrow	—	No
B. Other community samples (representativeness not known)														
Saunders et al. (1992)	Adults	391	100	42	No	No	No	No	DIS	Lifetime & 1 month	<18	Broad	4.5– ≥6	Yes[c]
C. Community subgroups or convenience samples														
Arata (1999)	College	92	100	24	No	No	No	No	SCID	Lifetime & 1 month	<14	Broad	—	No

Post-traumatic stress disorder													
Level 2: Prospective studies													
Silverman et al. (1996) (17-year follow-up)	Community	375	50	—	No	No	No	No	DIS	Lifetime	<18	—	—	Yes
Level 3: Cross-sectional studies													
A. Representative community samples													
Davidson et al. (1991)	Adults	2985	54	—	No	No	No	No	DIS	Lifetime	<16	Intercourse	9.5	Yes
Molnar et al. (2001)	Adults	5877	50	—	Yes	Yes	Yes	No	CIDI	Lifetime	<18	Broad	5.3–10.2	Yes
Saunders et al. (1999)	Adults	4008	100	45	Yes	No	No	No	Clinical interview	Lifetime & 12 months	<18	Intercourse	3.2–2.4	Yes
C. Community subgroups or convenience samples														
Arata (1999)	College	92	100	24	No	No	No	No	SCID	Lifetime & 1 month	<14	Broad	—	Yes
Hien and Bukszpan (1999)	Obs/gyn. clinic	98	100	33	No	No	No	No	SCID	Lifetime	<18	Broad	—	Yes
Robin et al. (1997)	American Indians	375	58	37	No	No	No	No	SADS-1	Lifetime & 1 month	<16	Broad	1.6–8.7	Yes[c]
Schaaf and McCanne (1998)	College	269	100	18	No	No	No	No	Clinical	Current	<15	Broad Interview	—	Yes
Alcohol abuse or dependence														
Level 1: Twin studies														
Dinwiddie et al. (2000)	Adult twins	5946	65	43	Yes	No	No	No	SSAGA	Lifetime	<18	Narrow	1.9–2.8	Yes
Kendler et al. (2000) (10-year follow-up)	Adult twins	1411	100	40	No	Yes	No	No	SCID	Lifetime	<16	Broad	1.9–6.5	Yes[c]
Nelson et al. (2002)	Adult twins	3892	58	30	Yes	Yes	No	Yes	SSAGA	Lifetime	<16	Narrow	1.3–1.7	Yes[c]

continued

Table 23.7 Characteristics of studies included in the risk factor–disease relationship analysis (*continued*)

	Sample				Adjusted for confounders									
Level of evidence	*Type*	*N*	*% female*	*Age (years)*	*Demo-graphics*	*Family function*	*Other abuse*	*Other disorders*	*Outcome measure*[a]	*Diagnosis time frame*	*Childhood definition*	*CSA definition*[b]	*OR*	*P <0.05*
Level 2: Prospective studies														
Fergusson et al. (1996b) (Follow-up U/K)	Community	1019	51	18	Yes	Yes	No	Yes	CIDI	2 years	<16	Broad	1.9–2.7	Yes[c]
Silverman et al. (1996) (17-year follow-up)	Community	375	50	—	No	No	No	No	Various	Lifetime	<18	—	—	Yes
Widom and White (1997) (20-year follow-up)	Community	1190	49	29	No	No	No	No	DIS	Lifetime	<11	Narrow	—	No
Level 3: Cross-sectional studies														
A: Representative community samples														
Fleming et al. (1998)	Adults	710	100	40	Yes	Yes	Yes	No	AUDIT	—	<16	Narrow	0.61	No
Kilpatrick et al. (2000)	Adults	4023	49	—	Yes	Yes	Yes	No	Clinical interview	12 month	<17	Narrow	2.4	Yes
Molnar et al. (2001)	Adults	5877	50	—	Yes	Yes	Yes	No	CIDI	Lifetime	<18	Broad	1.5–1.7	Yes
Saunders et al. (1999)	Adults	4008	100	45	Yes	No	No	No	Clinical interview	Lifetime & 12 months	<18	Intercourse	2.0–2.4	Yes
Spak et al. (1998)	Adults	316	100	—	Yes	No	No	Yes	CIDI-SAM	Lifetime	<18	Broad	3.5	Yes[c]
Stein et al. (1988)	Adults	2683	51	—	Yes	No	No	No	DIS	Lifetime & 6 months	<16	Narrow	1.8	Yes[c]
B. Other community samples (representativeness not known)														
Peters (1988)	Adults	119	100	—	Yes	No	No	No	DIS	Lifetime	<18	Narrow	—	Yes

C. Community subgroups or convenience samples													
Arata (1999)	College	92	100	24	No	No	No	No	SCID	Lifetime & 1 month	<14	Broad	—	No
Robin et al. (1997)	American Indians	375	58	37	No	No	No	No	SADS-I	Lifetime & 1 month	<16	Broad	1.8–2.8	Yes[c]
Drug abuse or dependence														
Level 1: Twin studies														
Kendler et al. (2000) (10-year follow-up)	Adult twins	1411	100	40	No	Yes	No	No	SCID	Lifetime	<16	Broad	1.2–6.6	Yes[c]
Level 2: Prospective studies														
Fergusson et al. (1996b) (Follow-up U/K)		1019	51	18	Yes	Yes	No	Yes	CIDI	2 years	<16	Broad	0.1–2.9	Yes[c]
Silverman et al. (1996) (17-year follow-up)		375	50	—	No	No	No	No	DIS	Lifetime	<18	—	—	No
Widom and White (1997) (20-year follow-up)	Community	1190	49	29	No	No	No	No	DIS	Lifetime	<11	Narrow	—	No
Level 3: Cross-sectional studies														
A. Representative community samples														
Molnar et al. (2001)	Adults	5877	50	—	Yes	Yes	Yes	No	CIDI	Lifetime	<18	Broad	2.0–2.0	Yes
Stein et al. (1988)	Adults	2683	51	—	Yes	No	No	No	DIS	Lifetime & 6 months	<16	Narrow	1.8–2.1	Yes[c]
C. Community subgroups or convenience samples														
Arata (1999)	College	92	100	24	No	No	No	No	SCID	Lifetime & 1 month	<14	Broad	—	No
Robin et al. (1997)	American Indians	375	58	37	No	No	No	No	SADS-I	Lifetime & 1 month	<16	Broad	1.6–4.8	Yes[c]

continued

Table 23.7 Characteristics of studies included in the risk factor–disease relationship analysis (*continued*)

	Sample				Adjusted for confounders									
Level of evidence	*Type*	*N*	*% female*	*Age (years)*	*Demo-graphics*	*Family function*	*Other abuse*	*Other disorders*	*Outcome measure*[a]	*Diagnosis time frame*	*Childhood definition*	*CSA definition*[b]	*OR*	*P <0.05*
Suicide attempts														
Level 1: Twin studies														
Dinwiddie et al. (2000)	Adult twins	5946	65	43	Yes	No	No	No	SSAGA	Lifetime	<18	Narrow	7.1–7.7	Yes
Nelson et al. (2002)	Adult twins	3892	58	30	Yes	Yes	No	Yes	—	Lifetime	<16	Narrow	0.97–1.1	Yes[c]
Level 2: Prospective studies														
Brown et al. (1999) (13-year follow-up)	Community	639	48	—	Yes	Yes	No	Yes	DISC	—	<18	Narrow	5.7	Yes
Ernst et al. (1993) (10-year follow-up)	Community	421	47	28	No	No	No	No	SPIKE interview	—	<16	Broad	—	No
Fergusson et al. (1996b) (Follow-up U/K)	Community	1019	51	18	Yes	Yes	No	Yes	CIDI	2 years	<16	Broad	0.8–5.0	Yes[c]
Silverman et al. (1996) (17-year follow-up)	Community	375	50	—	No	No	No	No	Various	Lifetime	<18	—	10.7	Yes
Level 3: Cross-sectional studies														
A. Representative community samples														
Bagley et al. (1995)	Adolescents	2112	49	—	No	No	No	No	Study-specific questions	6 month	<18	Broad	10–37	Yes
Bensley et al. (1999)	Adults	4790	48	16	Yes	No	Yes	No	YRBS	12 month	<18	Broad	2.7–47.1	Yes
Garnefski and Arends (1998)	Adults	13894	50	15	Yes	No	No	No	Study-specific questions	Lifetime	<19	Narrow	—	Yes[c]

B. Other community samples (representativeness not known)														
Bagley and Ramsay (1985)	Community	377	100	40	No	No	No	No	Paykel (1972)	—	<17	Narrow	eta = 0.16	Yes
Bagley et al. (1994)	Adults	750	0	23	No	No	No	No	Study-specific questions	Lifetime	<17	Narrow	—	Yes
Leth (2001)	Adults	1235	54	—	No	No	No	No	—	Lifetime	<18	Broad	—	Yes
Martin (1996)	Adolescents	352	43	15	No	No	No	No	Smith and Crawford (1986)	6 month	<18	Narrow	—	Yes
Mullen et al. (1996)	Community	497	100	—	No	Yes	Yes	No	Study-specific questions	Lifetime	<16	Broad	3.6	Yes
Saunders et al. (1992)	Adults	391	100	42	No	No	No	No	DIS	Lifetime & 1 month	<18	Broad	3.0	Yes[c]
C. Community subgroups or convenience samples														
Bendixen et al. (1994)	College	996	51	23	No	No	No	No	Study-specific questions	Lifetime	—	Broad	—	No
Boudewyn and Liem (1995)	College	438	61	25	No	No	No	No	Study-specific questions	Lifetime	<14	Narrow	—	Yes
Chandy et al. (1996)	School students	2022	100	15	No	No	No	No	Study-specific question	Lifetime	<18	Narrow	—	Yes
Chandy et al. (1997)	School students	740	0	15	No	No	No	No	Study-specific question	Lifetime	<18	Narrow	—	Yes

continued

Table 23.7 Characteristics of studies included in the risk factor–disease relationship analysis (*continued*)

	Sample				Adjusted for confounders									
Level of evidence	*Type*	*N*	*% female*	*Age (years)*	*Demographics*	*Family function*	*Other abuse*	*Other disorders*	*Outcome measure*[a]	*Diagnosis time frame*	*Childhood definition*	*CSA definition*[b]	*OR*	*P <0.05*
Borowsky et al. (1999)	American Indian and Alaska Native youth	11666	52	15	Yes	Yes	Yes	Yes	Study-specific question	Lifetime	<18	Narrow	—	Yes
Lazartigues et al. (1989)	College	963	57	—	No	No	No	No	—	Lifetime	<16	Broad	—	Yes
Robin et al. (1997)	American Indians	375	58	37	No	No	No	No	Clinical	Lifetime interview	<16	Broad	3.1–6.9	Yes
Sedney and Brooks (1984)	College	102	100	19	No	No	No	No	Study-specific questions	Lifetime	—	Broad	—	No
Hibbard et al. (1988)	School students	712	50	15	No	No	No	No	Study-specific questions	Lifetime	<15	Broad	3.1	Yes
Hibbard et al. (1990)	School students	3998	51	15	No	No	No	No	Study-specific questions	Lifetime	<15	Broad	9.2	Yes
Yama et al. (1995)	College	379	100	20	No	Yes	No	No	Study-specific questions	Lifetime	<16	Narrow	—	Yes

Level 4 studies: Prospective case–control studies														
Plunkett et al. (2001) (9-year follow-up)	CPU & controls from community	259	75	19	No	No	No	No	Study-specific questions	Lifetime	<15	Narrow	1.9	Yes
Level 5: Single wave case–control studies														
Briere and Runtz (1986)	Crisis centre	195	100	27	No	No	No	No	Study-specific questions	Lifetime	<17	Narrow	—	Yes
Molnar et al. (1998)	Street youth	775	35	18	Yes	No	No	No	Study-specific questions	Lifetime	<18	Broad	3.2–4.3	Yes
Mullen et al. (1993)	Community	492	100	—	No	Yes	No	No	Study-specific questions	Lifetime	<16	Intercourse	8.6–25.6	Yes
Completed suicide														
Level 4 studies: Prospective case–control studies														
Plunkett et al. (2001)	CPU & controls from community	259	75	19	No	No	No	No	Death certificate	NA	<15	Narrow	Rate = 0.18%	No

Key: CPU, Hospital Child Protection Units; NA, not applicable; U/K, unknown.

— No data.

[a] Diagnostic instrument definitions: AUDIT = Alcohol Use Disorders Identification Test; BDI = Beck Depression Inventory (Score ≥16 unless otherwise specified); CDI = Children's Depression Inventory (Score ≥20 unless otherwise specified); CESD = Center for Epidemiologic Studies Scale (Depression); CIDI = Composite International Diagnostic Interview; DIS (C) = Diagnostic Interview Schedule (for children) ; PSE = Present State Examination; SADS = Schedule for the Affective Disorders and Schizophrenia; SCID = Structured Clinical Interview for DSM-III-R; SPIKE = Structured Interview to assess psychiatric & psychosomatic symptoms & syndromes, social relationships, coping & life events; SSAGA = Semi-structured assessment for the genetics of alcohol; YRBS = Youth Risk Behaviour Schedule.

[b] Broad: non-contact, contact or intercourse; Narrow: contact or intercourse only.

[c] Significant for some relationships but not others.

- In the absence of a large number of estimates assume that reporting data from fewer sources is likely to be a greater source of error than extrapolation of data to fit the categories required for reporting (e.g. risk by type of exposure or by sex).

Sex

Within the literature, a large proportion of the research has examined the sequelae of CSA in females, leaving males underrepresented. This trend was reflected in the data set with many estimates for females and relatively few for males. Additionally, for males, there were no estimates in the severity categories of non-contact, contact and intercourse. Those studies that did examine sequelae in males only gave estimates for the categories of narrow or broad CSA. Therefore, no data were available in the categories required for analysis for males. Given the assumption that more error would be introduced if no estimates were available for males, a decision was made to extrapolate from data for females. To examine the validity of this decision a comparison between males and females was made between those studies that measured outcome in both (restricted to the narrow and broad categories since males only had data for these).

Comparisons were available across all disorder categories and 95% CIs were compared between male and female estimates within studies to determine if male and female estimates differed significantly. Confidence intervals for the RRs overlapped for depression, agoraphobia, panic, drug and alcohol abuse/dependence for all the studies available for comparison. For PTSD there were significant differences for both of the studies, but as the relative risk was higher for males in one study, and this effect was reversed in the other, no difference between males and females was assumed. For suicide attempts two studies found differences between males and females with the relative risk for males being higher in both. However, in the other six studies available for comparison for suicide attempts no significant differences were observed. Additionally, reviewers have implied that no difference exists between males and females in terms of consequences of CSA (Urquiza and Capra 1990; Watkins and Bentovim 1992). Overall, given that no significant differences were found between the sexes and no theoretical reason presents, no difference between the sexes for the relationship between CSA and mental disorders was assumed.

In light of this decision and in order to maximize the number of estimates available for analysis a hierarchy for selection of studies based on sex was constructed. Where estimates for males and females combined were available from a study these were selected first. Female estimates were then chosen followed by male estimates. This ensured that one estimate was available for each outcome that each study reported.

Diagnostic time frame

There was considerable variation in the time frame used to measure outcome across studies. While some studies determined the presence of mental disorders over an individual's lifetime, others determined the presence of disorder over the past 12 months, six months or one month. While current or one month estimates might be considered the most accurate measure of current risk, these were only presented for a small number of studies (see Table 23.7). In order to examine the differences between lifetime and current estimates the following analysis was undertaken.

Lifetime and current estimates of risk were compared in the six studies that presented both. All six samples were in the 30–44-year age group. There were five sets of estimates for depression, two for panic disorder, two for drug dependence, three for alcohol dependence and two for PTSD. Only one study presented estimates for males. Ratios of current to lifetime relative risk ranged from 0.57 to 3.21 with a trend for current risks to be greater than lifetime (ratios of >1). However, when confidence intervals around RRs were examined very few comparisons were significant.

A second set of comparisons was also undertaken. Relative risks across all studies were grouped according to diagnostic time frame and averaged within age groups and outcomes. Again no clear pattern emerged. On the basis of these analyses and in order to maximize the number of estimates it was decided to include all estimates of relative risk regardless of whether the diagnostic time frame used was lifetime, current or 12 months. Where studies presented both, current risk was used.

Levels of exposure

Many studies presented relative risk for exposure vs non-exposure only, rather than by levels of exposure. Moreover these studies varied in terms of whether they presented relative risk for broad CSA (non-contact, contact or intercourse) or for narrow CSA (contact or intercourse only). There is strong evidence in the literature to suggest that outcome varies with level of exposure, risk being the highest for those who have experienced abuse involving intercourse (Fergusson and Mullen 1999). In order to examine the relationship between relative risks for each level of exposure the following analysis was undertaken.

Relative risks for overall exposure to CSA were calculated for all studies ($N = 5$) that presented risk by level of exposure. Relative risks for each level of exposure were then expressed as a ratio of the overall risk. Estimates and ratios are presented in Table 23.8.

The ratios in Table 23.8 were then applied to those studies that only reported risk for exposed vs non-exposed. More specifically, the ratios derived from those studies reporting risk estimates for the contact and intercourse categories of abuse were applied to those studies reporting

Table 23.8 Relative risks of each disorder for each level of exposure, as a ratio of overall relative risk[a]

	Study characteristics		*Exposure level*			*Any CSA*		*Ratios*		
Outcome	*Sample*	*N*	RR_1	RR_2	RR_3	RR_{any} *broad*	RR_{any} *narrow*	RR_1/RR_{any}	RR_2/RR_{any}	RR_3/RR_{any}
Depression										
Saunders et al. (1992)	Community	391	1.24	1.65	1.76	1.57	—	0.79	1.05	1.12
Fergusson et al. (1996b)	Community	1019	2.19	2.04	3.83	2.56	—	0.86	0.80	1.50
Kendler et al. (2000)	Community	1411	1.25	1.39	1.83	1.48	—	0.84	0.94	1.24
Mullen et al. (1993)	Community	492	—	—	4.38	2.62	—	—	—	1.67
Panic disorder										
Saunders et al. (1992)	Community	391	1.49	0.91	2.67	1.59	—	0.94	0.57	1.68
Kendler et al. (2000)	Community	1411	1.41	1.76	2.42	1.86	—	0.76	0.95	1.30
Alcohol dependence										
Fergusson et al. (1996b)	Community	1019	1.63	2.19	2.02	2.01	—	0.81	1.09	1.00
Kendler et al. (1996b)	Community	1411	2.32	2.29	3.43	2.61	—	0.89	0.88	1.31
Drug dependence										
Fergusson et al. (1996b)	Community	1019	0.78	1.64	3.66	2.13	—	0.37	0.77	1.72
Kendler et al. (2000)	Community	1411	2.63	2.03	5.19	3.05	—	0.86	0.67	1.70
Suicide										
Saunders et al. (1992)	Community	391	0.50	2.74	3.11	2.25	—	0.22	1.22	1.38
Fergusson et al. (1996b)	Community	1019	1.03	2.15	3.53	2.33	—	0.44	0.92	1.47
Mean								0.71	0.90	1.42
Depression										
Saunders et al. (1992)	Community	391	—	1.65	1.76	—	1.69	—	0.98	1.04
Fergusson et al. (1996b)	Community	1019	—	2.04	3.83	—	2.68	—	0.76	1.43
Kendler et al. (2000)	Community	1411	—	1.39	1.83	—	1.56	—	0.89	1.17
Banyard (1999)	Low income women	518	—	1.17	2.38	—	2.04	—	0.57	1.17
Panic disorder										
Saunders et al. (1992)	Community	391	—	0.91	2.67	—	1.63	—	0.56	1.57

Table 23.8 Relative risks of each disorder for each level of exposure, as a ratio of overall relative risk[a] (*continued*)

						Any CSA		Ratios		
	Study characteristics		Exposure level			RR_{any}	RR_{any}	RR_1/RR_{any}	RR_2/RR_{any}	RR_3/RR_{any}
Outcome	Sample	N	RR_1	RR_2	RR_3	broad	narrow			
Kendler et al. (2000)	Community	1411	—	1.76	2.42	—	2.01	—	0.88	1.20
Agoraphobia										
Saunders et al. (1992)	Community	391	—	2.03	5.19	—	3.31	—	0.61	1.57
Alcohol dependence										
Fergusson et al. (1996b)	Community	1019	—	2.19	2.02	—	2.12	—	1.03	0.95
Kendler et al. (2000)	Community	1411	—	2.29	3.43	—	2.71	—	0.85	1.27
Drug dependence										
Fergusson et al. (1996b)	Community	1019	—	1.64	3.66	—	2.53	—	0.65	1.45
Kendler et al. (2000)	Community	1411	—	2.03	5.19	—	3.20	—	0.63	1.62
Suicide										
Saunders et al. (1992)	Community	391	—	2.74	3.11	—	2.89	—	0.95	1.08
Fergusson et al. (1996b)	Community	1019	—	2.15	3.43	—	2.71	—	0.79	1.27
Mean									0.78	1.29

— No data.

[a] RR_1, RR_2, RR_3 refers to RRs for level 1 exposure (non-contact), level 2 (contact) and level 3 (intercourse).

Note: 95% CIs were calculated but are not presented here. All estimates are for females or all persons. None are for males.

risk estimates for the narrow category of abuse. Ratios from studies reporting risk estimates in the non-contact, contact and intercourse categories were applied to the risk estimates for the broad category of abuse. This process generated relative risks for the three categories of abuse required for analysis in relation to the single estimate reported by each study. In this way, the extrapolated relative risks reflect an approximation of what the risk may have been if each study had reported risks for levels of exposure, rather than only risks for exposure vs non-exposure. This ensured that data from studies reporting risk in terms of exposed vs non-exposed, rather than risk by levels of exposure, could still be included in the analysis.

The accuracy of the extrapolated relative risks relies on the premise that risk increases with level of exposure to the same degree across

studies. It was decided, however, that the error introduced by the extrapolation process was less than the error introduced by pooling relative risk estimates from a small number of studies. After the extrapolation process each study that reported an estimate only for the broad category of abuse now had relative risks for the non-contact, contact and intercourse categories of abuse. Accordingly, each study that reported an estimate for the narrow category of abuse now had a relative risk for the contact and intercourse categories of abuse.

In order to calculate confidence intervals for those relative risks that had been extrapolated, standard errors had to be estimated. In this instance, the standard errors derived for the relative risks for the broad and narrow categories of abuse were used. This may not accurately reflect—indeed may underestimate—the true variance around the extrapolated estimates, but for the purposes of this analysis it was assumed to be a reasonable approximation.

Adjustment for confounders

There is strong evidence within the literature that child sexual abuse is often comorbid with other forms of child abuse and also that child abuse occurs within the context of other family dysfunction (Fergusson and Mullen 1999). Given this, any studies that do not control for these other childhood adversities may inflate the contribution child sexual abuse makes to the onset of our chosen outcomes. Only 13 studies in the data set controlled for confounders. In order to adjust uncontrolled estimates for the potential contribution of confounders a method proposed by Rothman and Greenland (1998) was utilized.

Each of the 13 studies varied in terms of outcomes measured and confounders controlled for. For each study the confounders were recorded and grouped into four categories: sociodemographic, other psychopathology, other abuse and family dysfunction. The studies varied according to the types of confounders controlled for and the measures used. In particular, the category of family dysfunction represented a wide variety of factors. The measures used varied from questionnaires concerning parental attachment and parent/child bonding through to those measuring markers of dysfunction, such as whether the subject grew up in a nuclear family or whether there was parental psychopathology. While the measures varied greatly it was assumed that a common underlying dimension was being measured, that is, the degree to which the family environment was impoverished or dysfunctional and so they were grouped together.

At this stage five studies were excluded from the analysis, as they were not consistent with the pattern of confounders measured by the other seven studies. Saunders et al. (1999), Molnar et al. (1998) and Wilsnack et al. (1997) were excluded since they only controlled for sociodemographic variables and might dilute the adjustment factor if they contributed to the average estimate. Stein et al. (1988) was excluded since

it did not control for any childhood adversity, instead only controlling for subsequent adult sexual abuse. Spak et al. (1998) controlled for childhood behavioural difficulties and childhood psychopathology. These may be important factors confounding the relationship between CSA and alcohol dependence, but, it is unclear to what extent childhood behavioural difficulties and childhood psychopathology are additive to the effects of other confounders. It was therefore excluded.

The eight studies left in the analysis controlled for family dysfunction or other types of abuse, including physical and emotional abuse. Therefore the adjustment factor derived from these studies reflects an adjustment for childhood adversity stemming from dysfunctional home and family environments and is the confound identified as being most important in the literature (Chandler and Jackson 1997; Fergusson and Mullen 1999).

For each of the eight studies adjusted odds ratios and unadjusted odds ratios could be derived. Using the odds ratios an adjustment factor was calculated using the following formula:

$$U = OR_u / OR_a$$

OR_u represents the unadjusted odds ratio, OR_a represents the odds ratio adjusted for confounders and U is the bias produced from failure to control for the confounders. Since there were only eight studies some assumptions about the commonality of effect across sex, disorder categories and abuse categories had to be made. Table 23.9 presents the final adjustment factors by study and disorder averaged across abuse category and sex.

Extrapolation across sex

Of the eight studies, seven provided estimates for females but only two provided estimates for males. One study provided estimates for males and females combined. Of the two studies that presented data for both males and females there was no clear pattern of differences between the sexes. Additionally, there is no theoretical reason to expect that the confounders in question would differentially affect the relationship between exposure to CSA and disorder; and due to the paucity of data for males we assumed no difference.

Extrapolation across levels of exposure

With the exception of three studies, all of the studies provided estimates for only the narrow or broad categories of abuse. In the absence of any data it was assumed that the effects of confounders would be the same across abuse categories.

Table 23.9 Adjustment factors for family dysfunction according to disorder type and study

		Adjustment factor (U)						
Study	*Sample type*	*Depression*	*Panic*	*Drug*	*Alcohol*	*PTSD*	*Disorder mean*	*Suicide*
Molnar et al. (2001)[a]	Community	1.68	1.69	1.84	1.55	1.53	1.62	—
Kendler et al. (2000)[b]	Community	1.21	1.02	1.19	0.92	—	1.08	—
Fergusson et al. (1996b)[b]	Community	1.43	—	0.86	1.06	—	1.12	1.64
Mullen et al. (1996)[a]	Community	2.11	—	—	—	—	2.11	5.48
Zuravin and Fontanella (1999)[a]	Low income mothers	1.28	—	—	—	—	1.28	—
Mullen et al. (1993)[a]	Community	—	—	—	—	—	—	2.39
Yama et al. (1995)[a]	College	—	—	—	—	—	—	1.71
Borowsky et al. (1999)[b]	American Indian and Alaska Native youth	—	—	—	—	—	—	2.08
Mean		1.54	1.35	1.29	1.18	1.53	1.39[c]	2.66

— No data.

[a] Controlled for both other abuse and family factors.

[b] Controlled only for family factors.

[c] Excludes suicide.

Extrapolation across outcomes

Most of the eight studies in the analysis only provided estimates for one or two disorders. There were three studies that gave estimates for most of the disorders, excluding suicide attempts; and within these studies there was no definitive variation across disorder. This conclusion is again limited by paucity of data, but there is no theoretical reason for the confounders to act differentially according to disorder. Suicide attempts appear to be an exception since when the means across disorder and study are compared the adjustment factor for suicide is higher. This may indicate that the confounding variables are more predictive of suicide attempts than the other disorders considered and hence, to be conservative, a different adjustment factor is applied to the suicide estimates.

Generalizability of the adjustment factor

The validity of adjusting estimates from uncontrolled studies using an adjustment factor derived from studies that do control for confounders will be accurate only to the extent that the confounding effects of the covariates are similar across both the controlled and uncontrolled studies (Greenland 1987). While there is no way to assess this issue quantitatively, we can consider the samples from which our adjustment factors were derived. If the samples from controlled studies are drawn from significantly different groups within the community then generalizability to the uncontrolled studies may be limited. In the current analysis the controlled samples included four community and one college sample that were representative of the samples from uncontrolled studies. Moreover, the two samples from community subgroups, low-income mothers, and American Indian and Alaska Native youth, provided estimates that are comparable to the community estimates.

Application of adjustment factors

Adjustment factors were applied differentially across the risk estimates according to several criteria. For those eight studies reporting both adjusted and unadjusted risk estimates from which an adjustment factor could be calculated, its own individual adjustment factor was applied to each. For those studies that reported no adjusted estimates and from which an adjustment factor could not be derived, the average of the adjustment factors was applied. The exception was where the unadjusted relative risks were non-significant, and in this instance the relative risks were not adjusted for the presence of confounders. This ensured that a significant protective relationship between CSA and psychiatric outcome was not created artificially.

Meta-analysis

The relative risks were combined using meta-analysis with STATA Intercooled 7. For ease of calculation the macro "meta" was utilized (Sharp and Sterne 1997). Estimates were grouped according to psychiatric outcome and then combined. In most cases the studies combined within each group were significantly heterogeneous according to Cochrane's Q statistic, indicating that moderator variables other than psychiatric outcome were still accounting for significant variation. However the small number of studies prevented further partitioning according to other hypothesized moderator variables.

The presence of significant, unexplained heterogeneity generally indicates preference for a random-effects model to take into account the between-study heterogeneity (Cooper and Hedges 1994; Rothman and Greenland 1998). However, when only two or three studies are available for combination, the between-study variance is estimated with poor precision (Cooper and Hedges 1994). In this instance it was decided that

Table 23.10 Number of studies contributing to each estimate within disorder category

	No. of studies		
Outcome	*Non-contact*	*Contact*	*Intercourse*
Depression	14	23	25
Panic disorder	5	8	8
Alcohol abuse/dependence	7	13	15
Drug abuse/dependence	4	7	7
PTSD	5	6	8
Suicide attempts	13	29	29

groups with five or more studies would be combined using a random-effects model and those with less than five would use a fixed-effects model.

3.4 Assessment of causality for each outcome

The following summarizes the evidence of a causal relationship between CSA and each of the seven outcomes examined. A schema or hierarchy was developed to organize this evidence and is outlined below. All of the studies that reported relevant data on proportions of persons in the clinical ranges for each outcome have been tabulated by level of evidence. For some, the χ^2 statistic was calculated to test for significance of the odds ratios where these were not available from the studies themselves. Where research on a given psychiatric outcome focused on a subsample of a study group that has been described previously, the findings of the main sample are reported.

Level 1: Studies controlling for both genetic background and family environment.

Level 2: Prospective studies where family environment measured prospectively was used to control confounding of deprivation and CSA.

Level 3: Retrospective cross-sectional studies in which the occurrence of CSA was determined at the time illness was ascertained. Family environment measured reliably was used to control confounding of deprivation and CSA.

Level 3a: Representative community samples: samples of adolescents or adults where either sampling strategy or weighting procedures ensured representativeness of sample.

Level 3b: Non-representative community samples: samples of adolescents or adults where methodology did not necessarily ensure representativeness of sample.

Level 3c: Community subgroup samples: samples of college students, general practice attendees or other community subgroups.

Level 4: Prospective case–control studies in which a CSA group was compared with controls matched for family environment (actually or statistically) and followed over time to measure the onset of mental disorders.

Level 5: Single wave case–control studies in which a CSA group was compared with controls matched for family environment (actually or statistically).

Level 6: Studies of special groups such as foster care children, street youth and juvenile detainees.

Depression

The most convincing evidence for a relationship between child sexual abuse and adult depression is from three level 1 studies of adult twins. Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between contact or intercourse CSA and depression for female (OR = 2.20) and male twins (OR = 3.93). Restricting the sample to twin pairs discordant for CSA (and thereby substantially reducing the power of the analysis), the relationship between CSA and depression was no longer statistically significant for either males or females even though both were at increased risk for depression.

Kendler et al. (2000) adjusted for family functioning and parental psychopathology in a sample of female twins. They found that odds ratios for the risk of depression in abused twins compared to non-abused twins were modest but not significant for non-genital abuse (OR = 1.08), but increased with the severity of the abuse. For abuse involving genital contact (OR = 1.58) and intercourse (OR = 2.79), odds ratios were significant. When analyses were restricted to twins discordant for CSA, odds ratios were of similar magnitude to the previous analysis but only risk in the intercourse category of abuse remained statistically significant (Kendler et al. 2000). It was noted by the author that the discordant twin analyses were limited by small sample size.

In a sample consisting of monozygotic and dizygotic twins who were discordant for CSA, the risk of developing major depression was 1.68 for women and 1.25 in men; however, this relationship was significant only for women (Nelson et al. 2002). Of note was the non-significant trend for non-abused co-twins to also be at greater risk of having a history of depression in comparison to twin pairs with no history of abuse, providing evidence for the contribution of familial factors to the onset of disorder. Therefore, the finding that abused co-twins have a higher risk of depression than their non-abused co-twins demonstrates the increased risk that CSA contributes over and above family background.

Across the three twin studies a significant relationship between CSA and depression has been found. Particularly, abuse involving contact and penetration has been found to significantly increase the risk of a depressive disorder. Discordant twin analyses were conducted for three studies and have the advantage of controlling more tightly for familial and genetic factors. These analyses also found CSA to contribute significantly to onset of depression, but in the case of Dinwiddie et al. (2000) and Kendler et al. (2000) only when the abuse involved intercourse. While the small sample sizes of the discordant twins restricted the power of the analyses they do provide strong support for a causal relationship between depression and CSA involving intercourse.

To conclude, there have been three level 1, and 19 level 2 and 3 studies examining the relationship between CSA and depression; 19 of them supported a significant relationship. There is strong, consistent evidence that, after adjustment for confounders, there is a significant relationship between CSA and depression in adults, particularly for women who experienced more severe forms of abuse.

Panic disorder

There is evidence from level 1 twin studies for a relationship between CSA and panic disorder. Defining sexual abuse as contact or intercourse, Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between CSA and panic disorder for female (OR = 3.54) and male twins (OR = 5.02). In the twin pairs discordant for CSA, the relationship between CSA and panic disorder was no longer significant for women even though they were at increased risk for the disorder (OR = 2.00). Similar analyses were unable to be conducted for men because of sample size.

In their analyses of panic disorder, Kendler et al. (2000) adjusted for family functioning (but not parental psychopathology) and found that odds ratios in abused twins compared to non-abused twins were modest but not significant for either non-genital abuse (OR = 1.25) or for abuse involving genital contact (OR = 1.92). Odds ratios for intercourse (OR = 2.62) were significant. Discordant co-twin analyses were not performed due to small sample sizes. Prospective studies have found equivocal evidence of a relationship (Brown and Harris 1993; Ernst et al. 1993) and studies using community samples have found significant associations between some forms of CSA but not others (Molnar et al. 2001; Saunders et al. 1992; Stein et al. 1988). Although few studies have examined panic disorder as an outcome of CSA, there is evidence from two level 1 studies, one level 2 study and three level 3 studies that the rates of panic disorder are increased in sexually abused adults and are more strongly predicted by abuse involving penetration.

Obsessive–compulsive disorder (OCD)

Only three studies explored the relationship between CSA and OCD and only one of these controlled for confounders. This representative community sample (level 3a) adjusted for demographics and other abuse but found no significant relationships between CSA and lifetime or 6-month history of OCD (Stein et al. 1988). Results from the two samples that did not adjust for confounders were mixed. In a sample of female college students (level 3c) (Arata 1999), no significant relationships were reported between CSA and lifetime or 1-month history of OCD. However, childhood was defined as aged <14 years and was therefore quite restrictive. Saunders et al. (1992) (level 3b) defined childhood more broadly, as aged <18 years. While results for non-contact abuse were not significant, women were 4.5 times more likely to have a lifetime history of OCD if they had experienced CSA in the form of contact abuse and over six times more likely if they had experienced intercourse; and these findings were significant. This pattern of significant results was also the case for the relationship between CSA and women who met the diagnostic criteria for OCD at interview. In summary, only one study supported a relationship between CSA and OCD. Further research is required in order to confirm this association. For this reason the risk for OCD will not be calculated in this report.

Post-traumatic stress disorder (PTSD)

There was one study of PTSD that provided level 2 evidence (Silverman et al. 1996). This was a 17-year follow-up of a community sample of children in a working class area. There was no adjustment for confounders; however, a significant relationship was found between CSA and lifetime history of PTSD for females at age 21 years.

Of the remaining studies of PTSD, all were from level 3 evidence and only two of these adjusted for confounders. These studies were representative community samples (level 3a). One study restricted CSA to intercourse only, and only adjusted for age, but found that CSA significantly increased the risk of lifetime (OR = 3.42) and recent (OR = 3.17) PTSD (Saunders et al. 1999). The other study adjusted for demographic and family variables such as parental substance abuse and psychopathology, and the presence of physical abuse (Molnar et al. 2001). Child sexual abuse was defined as contact sexual abuse or intercourse and was found to be significantly related to lifetime history of PTSD in women (OR = 10.2) and men (OR = 5.3). A notable finding was that the risk of PTSD was significantly higher for penetrative abuse than contact abuse.

In summary, there were eight level 2 and 3 studies; all showed a significant relationship between child sexual abuse and adult PTSD. There is strong, consistent evidence that, after adjustment for confounders, there is a significant relationship between CSA and PTSD in adults, particularly for those who have experienced more severe forms of abuse.

Alcohol abuse or dependence

The evidence for a relationship between childhood sexual abuse and adult alcohol abuse/dependence is from the three level 1 studies of adult twins. Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between contact or intercourse CSA and alcohol abuse/dependence for female (OR = 2.81) and male twins (OR = 1.91). Restricting the sample to twin pairs discordant for CSA, the relationship between CSA and alcohol abuse/dependence was no longer significant for either males or females; however females remained at increased risk (OR = 2.50).

Kendler et al. (2000) adjusted for family functioning and parental psychopathology in a sample of female twins. The odds ratios for the risk of alcohol abuse/dependence in abused twins compared to non-abused twins were modest but not significant for genital contact abuse (OR = 1.91). For non-genital abuse (OR = 3.20) and intercourse (OR = 6.48), odds ratios were significant. Analyses for twins discordant for CSA, or where the co-twin had experienced a less severe form of abuse, showed that intercourse significantly increased the risk of alcohol abuse/dependence. In another sample that also presented results for CSA among discordant twins the risk of developing alcohol abuse/dependence was 1.73 for women and 1.25 in men; however this relationship was significant only for women (Nelson et al. 2002). In particular, the findings from the discordant twin analyses suggest a causal relationship between CSA and alcohol abuse/dependence, and indicate that CSA increases the risk over and above that arising from family background. Low sample size of the discordant twin analyses may have produced the non-significant results for contact forms of abuse but the significant finding for intercourse in spite of low power makes this finding notable. Other prospective and community studies also provide evidence for a relationship between CSA and alcohol abuse/dependence with two out of three level 2 studies and seven out of nine level 3 studies presenting significant odds ratios.

Drug abuse or dependence

There was only one level 1 study that explored drug abuse or dependence. In their study of female twins, Kendler et al. (2000) adjusted for family functioning and parental psychopathology. They found that odds ratios for the risk of drug abuse/dependence in abused twins compared to non-abused twins were modest but not significant for genital contact abuse (OR = 1.21). For non-genital abuse (OR = 3.57) and intercourse (OR = 6.55), odds ratios were significant. Analyses of twins discordant for CSA, or where the co-twin had experienced a less severe form of abuse, showed a similar pattern of results with non-genital CSA and CSA involving intercourse placing subjects at increased risk (ORs 4.29 and 2.85, respectively). However, these results were not significant, and this is likely to be a function of small sample sizes.

Of the three level 2 studies, only one showed a significant association and was significant only for the intercourse category of abuse (OR = 5.1) (Fergusson et al. 1996b). The two other prospective studies failed to show a relationship (Silverman et al. 1996; Widom and White 1997). Of the four level 3 studies using community samples, three demonstrated a significant association but not across all three levels of abuse. In summary, there is evidence to suggest a relationship between CSA and drug dependence but only for the more severe forms of abuse. However, the evidence is more equivocal compared to other outcomes and more research is required to confirm this relationship.

Suicide attempts

There were two level 1 studies of adult twins that addressed suicide attempts. In the first, Dinwiddie et al. (2000) adjusted for demographic factors and found significant relationships between contact or intercourse CSA and serious suicide attempts for female (OR = 7.74) and male twins (OR = 7.07). Restricting the sample to twin pairs discordant for CSA, the relationship between CSA and suicide attempts was no longer significant for females; however they were still at increased risk (OR = 2.33). Risk estimates could not be computed for males in the discordant twin analysis due to small numbers. In the second study of twins who were discordant for CSA (Nelson et al. 2002) the risk of suicide attempts was 2.33 for women and 4.50 in men. This relationship was significant for women, and almost reached significance in men (95% CI 0.97–20.83), even after adjustment for demographic and family factors. Furthermore, three of four prospective studies (level 2) and 16 of 18 level 3 studies also found significant associations between CSA and suicide attempts. There is strong evidence for a relationship between CSA and suicide attempts.

Completed suicide

Completed suicide, by definition, can only be studied prospectively. Only one study to date has examined the relationship between completed suicide and child sexual abuse. This Australian study was based on a relatively small sample of sexually abused young people, the majority of them female, who presented to hospitals for the abuse, and a control group of non-abused young people from the community (Plunkett et al. 2001). Without controlling for confounders, CSA was found to increase the risk of completed suicide; and the rate was found to be very high, 179.5 per 100 000 person-years, or 1.8%, but this relationship was not significant. It should be noted that the number of completed suicides was 3 out of a sample of 259, and none of the controls had committed suicide, so analyses were somewhat limited. The national suicide death rates for 15–24 year olds during the same time period ranged from 13.8 to 16.7, so the observed rate in the study was 10.7 to 13.0 times that of the Australian national suicide death rate (Dudley et al. 1998).

Relationship between attempted and completed suicide

Attempted suicide has consistently been shown to be a strong predictor of completed suicide (see Graham et al. 2000 for review). Estimates of the magnitude of this risk however, vary considerably. The prevalence of completed suicides among those who have attempted has been estimated at 1% to 19% in the 12 months after the attempt (Diekstra 1992; Graham et al. 2000), 2.8% after 8 years (Hawton and Fagg 1988) and 10% after 10 years (Tejedor et al. 1999). Lifetime prevalence of completed suicide in those who have ever been hospitalized for suicidality has been estimated at 8.6% (Bostwick and Pankratz 2000). Although difficult to quantify on the basis of available data, the evidence suggests that approximately 1 in 10 individuals who attempt suicide will die by suicide at some point following the attempt. These reviews also pointed out, however, that the majority of completed suicides are not preceded by a suicide attempt (Graham et al. 2000). Estimates of risk for completed suicide that are derived only from estimates of previous attempts are therefore likely to underestimate the risk of completed suicide.

The issue is further complicated in that mental disorders have also been shown to be strong and consistent predictors of both suicide attempts and completed suicide (Brent et al. 1999; Graham et al. 2000; Harris and Barraclough 1997; Hawton and Fagg 1988; Kessler et al. 1999). Within this context, the exact nature of the relationship between CSA, mental disorder, suicide attempts and completed suicide is likely to be complex (Beautrais 2000). Unfortunately, none of the studies in the present report that examined the relationship between CSA and suicide attempts controlled for concurrent psychopathology.

From the above it is concluded that there is evidence of a relationship between attempted suicide and completed suicide. However, for the purposes of the current analysis it is necessary to determine whether a person who has been subject to CSA has any higher or lower chance of completing suicide conditional on having attempted it. To our knowledge these data do not exist. In the absence of these data the relative risk for suicide attempts will be used as proxies for the relative risk for completed suicide. This assumes a constant relationship between attempted and completed suicide given exposure to CSA. The extent to which this underestimates or overestimates the relative risk is unknown.

3.5 Estimates of risk factor–disease relationships by age, sex and subregion

Estimates of risk

Table 23.11 summarizes the results of the meta-analysis and reports the relative risks for the psychiatric outcomes examined. No estimates are available for different age and sex groups. All studies were from AMR-A, WPR-A or EUR-A. No studies were available from other subregions that may have different cultural norms and socioeconomic

Table 23.11 Unadjusted and adjusted relative risks across disorder and abuse categories

	Non-contact abuse				*Contact abuse*				*Intercourse*			
	Adjusted[a]		*Unadjusted*		*Adjusted*[a]		*Unadjusted*		*Adjusted*[a]		*Unadjusted*	
	Pooled RR	*95% CI*	*Pooled RR*	*95% CI*	*Pooled RR*	*95% CI*	*Pooled RR*	*95% CI*	*Pooled RR*	*95% CI*	*Pooled RR*	*95% CI*
Depression	1.06	0.91–1.24	1.37	1.16–1.61	1.32	1.16–1.51	1.80	1.61–2.02	2.04	1.78–2.35	2.80	2.44–3.22
Panic disorder	1.01	0.76–1.35	1.52	1.13–2.04	1.64	1.12–2.42	2.27	1.56–3.32	2.60	1.70–3.97	3.58	2.34–5.48
Alcohol abuse/dependence	1.19	0.85–1.67	1.35	0.91–2.02	1.32	1.07–1.63	1.61	1.24–2.09	1.87	1.47–2.39	2.58	2.04–3.24
Drug abuse/dependence	1.03[b]	0.84–1.26	1.57[b]	1.28–1.93	1.31	0.90–1.91	1.68	1.03–2.72	2.40	1.46–3.96	3.04	1.83–5.04
PTSD	1.95	0.95–3.98	2.70	1.32–5.54	2.95	1.53–5.68	4.10	2.12–7.90	4.48	2.33–8.65	6.23	3.23–12.02
Suicide attempts	1.02	0.71–1.45	2.80	1.89–4.15	1.32	1.08–1.60	3.25	2.53–4.18	2.21	1.77–2.76	5.56	4.32–7.16

[a] These relative risks have been adjusted for family dysfunction and other types of abuse. The RRs for suicide attempts have been divided by 2.66 and the RRs for the mental disorders have been adjusted by 1.39.

[b] Fixed-effects model used to combine estimates due to small number of studies available.

circumstances. No differences in risk were assumed across sex and age breakdowns since numbers were few and many studies used lifetime diagnoses.

On the whole several conclusions can be drawn from these results. First, the relative risks were not significantly different across types of mental disorder, suggesting that CSA is not particularly associated with any one disorder. Rather the risk appears to be pervasive across the whole spectrum of mental disorders examined. This lack of specificity makes CSA particularly damaging, putting individuals at risk for a wide range of mental disorders.

Second, with the exception of suicide attempts, risks did not vary significantly across categories of abuse. This may reflect the small number of studies available for analysis, as there is a general trend for increased risk to be associated with "increased" exposure to CSA. That is, as more severe forms of CSA are experienced the risks for developing a mental disorder increase. This may indicate that those exposed to CSA do not represent a homogeneous group but instead reflect a group that varies in terms of exposure and subsequent risk for psychiatric disorder (Fergusson and Mullen 1999). However, this finding is largely an artefact of the extrapolation process. Further research will enable this "dose–response relationship" to be confirmed.

Non-contact abuse was not a significant predictor of risk after external adjustment for confounders. Non-contact abuse may constitute a more heterogeneous category of abuse compared to the contact and intercourse categories. Certainly, in the studies that contributed to this analysis the category of non-contact abuse encompassed a variety of acts. Such heterogeneity may make the results of the analysis hard to interpret. The question is whether non-contact does not place individuals at increased risk for disorders or whether some non-contact forms of abuse are more or less harmful than others. Cultural factors are likely to play a role in this and further investigation is required. Across the majority of disorders (excluding drug abuse/dependence) the relative risks for contact and intercourse forms of abuse remained significant after external adjustment. These results are more easily interpreted as the two categories are a more homogeneous group of acts and can be more tightly defined.

Results of the current meta-analysis were consistent with the only other review to look systematically at psychiatric diagnosis as an outcome of CSA. Fergusson and Mullen (1999) collated and reanalysed data from 12 studies reporting on the relationship between CSA and psychiatric dysfunction. While Fergusson and Mullen (1999) did not statistically combine the estimates, consistent and pervasive relationships between CSA and adult psychopathology were apparent. Other meta-analyses have been conducted in this area (Jumper 1995; Paolucci et al. 2001; Rind 1997; Rind et al. 1998) but have focused on continuous measures of adjustment. Jumper (1995) and Paolucci et al. (2001) have com-

mented on the difficulty in quantifying the impact that confounders have since many large-scale community studies do not measure them. Rind et al. (1998) was the only meta-analysis to control for the confounding effect of family environment using college samples since confounds are more likely to be measured in studies using these samples. Results of their analysis showed that the relationship between CSA and adjustment in adulthood disappeared after controlling for family environment. Adjustment in their analysis refers to the various psychological correlates of CSA measured in the studies that contributed to their analysis. Eighteen categories of psychological correlates were coded from the studies and included anxiety and depressive symptoms, plus broader areas of adjustment such as sexual adjustment and social adjustment. When commenting on Rind et al. (1998) results, Kendler et al. (2000) stated that CSA appears to be more related to lifetime psychiatric disorder rather than cross-sectional measures of adjustment which focus solely on current well-being. Certainly, the results of the current meta-analysis do suggest a significant relationship between CSA and psychiatric disorder after adjustment for important confounders.

Several caveats should be mentioned in regard to the current analysis. The external method of adjustment used was limited by the methodology of the studies from which the adjustment factors were derived. Across the studies family dysfunction was not defined or measured consistently and the measures that were used may not be reliable. It should also be noted that current research methodologies make it impossible to determine whether any antecedent signs or symptoms of psychiatric disorders mean that the child is more vulnerable to abuse, thus inflating the risk factor–disease relationship. Additionally, the majority of studies contributing to this analysis were conducted during a time when CSA was ignored and the victims were stigmatized. While no cohort effects were found for reporting episodes of CSA the stigmatization could possibly have had an impact on the degree to which people were affected.

To conclude, it appears that CSA is particularly damaging with effects evident over and above other forms of childhood adversity. Contact and intercourse categories of abuse remained significant after controlling for confounders but the category of non-contact was non-significant. Further research is required to understand the effects of various forms of non-contact abuse.

Risk reversibility

By definition, exposure to CSA stops at the age of 17 years. The reversibility of risk therefore refers simply to decline in the risk of developing mental disorder given exposure to CSA over the lifespan. Problems in quantifying the decline arise from both paucity of data and competing theoretical arguments. There are two arguments. Either assume that risk remains constant throughout the lifespan or assume that

risk subsides over time and eventually becomes that of people who were never abused. One can argue that risk remains constant over time as there is evidence to suggest that CSA alters your life trajectory such that you are more likely to experience problems with relationships, self-esteem and sexual adjustment (Mullen et al. 2000). These problems in themselves are associated with mental disorder and so potentially mediate the relationship between CSA and disorder in later life. Therefore, those exposed to CSA remain at increased risk for mental disorder compared to those who have not been exposed. Alternatively one can argue that risk decreases over the lifespan. As a person moves away from a traumatic life event, its power to inflict psychological harm is lessened and so those exposed to CSA eventually have the same risk for mental disorders as those not exposed.

The data collected for this report do not inform either argument further. Preliminary analyses divided the relative risks into two age groups, 15–29 and 30–44 years. Confidence intervals overlapped suggesting no significant difference between the two age groups. Two explanations can be given. First, the numbers in the groups were small limiting the robustness of the estimates. Second, most studies reported estimates of risk for lifetime mental disorders thereby preventing any clear distinction in onset of mental disorders between the two age groups from being made. In the absence of data we have assumed that risk remains constant over the lifespan, as it becomes mere speculation if one tries to estimate the amount risk reduces over time. In order to better inform this decision, data on risk for current disorder are required, broken down into appropriate age groups. Further analyses of the large community and prospective samples are required to address this.

3.6 Reasons and implications for extrapolation of risk factor–disease relationships from one subregion to another

This represents perhaps the biggest threat to validity to the present study. The studies in the risk analysis were overrepresented by samples from Australia, Canada, New Zealand, some European countries and the United States of America. The prevalence of psychiatric disorders varies from country to country and this variability represents the myriad social, economic and cultural factors that interplay with the development of disorder. The degree to which these same factors mediate the relationship between CSA and mental disorder cannot be quantified and is difficult to even speculate on. The answer to this is not likely to be reached in the near future as extensive research is required before sound conclusions can be drawn. The paucity of data and theoretical complexity necessitated the assumption that the relative risks remain constant across subregions.

3.7 Quantitative and qualitative sources of uncertainty

Uncertainty in the current analysis came from several sources. Meta-analysis was used as a method of quantifying the quantitative uncertainty around the final prevalence estimates taking into account sample size or variability between studies, whichever is appropriate given the homogeneity of the estimates being combined.

The methods of extrapolation also introduce uncertainty into the analysis. In particular, the extrapolation conducted to get relative risks for the three levels of exposure will have introduced error. The ratios used for extrapolation came from only five studies and the relationship between the relative risks for each level of exposure may vary according to the population studied.

The external method employed to adjust for confounders introduces three sources of uncertainty. First, the instruments used to measure the confounders may have variable reliability. Second, the 12 studies that contributed to the adjustment factor all measured family dysfunction differently thereby creating another source of uncertainty. Third, the ability of the adjustment factor to be generalized to other studies conducted on different populations can be questioned. It is possible that confounders may mediate the relationship between CSA and outcomes differentially depending upon the population studied. Not enough studies that controlled for confounds were available to see if the adjustment factor varied in any systematic way.

Additionally, while all estimates came from AMR-A, WPR-A and EUR-A not all countries from those subregions were represented. Inter-country variations within subregions are possible and may affect the generalizability of our relative risk estimates. Even greater uncertainty is introduced when estimates from one subregion are extrapolated to another, a decision necessitated by the paucity of data from other subregions.

4. Results and discussion

The discussion of the estimates of the burden in mental disorder outcome that is attributable to CSA will focus on differences between mental disorders, followed by an examination of sex, age and subregional differences. Across the world, CSA contributed to between 4% and 5% of the disability-adjusted life years (DALYs) in males and between 7% and 8% of the DALYs in females for each of depression, alcohol abuse/dependence and drug abuse/dependence. The attributable fractions were higher for panic disorder (7% for males and 13% for females) and higher still for PTSD (21% for males and 33% for females). For suicide attempts the attributable fractions were 6% for males and 11% for females. As discussed above, the confidence intervals around the relative risks of each of the mental disorders overlap and thus the apparent differences may not be real. The same applies to the attributable fractions.

On the whole CSA contributes to a higher percentage of DALYs for females than for males. This difference is driven by the difference in the prevalence of CSA for females and males. There are slight subregional variations in the attributable fractions. In particular, AFR-E and SEAR-D have higher attributable fractions than the other subregions. This is a function of the higher prevalence rates for these subregions. However, data for these subregions came from a few studies that were poor methodologically.

The number of DALYs attributable to CSA varies as this is a function of both the attributable fractions and the amount of burden of disease accounted for by psychiatric disorders in the various subregional groups. One pattern is evident, however: the number of DALYs are greater in the younger age groups and decline in the older age groups. Since risk was assumed to be constant across age, this merely reflected the distribution of DALYs for mental disorders, which impact largely in the younger age groups due to their early onset and chronic nature.

5. Methods for projection of exposure forward

Estimates of avoidable burden, the amount of burden that could be prevented if exposure to CSA was curtailed, are dependent upon the expected prevalence of CSA in the future. As outlined above, we have assumed that CSA will not vary over time. The reviews conducted on cohort effects have provided equivocal evidence (Bagley 1990, 1995; Bagley and Ramsay 1985; Bickerton et al. 1991; Feldman et al. 1991; Fergusson et al. 2000). Three reviews concluded that the prevalence of CSA could increase over time while three reviews also conclude that there is no evidence to support a change in prevalence over time. Analysis of this phenomenon in our own data set also provided no evidence of a change in prevalence over time. Additionally, it is difficult to speculate on what future factors may arise to influence prevalence. Therefore, the estimates of avoidable burden should be the same as those for attributable burden.

Acknowledgements

Steering Committee: Majid Ezzati, David Fergusson, Bronwyn Gould, Beth Kotze, Ron Kessler, Alan Lopez, Michael Lynskey, Judy Martin, Paul Mullen, Anthony Rodgers, Bedirhan Üstün.

Note

1 See preface for an explanation of this term.

References

Agathonos H, Alexandridis A, Fereti I (1992) *A retrospective study of child sexual abuse among Greek university students: implications for prevention and future research*. (Proceedings of the Ninth International Congress on Child Abuse and Neglect.) Chicago, IL.

American School Health Association (1989) *The national student health survey: a report on the health of America's youth*. Third Party Publishing Co., Oakland, CA.

Anderson J, Martin J, Mullen P, Romans S, Herbison P (1993) Prevalence of childhood sexual abuse experiences in a community sample of women. *Journal of the American Academy of Child and Adolescent Psychiatry*, **32**:911–919.

Andrews G, Henderson S, Hall W (2001) Australia's mental health: an overview of the general population survey. *British Journal of Psychiatry*, **178**:145–153.

Arata CM (1999) Repeated sexual victimization and mental disorders in women. *Journal of Child Sexual Abuse*, **7**(3):1–17.

Arroyo JA (1997) Childhood sexual abuse among Hispanic and non-Hispanic white college women. *Hispanic Journal of Behavioral Sciences*, **19**:57–68.

Bagley C (1990) Is the prevalence of child sexual abuse decreasing? Evidence from a random sample of 750 young adult women. *Psychological Reports*, **66**:1037–1038.

Bagley C (1991) The prevalence and mental health sequels of child sexual abuse in a community sample of women aged 18 to 27. *Canadian Journal of Community Mental Health*, **10**:103–116.

Bagley C (1995) Child sexual abuse in Canada: further analysis of the 1983 national survey. In: *Child sexual abuse and mental health in adolescents and adults: British and Canadian perspectives*. Bagley C, ed. Ashgate Publishing, Brookfield, VT.

Bagley C, Bolitho F, Bertrand L (1995) Mental health profiles, suicidal behavior, and community sexual assault in 2112 Canadian adolescents. *Crisis*, **16**: 126–131.

Bagley C, Mallick K (2000) Prediction of sexual, emotional and physical maltreatment and mental health outcomes in a longitudinal cohort of 290 adolescent women. *Child Maltreatment*, **5**:218–226.

Bagley C, Ramsay R (1985) Sexual abuse in childhood: psychosocial outcomes and implications for social work practice. *Journal of Social Work and Human Sexuality*, **4**:33–47.

Bagley C, Wood M, Young L (1994) Victim to abuser: mental health and behavioral sequels of child sexual abuse in a community survey of young adult males. *Child Abuse and Neglect*, **18**:683–697.

Baker AW, Duncan SP (1985) Child sexual abuse: a study of prevalence in Great Britain. *Child Abuse and Neglect*, **9**:457–467.

Baldini G (1996) *Rape and sexual abuse within the Aboriginal communities*. Sexual Assault Referral Centre, Perth, WA.

Banyard VL (1999) Childhood maltreatment and the mental health of low-income women. *American Journal of Orthopsychiatry*, **69**:161–171.

Barthauer LM, Leventhal JM (1999) Prevalence and effects of child sexual abuse in a poor, rural community in El Salvador: a retrospective study of women after 12 years of civil war. *Child Abuse and Neglect*, **23**:1117–1126.

Barton G (1987) Child sexual abuse: an Australian view; the statistics. *IANUA, The Official Journal of the Kindergarten Teachers' Association of Victoria*, **1**:15.

Bayatpour M, Wells RD, Holford S (1992) Physical and sexual abuse as predictors of substance use and suicide among pregnant teenagers. *Journal of Adolescent Health*, **13**:128–132.

Beautrais AL (2000) Risk factors for suicide and attempted suicide among young people. *Australian and New Zealand Journal of Psychiatry*, **34**:420–436.

Bendixen M, Muus KM, Schei B (1994) The impact of child sexual abuse—a study of a random sample of Norwegian students. *Child Abuse and Neglect*, **18**:837–847.

Bensley LS, Van Eenwyk J, Spieker SJ, Schoder J (1999) Self-reported abuse history and adolescent problem behaviors. I. Antisocial and suicidal behaviors. *Journal of Adolescent Health*, **24**:163–172.

Bentovim A, Boston P, van Elburg A (1987) Child sexual abuse—children and families referred to a treatment project and the effects of intervention. *British Medical Journal*, **295**:1453–1457.

Berry C (1997) Attitudes toward child sexual abuse, gender differences and prevalence: a comparative study. *MAI*, **35**:1905.

Bickerton D, Hall R, Williams ALJ (1991) Women's experiences of sexual abuse in childhood. *Public Health*, **105**:447–453.

Bifulco A, Brown GW, Adler Z (1991) Early sexual abuse and clinical depression in adult life. *British Journal of Psychiatry*, **159**:115–122.

Blum RW, Geer L, Hutton L et al. (1988) The Minnesota adolescent health survey: implications for physicians. *Minnesota Medicine*, **71**:143–149.

Blum RW, Harmon B, Harris L, Bergeisen L, Resnick MD (1992) American Indian-Alaska Native youth health. *Journal of the American Medical Association*, **267**:1637–1644.

Bolen RM, Scannapieco M (1999) Prevalence of child sexual abuse: a corrective meta-analysis. *Social Service Review*, **73**:281–313.

Bolstad BR, Zinbarg RE (1997) Sexual victimization, generalized perception of control, and posttraumatic stress disorder symptom severity. *Journal of Anxiety Disorders*, **11**:523–540.

Boney-McCoy S, Finkelhor D (1995) Psychosocial sequelae of violent victimization in a national youth sample. *Journal of Consulting and Clinical Psychology*, **63**:726–736.

Boney-McCoy S, Finkelhor D (1996) Is youth victimization related to trauma symptoms and depression after controlling for prior symptoms and family relationships? A longitudinal, prospective study. *Journal of Consulting and Clinical Psychology*, **64**:1406–1416.

Borowsky IW, Resnick MD, Ireland M, Blum RW (1999) Suicide attempts among American Indian and Alaska Native youth: risk and protective factors. *Archives of Pediatrics and Adolescent Medicine*, **153**:573–580.

Bostwick JM, Pankratz VS (2000) Affective disorders and suicide risk: a reexamination. *American Journal of Psychiatry*, **157**:1925–1932.

Boudewyn AC, Liem JH (1995) Childhood sexual abuse as a precursor to depression and self-destructive behavior in adulthood. *Journal of Traumatic Stress*, **8**:445–459.

Bouhet B, Perard D, Zorman M (1992) De l'importance des abus sexuels en France. In: *Les enfants victimes d'abus sexuels*. Gabel M, ed. Puf, Paris.

Brent DA, Baugher M, Bridge J, Chen T, Chiappetta L (1999) Age- and sex-related risk factors for adolescent suicide. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**:1497–1505.

Briere J, Runtz M (1986) Suicidal thoughts and behaviours in former sexual abuse victims. *Canadian Journal of Behavioural Science*, **18**:413–423.

Briere J, Runtz M (1988) Symptomatology associated with childhood sexual victimization in a nonclinical adult sample. *Child Abuse and Neglect*, **12**:51–59.

Briere J, Runtz M (1990) Differential adult symptomatology associated with three types of child abuse histories. *Child Abuse and Neglect*, **14**:357–364.

Brown GW, Harris TO (1993) Aetiology of anxiety and depressive disorders in an inner-city population: 1—Early adversity. *Psychological Medicine*, **23**: 143–154.

Brown GW, Moran P (1994) Clinical and psychosocial origins of chronic depressive episodes. I: a community survey. *British Journal of Psychiatry*, **165**:447–456.

Brown J, Cohen P, Johnson JG, Smailes EM (1999) Childhood abuse and neglect: specificity of effects on adolescent and young adult depression and suicidality. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**:1490–1496.

Browne A, Finkelhor DA (1986) Impact of child sexual abuse: a review of the research. *Psychological Bulletin*, **99**:66–77.

Bryant SL, Range LM (1995) Suicidality in college women who were sexually and physically abused and physically punished by parents. *Violence and Victims*, **10**:195–201.

Bryant SL, Range LM (1997) Type and severity of child abuse and college students' lifetime suicidality. *Child Abuse and Neglect*, **21**:1169–1176.

Caspi A, Moffitt TE, Newman DL, Silva PA (1996) Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. *Archives of General Psychiatry*, **53**:1033–1039.

Castelino CT (1985) *Child sexual abuse: a retrospective study*. Tate Institute of Social Sciences, Bombay.

Cawson P, Wattam C, Brooker S, Kelly G (2000) *Child maltreatment in the United Kingdom: a study of the prevalence of child abuse and neglect*. National Society for the Prevention of Cruelty to Children (NSPCC), London.

Chandler RK, Jackson JL (1997) Family environment and childhood sexual victimization: a test of the buffering hypothesis. *Journal of Interpersonal Violence*, **12**:3–17.

Chandy JM, Blum RW, Resnick MD (1996) Female adolescents with a history of sexual abuse: risk outcome and protective factors. *Journal of Interpersonal Violence*, **11**:503–518.

Chandy JM, Blum RW, Resnick MD (1997) Sexually abused male adolescents: how vulnerable are they? *Journal of Child Sexual Abuse*, **6**:1–16.

Choquet M, Darves-Bornoz JM, Ledoux S, Manfredi R, Hassler C (1997) Self-reported health and behavioral problems among adolescent victims of rape in France: results of a cross-sectional survey. *Child Abuse and Neglect*, **21**:823–832.

Collings SJ (1991) Childhood sexual abuse in a sample of South African university males: prevalence and risk factors. *South African Journal of Psychology*, **21**:153–158.

Collings SJ (1997) Child sexual abuse in a sample of South African women students: prevalence, characteristics, and long-term effects. *South African Journal of Psychology*, **27**:37–42.

Committee on Sexual Offences Against Children and Youths (1984) *Sexual offences against children. Volume 1*. Minister of Supply and Services, Ottawa.

Conte JR, Schuerman JR (1987) Factors associated with an increased impact of child sexual abuse. *Child Abuse and Neglect*, **11**:201–211.

Cooper H, Hedges LV (1994) *The handbook of research synthesis*. Cooper H, Hedges LV, eds. Russell Sage Foundation, New York.

Coxell A, King M, Mezey G, Gordon D (1999) Lifetime prevalence, characteristics, and associated problems of non-consensual sex in men: cross sectional survey. *British Medical Journal*, **318**:846–850.

Davidson JRT, Hughes D, Blazer DG, George LK (1991) Post-traumatic stress disorder in the community: an epidemiological study. *Psychological Medicine*, **21**:713–721.

deLahunta EA (1996) Personal exposure of faculty and medical students to family violence. *Journal of the American Medical Association*, **275**: 1903–1906.

de Paul J, Milner JS, Mugica P (1995) Childhood maltreatment, childhood social support, and child abuse potential in a Basque sample. *Child Abuse and Neglect*, **19**:907–920.

Descamps MJ, Rothblum ED, Bradford J, Ryan C (2000) Mental health impact of child sexual abuse, rape, intimate partner violence, and hate crimes in the

National Lesbian Health Care Survey. *Journal of Gay and Lesbian Social Services*, **11**:27–55.

Diekstra RF (1992) The prevention of suicidal behaviour: evidence for the efficacy of clinical community-based programs. *International Journal of Mental Health*, **21**:69–87.

Dinwiddie S, Heath AC, Dunne MP et al. (2000) Early sexual abuse and lifetime psychopathology: a co-twin-control study. *Psychological Medicine*, **30**:41–52.

DiVasto PV, Kaufman A, Rosner L et al. (1984) The prevalence of sexually stressful events among females in the general population. *Archives of Sexual Behavior*, **13**:59–67.

Duane EA, Stewart CS, Bridgeland WM (1997) Consequences of childhood sexual abuse for college students. *Journal of College Student Development*, **38**:13–23.

Dudley MJ, Kelk NJ, Florio TM, Howard JP, Waters BG (1998) Suicide among young Australians, 1964–1993: an interstate comparison of metropolitan and rural trends. *Medical Journal of Australia*, **169**:77–80.

Edgardh K, Ormstad K (2000) Prevalence and characteristics of sexual abuse in a national sample of Swedish seventeen-year-old boys and girls. *Acta Paediatrica*, **89**:310–319.

Edwards JJ, Alexander PC (1992) The contribution of family background to the long-term adjustment of women sexually abused as children. *Journal of Interpersonal Violence*, **7**:306–320.

Elal G, Oral G, Atamer A, Sabol E (2000) *A comparative study of child sexual abuse among university students*. (Paper presented at the 16th Annual Meeting of Traumatic Stress Studies). San Antonio, TX.

Erickson PI, Rapkin AJ (1991) Unwanted sexual experiences among middle and high school youth. *Journal of Adolescent Health*, **12**:319–325.

Ernst C, Angst J, Foldenyi M (1993) The Zurich Study. XVII. Sexual abuse in childhood. Frequency and relevance for adult morbidity data of a longitudinal epidemiological study. *European Archives of Psychiatry and Clinical Neuroscience*, **242**:293–300.

Essock-Vitale SM, McGuire MT (1985) Women's lives viewed from an evolutionary perspective. I. Sexual histories, reproductive success, and demographic characteristics of a random sample of American women. *Ethology and Sociobiology*, **6**:137–154.

Feldman W, Feldman E, Goodman JT et al. (1991) Is childhood sexual abuse really increasing in prevalence? An analysis of the evidence. *Pediatrics*, **88**:29–33.

Felitti VJ, Anda RF, Nordenberg D et al. (1998) Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: the adverse childhood experiences (ACE) study. *American Journal of Preventive Medicine*, **14**:245–258.

Fergusson DM, Horwood LJ, Lynskey MT (1996b) Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**:1365–1374.

Fergusson DM, Horwood LJ, Woodward LJ (2000) The stability of child abuse reports: a longitudinal study of the reporting behaviour of young adults. *Psychological Medicine*, **30**:529–544.

Fergusson DM, Lynskey MT, Horwood LJ (1996a) Childhood sexual abuse and psychiatric disorder in young adulthood: I. Prevalence of sexual abuse and factors associated with sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**:1355–1364.

Fergusson DM, Mullen PE (1999) *Childhood sexual abuse: an evidence based perspective*. Sage Publications, Thousand Oaks, CA.

Finkelhor D (1979) *Sexually victimized children*. The Free Press, New York.

Finkelhor D (1994) The international epidemiology of child sexual abuse. *Child Abuse and Neglect*, **18**:409–417.

Finkelhor D, Browne A (1988) Assessing the long-term impact of child sexual abuse: a review and conceptualization. In: *Handbook on sexual abuse of children. Assessment and treatment issues*. Walker LEA, ed. Springer Publishing Company, New York.

Finkelhor D, Dziuba-Leatherman J (1994) Children as victims of violence: a national survey. *Pediatrics*, **94**:413–420.

Finkelhor D, Hotaling G, Lewis IA, Smith C (1990) Sexual abuse in a national survey of adult men and women: prevalence, characteristics, and risk factors. *Child Abuse and Neglect*, **14**:19–28.

Fleming J, Mullen PE, Sibthorpe B, Attewell R, Bammer G (1998) The relationship between childhood sexual abuse and alcohol abuse in women—a case–control study. *Addiction*, **93**:1787–1798.

Fleming J, Mullen PE, Sibthorpe B, Bammer G (1999) The long-term impact of childhood sexual abuse in Australian women. *Child Abuse and Neglect*, **23**:145–159.

Fleming JM (1997) Prevalence of childhood sexual abuse in a community sample of Australian women. *Medical Journal of Australia*, **166**:65–68.

Friedrich E, Klupp N, Akkaya T (1997) Child abuse in Vienna over the last 10 years. *Journal de Médecine Legale Droit Medical*, **40**:337–338.

Fritz GS, Stoll K, Wagner NN (1981) A comparison of males and females who were sexually molested as children. *Journal of Sex and Marital Therapy*, **7**:54–59.

Fromuth ME (1986) The relationship of childhood sexual abuse with later psychological and sexual adjustment in a sample of college women. *Child Abuse and Neglect*, **10**:5–15.

Fromuth ME, Burkhart BR (1989) Long-term psychological correlates of childhood sexual abuse in two samples of college men. *Child Abuse and Neglect*, **13**:533–542.

Fryer GE, Miyoshi TJ (1994) A survival analysis of the revictimization of children: the case of Colorado. *Child Abuse and Neglect*, **18**:1063–1071.

Garabedian MJ (1994) Relationship of child sexual, physical, and psychological abuse to eating disorders and post-traumatic stress disorder in adult women. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, **54**:2465–2465.

Gardner MJ, Altman DG, eds. (1989) *Statistics with confidence: confidence intervals and statistical guidelines*. British Medical Journal Books, London.

Garnefski N, Arends E (1998) Sexual abuse and adolescent maladjustment: differences between male and female victims. *Journal of Adolescence*, **21**:99–107.

Goldman JDG, Padayachi UK (1997) The prevalence and nature of child sexual abuse in Queensland, Australia. *Child Abuse and Neglect*, **21**:489–498.

Goldman RJ, Goldman JDG (1988) The prevalence and nature of child sexual abuse in Australia. *Australian Journal of Sex, Marriage and Family*, **9**:94–106.

Gorcey M, Santiago JM, McCall-Perez F (1986) Psychological consequences for women sexually abused in childhood. *Social Psychiatry*, **21**:129–133.

Gorey KM, Leslie DR (1997) The prevalence of child sexual abuse: integrative review adjustment for potential response and measurement biases. *Child Abuse and Neglect*, **21**:391–398.

Gould DA, Stevens NG, Ward NG, Carlin AS, Sowell HE, Gustafson B (1994) Self-reported childhood abuse in an adult population in a primary care setting. Prevalence, correlates, and associated suicide attempts. *Archives of Family Medicine*, **3**:252–256.

Graham A, Reser J, Scuderi C, Zubrick S, Smith M, Turley B (2000) Suicide: an Australian Psychological Society discussion paper. *Australian Psychologist*, **35**:1–28.

Green AH (1988) Child maltreatment and its victims. A comparison of physical and sexual abuse. *Psychiatric Clinics of North America*, **11**:591–610.

Greenland S (1987) Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews*, **9**:1–30.

Greenwald E, Leitenberg H (1990) Posttraumatic stress disorder in a nonclinical and nonstudent sample of adult women sexually abused as children. *Journal of Interpersonal Violence*, **5**:217–228.

Greenwood CL, Tangalos EG, Maruta T (1990) Prevalence of sexual abuse, physical abuse, and concurrent traumatic life events in a general medical population. *Mayo Clinic Proceedings*, **65**:1067–1071.

Hall LA, Sachs B, Rayens MK, Lutenbacher M (1993) Childhood physical and sexual abuse: their relationship with depressive symptoms in adulthood. *Image—The Journal of Nursing Scholarship*, **25**:317–323.

Halperin DS, Bouvier P, Jaffe PD et al. (1996) Prevalence of child sexual abuse among adolescents in Geneva: results of a cross sectional survey. *British Medical Journal*, **312**:1326–1329.

Harris EC, Barraclough B (1997) Suicide as an outcome for mental disorders. A meta-analysis. *British Journal of Psychiatry*, **170**:205–228.

Harrison PA, Fulkerson JA, Beebe TJ (1997) Multiple substance use among adolescent physical and sexual abuse victims. *Child Abuse and Neglect*, **21**:529–539.

Haugaard JJ, Emery RE (1989) Methodological issues in child sexual abuse research. *Child Abuse and Neglect*, **13**:89–100.

Hawton K, Fagg J (1988) Suicide, and other causes of death, following attempted suicide. *British Journal of Psychiatry*, **152**:359–366.

Hernandez JT (1992) Substance abuse among sexually abused adolescents and their families. *Journal of Adolescent Health*, **13**:658–662.

Hernandez JT, Lodico M, DiClemente RJ (1993) The effects of child abuse and race on risk-taking in male adolescents. *Journal of the National Medical Association*, **85**:593–597.

Hibbard RA, Brack CJ, Rauch S, Orr DP (1988) Abuse, feelings, and health behaviors in a student population. *American Journal of Diseases of Children*, **142**:326–330.

Hibbard RA, Ingersall GM, Orr DP (1990) Behavioral risk, emotional risk, and child abuse among adolescents in a nonclinical setting. *Pediatrics*, **86**:896–901.

Hien D, Bukszpan C (1999) Interpersonal violence in a "normal" low-income control group. *Women and Health*, **29**:1–16.

Higgins DJ, McCabe MP (1994) The relationship of child sexual abuse and family violence to adult adjustment: toward an integrated risk-sequelae model. *Journal of Sex Research*, **31**:255–266.

Hill AB (1965) The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, **58**:295–300.

Hill J, Davis R, Byatt M, Burnside E, Rollinson L, Fear S (2000) Childhood sexual abuse and affective symptoms in women: a general population study. *Psychological Medicine*, **30**:1283–1291.

Interventions for support (2001) *Child sexual abuse: statistics from IFSHA, New Delhi, 1996 to 2001*. Available at http://www.ifsha.org.

Jackson JL, Calhoun KS, Amick AE, Maddever HM, Habif VL (1990) Young adult women who report childhood intrafamilial sexual abuse: subsequent adjustment. *Archives of Sexual Behavior*, **19**:211–221.

Johnson JG, Cohen P, Brown J, Smailes EM, Bernstein DP (1999) Childhood maltreatment increases risk for personality disorders during early adulthood. *Archives of General Psychiatry*, **56**:600–606.

Jumper SA (1995) A meta-analysis of the relationship of child sexual abuse to adult psychological adjustment. *Child Abuse and Neglect*, **19**:715–728.

Kellogg ND, Hoffman TJ (1995) Unwanted and illegal sexual experiences in childhood and adolescence. *Child Abuse and Neglect*, **19**:1457–1468.

Kelly L, Regan L, Burton S (1991) *An exploratory study of the prevalence of sexual abuse in a sample of 16–21 year olds*. Child Abuse Studies Unit, University of North London, London.

Kendall-Tackett KA, Williams LM, Finkelhor D (1993) Impact of sexual abuse on children: a review and synthesis of recent empirical studies. *Psychological Bulletin*, **113**:164–180.

Kendler KS, Bulik CM, Silberg J, Hettema JM, Myers J, Prescott CA (2000) Childhood sexual abuse and adult psychiatric and substance use disorders in women: an epidemiological and cotwin control analysis. *Archives of General Psychiatry*, **57**:953–959.

Kercher GA, McShane M (1984) The prevalence of child sexual abuse victimization in an adult sample of Texas residents. *Child Abuse and Neglect*, **8**:495–501.

Kessler RC, Borges G, Walters EE (1999) Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. *Archives of General Psychiatry*, **56**:617–626.

Kessler RC, Davis CG, Kendler KS (1997) Childhood adversity and adult psychiatric disorder in the US National Comorbidity Survey. *Psychological Medicine*, **27**:1101–1119.

Kessler RC, Magee WJ (1993) Childhood adversities and adult depression: basic patterns of association in a US national survey. *Psychological Medicine*, **23**:679–690.

Kessler RC, McGonagle KA, Zhao S et al. (1994) Lifetime and 12-month prevalence of DSM–III–R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry*, **51**:8–19.

Kilpatrick AC (1986) Some correlates of women's childhood sexual experiences: a retrospective study. *Journal of Sex Research*, **22**:221–242.

Kilpatrick DG, Schnurr PP, Acierno R, Resnick HS, Best CL (2000) Risk factors for adolescent substance abuse and dependence: data from a national sample. *Journal of Consulting and Clinical Psychology*, **68**:19–30.

Krugman S, Mata L, Krugman R (1992) Sexual abuse and corporal punishment during childhood: a pilot retrospective survey of university students in Costa Rica. *Pediatrics*, **90**:157–161.

Lazartigues A, Perard D, Lisandre H, Pailleux T (1989) Sexual abuse. Study of a population of 1000 students. *Neuropsychiatrie de l'enfance et de l'adolescence*, **37**:223–229.

Lenihan GO (1996) *Childhood sexual abuse: gender differences in prevalence, experience, and outcomes*. (Presented at the 104th Annual Convention of the American Psychological Association.) Toronto.

Leth I (2001) *Sexual abuse of children and adolescents: results from a Scandinavian research concerning the extent and character of sexual abuse*. International Planned Parenthood Federation, London.

Levy HB, Markovic J, Chaudhry U, Ahart S, Torres H (1995) Reabuse rates in a sample of children followed for 5 years after discharge from a child abuse inpatient assessment program. *Child Abuse and Neglect*, **19**:1363–1377.

Locke SD (1996) The long-term psychological effects of childhood sexual abuse among black and white female college students (A retrospective analysis of socio-demographic factors). *Dissertation Abstracts International, A: The Humanities and Social Sciences*, **57**:875.

Lodico MA, Gruber E, DiClemente RJ (1996) Childhood sexual abuse and coercive sex among school-based adolescents in a Midwestern State. *Journal of Adolescent Health*, **18**:211–217.

Lopez F, Carpintero E, Hernandez A, Martin MJ (1995) The prevalence and consequences of child sexual abuse in Spain. *Child Abuse and Neglect*, **19**:1039–1050.

Lynskey MT, Fergusson DM (1997) Factors protecting against the development of adjustment difficulties in young adults exposed to childhood sexual abuse. *Child Abuse and Neglect*, **21**:1177–1190.

MacMillan HL, Fleming JE, Trocme N et al. (1997) Prevalence of child physical and sexual abuse in the community. Results from the Ontario Health Supplement. *Journal of the American Medical Association*, **278**:131–135.

Madu SN, Peltzer K (2000) Risk factors and child sexual abuse among secondary school students in the Northern Province (South Africa). *Child Abuse and Neglect*, **24**:259–268.

Market Research Bureau of Ireland (1987) *Child sexual abuse in Dublin (Pilot survey report)*. Market Research Bureau of Ireland, Ltd., Dublin.

Martin G (1996) Reported family dynamics, sexual abuse, and suicidal behaviours in community adolescents. *Archives of Suicide Research*, **2**:183–195.

Martin J, Anderson J, Romans S, Mullen P, O'Shea M (1993) Asking about child sexual abuse: methodological implications of a two stage survey. *Child Abuse and Neglect*, **17**:383–392.

Mazza D, Dennerstein L, Ryan V (1996) Physical, sexual and emotional violence against women: a general practice-based prevalence study. *Medical Journal of Australia*, **164**:14–17.

Meledandri G, Zantedeschi E, Cattaruzza MS, Signorelli C, Fara GM (1996) Child sexual abuse: Italian aspects of the phenomenon. *Annali d'Igiene*, **8**:239–244.

Menick DM, Ngoh F (1998) Child sexual abuse at school: results from an action research study in Cameroon [in French]. *Médecine Tropicale*, **58**:249–252.

Miles GM (2000) Children don't do sex with adults for pleasure: Sri Lankan children's views on sex and sexual exploitation. *Child Abuse and Neglect*, **24**:995–1003.

Moeller TP, Bachmann GA, Moeller JR (1993) The combined effects of physical, sexual, and emotional abuse during childhood: long-term health consequences for women. *Child Abuse and Neglect*, **17**:623–640.

Molnar BE, Buka SL, Kessler RC (2001) Child sexual abuse and subsequent psychopathology: results from the National Comorbidity survey. *American Journal of Public Health*, **91**:753–760.

Molnar BE, Shade SB, Kral AH, Booth RE, Watters JK (1998) Suicidal behavior and sexual–physical abuse among street youth. *Child Abuse and Neglect*, **22**:213–222.

Mullen PE, Fleming J (1998) Long-term effects of child sexual abuse. *Issues in Child Abuse Prevention*, **9**. National Child Protection Clearinghouse, Australia.

Mullen PE, King NJ, Tonge BJ (2000) Child sexual abuse: an overview. *Behaviour Change*, **17**:2–14.

Mullen PE, Martin JL, Anderson JC, Romans SE, Herbison GP (1996) The long-term impact of the physical, emotional, and sexual abuse of children: a community study. *Child Abuse and Neglect*, **20**:7–21.

Mullen PE, Martin JL, Anderson JC, Romans SE, Herbison GP (1993) Childhood sexual abuse and mental health in adult life. *British Journal of Psychiatry*, **163**:721–732.

Mullen PE, Romans-Clarkson SE, Walton VA, Herbison GP (1988) Impact of sexual and physical abuse on women's mental health. *The Lancet*, **1**:841–845.

Mulugeta E, Kassaye M, Berhane Y (1998) Prevalence and outcomes of sexual violence among high school students. *Ethiopian Medical Journal*, **36**:167–174.

Murphy J (1997) *Prevalence of child sexual abuse and consequent victimization in the general population*. (Paper presented at the Third National Family Violence Research Conference, Durham.) University of New Hampshire, Durham, NH.

Nagy S, DiClemente R, Adcock AG (1995) Adverse factors associated with forced sex among southern adolescent girls. *Pediatrics*, **96**:944–946.

Nelson DE, Higginson GK, Grant-Worley JA (1994) Using the youth risk behavior survey to estimate prevalence of sexual abuse among Oregon high school students. *Journal of School Health*, **64**:413–416.

Nelson EC, Heath AC, Madden PA et al. (2002) Association between self-reported childhood sexual abuse and adverse psychosocial outcomes: results from a twin study. *Archive of General Psychiatry*, **59**:139–145.

Neumann DA, Houskamp BM, Pollock VE, Briere J (1996) The long-term sequelae of childhood sexual abuse in women: a meta-analytic review. *Child Maltreatment*, **1**:6–16.

Olsson A, Ellsberg M, Berglund S et al. (2000) Sexual abuse during childhood and adolescence among Nicaraguan men and women: a population-based anonymous survey. *Child Abuse and Neglect*, **24**:1579–1589.

Palmer RL, Bramble D, Metcalfe M, Oppenheimer R, Smith J (1994) Childhood sexual experiences with adults: adult male psychiatric patients and general practice attenders. *British Journal of Psychiatry*, **165**:675–679.

Palmer RL, Coleman L, Chaloner D, Oppenheimer R, Smith J (1993) Childhood sexual experiences with adults. A comparison of reports by women psychiatric patients and general-practice attenders. *British Journal of Psychiatry*, **163**:499–504.

Paolucci EO, Genuis ML, Violato C (2001) A meta-analysis of the published research on the effects of child sexual abuse. *Journal of Psychology*, **135**:17–36.

Paradise JE, Rose L, Sleeper LA, Nathanson M (1994) Behavior, family function, school performance, and predictors of persistent disturbance in sexually abused children. *Pediatrics*, **93**:452–459.

Pederson W, Aas H (1995) Sexual victimization in Norwegian children and adolescents: victims, offenders, assaults. *Scandinavian Journal of Social Medicine*, **23**:173–178.

Peters DK, Range LM (1995) Childhood sexual abuse and current suicidality in college women and men. *Child Abuse and Neglect*, **19**:335–341.

Peters SD (1988) Child sexual abuse and later psychological problems. In: *Lasting effects of child sexual abuse*. Wyatt GE, Powell GL, eds. Sage Publications, Thousand Oaks, CA.

Peters SD, Wyatt GE, Finkelhor D (1986) Prevalence. In: *Sourcebook on child sexual abuse*. Finkelhor DA, ed. Sage Publications, Thousand Oaks, CA.

Plunkett A, Oates RK (1990) Methodological considerations in research on child sexual abuse. *Paediatric and Perinatal Epidemiology*, **4**:351–360.

Plunkett A, O'Toole BI, Swanston H, Oates RK, Shrimpton S, Parkinson P (2001) Suicide risk following child sexual abuse. *Ambulatory Pediatrics*, **1**:262–266.

Polusny MA, Follette VM (1995) Long-term correlates of child sexual abuse: theory and review of the empirical literature. *Applied and Preventive Psychology*, **4**:143–166.

Pothe P, Csemy L, Halfarova H, Bosak V (2000) Sexual abuse of children in the Czech Republic—retrospective study. *Ceska a Slovenska Psychiatrie*, **96**:131–135.

Priest R (1991) An assessment of the prevalence of child sexual victimization and the utilization of mental health services in selected African-American college samples. *Dissertation Abstracts International*, **51**:2641.

Ramos-Lira L, Saldivar-Hernandez G, Medina-Mora ME, Rojas-Guiot E, Villatoro-Velazquez J (1998) Prevalence of sexual abuse in students and its relationship with drug consumption. *Salud Publica de Mexico*, **40**:221–233.

Raupp U, Eggers C (1993) Sexual abuse of children. A regional study of the prevalence and characteristics. *Monatsschrift Kinderheilkunde*, **141**:316–322.

Richter-Appelt H, Tiefensee J (1996) Soziale und familiäre Gegebenheiten bei körperlichen Misshandlungen und sexuellen Missbrauchserfahrungen in der Kindheit aus der Sicht junger Erwachsener. *Psychotherapie, Psychosomatik und Medizinische Psychologie*, **46**:367–378.

Riggs S, Alario AJ, McHorney C (1990) Health risk behaviors and attempted suicide in adolescents who report prior maltreatment. *Journal of Pediatrics*, **116**:815–821.

Rind B (1997) A meta-analytic review of findings from national samples on psychological correlates of child sexual abuse. *The Journal of Sex Research*, **34**:237–255.

Rind B, Tromovitch P, Bauserman R (1998) A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin*, **124**:22–53.

Risberg G, Lundgren E, Westman G (1999) Prevalence of sexualized violence among women. A population-based study in a primary healthcare district. *Scandinavian Journal of Public Health*, **27**:247–253.

Risin LI, Koss MP (1987) The sexual abuse of boys: prevalence and descriptive characteristics of childhood victimizations. *Journal of Interpersonal Violence*, **2**:309–323.

Robin RW, Chester B, Rasmussen JK, Jaranson JM, Goldman D (1997) Prevalence, characteristics, and impact of childhood sexual abuse in a Southwestern American Indian tribe. *Child Abuse and Neglect*, **21**:769–787.

Romans S, Martin J, Mullen P (1997) Childhood sexual abuse and later psychological problems: neither necessary, sufficient nor acting alone. *Criminal Behaviour and Mental Health*, **7**:327–338.

Romero GJ, Wyatt GE, Loeb TB, Carmona JV, Solis BM (1999) The prevalence and circumstances of child sexual abuse among Latina women. *Hispanic Journal of Behavioral Sciences*, **21**:351–365.

Roosa MW, Reyes L, Reinholtz C, Angelini PJ (1998) Measurement of women's child sexual abuse experiences: an empirical demonstration of the impact of choice of measure on estimates of incidence rates and of relationships with pathology. *Journal of Sex Research*, **35**:225–233.

Rothman KJ, Greenland S (1998) *Modern epidemiology.* 2nd edn. Lippincott-Raven Publishers, Philadelphia, PA.

Rönström A (1985) *Sexual abuse of children in Sweden: perspectives on research, interventions and consequences*, Radda Barnen, Stockholm.

Russell DE (1983) The incidence and prevalence of intrafamilial and extrafamilial sexual abuse of female children. *Child Abuse and Neglect,* **7**: 133–146.

Rutter ML (1999) Psychosocial adversity and child psychopathology. *British Journal of Psychiatry*, **174**:480–493.

Sariola H, Uutela A (1994) The prevalence of child sexual abuse in Finland. *Child Abuse and Neglect*, **18**:827–835.

Saunders BE, Kilpatrick DG, Hanson RF, Resnick HS, Walker ME (1999) Prevalence, case characteristics, and long-term psychological correlates of child rape among women: a national survey. *Child Maltreatment*, **4**:187–200.

Saunders BE, Villeponteaux LA, Lipovsky JA, Kilpatrick DG (1992) Child sexual assault as a risk factor for mental disorders among women: a community survey. *Journal of Interpersonal Violence*, **7**:189–204.

Schaaf KK, McCanne TR (1998) Relationship of childhood sexual, physical, and combined sexual and physical abuse to adult victimization and posttraumatic stress disorder. *Child Abuse and Neglect*, **22**:1119–1133.

Schei B (1990) Prevalence of sexual abuse history in a random sample of Norwegian women. *Scandinavian Journal of Social Medicine*, **18**:63–68.

Schein M, Biderman A, Baras M et al. (2000) The prevalence of a history of child sexual abuse among adults visiting family practitioners in Israel. *Child Abuse and Neglect*, **24**:667–675.

Schoetensack K, Elliger T, Gross A, Nissen G (1992) Prevalence of sexual abuse of children in Germany. *Acta Paedopsychiatrica*, **55**:211–216.

Scott KD (1992) Childhood sexual abuse: impact on a community's mental health status. *Child Abuse and Neglect*, **16**:285–295.

Sedney MA, Brooks B (1984) Factors associated with a history of childhood sexual experience in a nonclinical female population. *Journal of the American Academy of Child and Adolescent Psychiatry*, **23**:215–218.

Sharp SJ, Sterne JAC (1997) Meta-analysis. *Stata Technical Bulletin*, **38**:9–14.

Siegel JM, Sorenson SB, Golding JM, Burnam MA, Stein JA (1987) The prevalence of childhood sexual assault. The Los Angeles Epidemiologic Catchment Area Project. *American Journal of Epidemiology*, **126**:1141–1153.

Silverman AB, Reinherz HZ, Giaconia RM (1996) The long-term sequelae of child and adolescent abuse: a longitudinal community study. *Child Abuse and Neglect*, **20**:709–723.

Singh HS, Yiing WW, Nurani HN (1996) Prevalence of childhood sexual abuse among Malaysian paramedical students. *Child Abuse and Neglect*, **20**:487–492.

So-kum Tang C (2000) *Survey study on child sexual abuse among Chinese college students*. End Child Sexual Abuse Foundation, Hong Kong.

Spaccarelli S, Kim S (1995) Resilience criteria and factors associated with resilience in sexually abused girls. *Child Abuse and Neglect*, **19**:1171–1182.

Spak L, Spak F, Allebeck P (1998) Sexual abuse and alcoholism in a female population. *Addiction*, **93**:1365–1373.

Stein JA, Golding JM, Siegel JM, Burnam MA, Sorenson SB (1988) Long-term psychological sequelae of child sexual abuse: the Los Angeles Epidemiologic Catchment Area study. In: *Lasting effects of child sexual abuse*. Wyatt GE, Powell GL, eds. Sage Focus Editions, Vol. 100. Sage Publications, Newbury Park, CA.

Stepakoff S (1998) Effects of sexual victimization on suicidal ideation and behavior in US college women. *Suicide and Life-Threatening Behavior*, **28**:107–126.

Streiner DL (1998) Risky business: making sense of estimates of risk. *Canadian Journal of Psychiatry*, **43**:411–415.

Swanston HY, Tebbutt JS, O'Toole BI, Oates RK (1997) Sexually abused children 5 years after presentation: a case–control study. *Pediatrics*, **100**:600–608.

Tejedor MC, Diaz A, Castillon JJ, Pericay JM (1999) Attempted suicide: repetition and survival—findings of a follow-up study. *Acta Psychiatrica Scandanavica*, **100**:205–211.

Thakkar RR, Gutierrez PM, Kuczen CL, McCanne TR (2000) History of physical and/or sexual abuse and current suicidality in college women. *Child Abuse and Neglect*, **24**:1345–1354.

Urquiza AJ, Capra M (1990) The impact of sexual abuse: initial and long-term effects. In: *The sexually abused male. Vol. I: Prevalence, impact and treatment*. Hunter M, ed. Lexington Books, New York.

Vandewege R, Bruynooghe R, Opdebeeck S (1988) *Ervaringen van vrouwen met fysiek en sexsueel geweld: prevalentie en gevolgen*. Limburgs Universitair Centrum, Diepenbeek.

Vogeltanz ND, Wilsnack SC, Harris TR, Wilsnack RW, Wonderlich SA, Kristjanson AF (1999) Prevalence and risk factors for childhood sexual abuse in women: national survey findings. *Child Abuse and Neglect*, **23**:579–592.

Walch AG, Broadhead WE (1992) Prevalence of lifetime sexual victimization among female patients. *Journal of Family Practice*, **35**:511–516.

Wang J, Liu XZ, Mai QG, Xu JL, Chen LB (1994) The analysis of the frequency of child abuse. *Chinese Social Medicine*, **50**:23–27.

Watkins B, Bentovim A (1992) The sexual abuse of male children and adolescents: a review of current research. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **33**:197–248.

Watts WD, Ellis AM (1993) Sexual abuse and drinking and drug use: implications for prevention. *Journal of Drug Education*, **23**:183–200.

Weiss P, Zverina J (1997) Prevalence of sexual abuse in childhood in the general population: results of a national survey. *Cesko-Slovenska Psychiatrie*, **93**:66–74.

Wellman MM (1993) Child sexual abuse and gender differences: attitudes and prevalence. *Child Abuse and Neglect*, **17**:539–547.

White K, Strange C (1993) Effects of unwanted childhood sexual experiences on psychosocial development of college women. *Journal of College Student Development*, **34**:289–294.

WHO (2001) *WHO multi-country study on women's health and domestic violence*. (Preliminary results.) World Health Organization, Geneva.

Widom CS, White HR (1997) Problem behaviours in abused and neglected children grown up: prevalence and co-occurrence of substance abuse, crime and violence. *Criminal Behaviour and Mental Health*, **7**:287–310.

Wilsnack SC, Vogeltanz ND, Klassen AD, Harris TR (1997) Childhood sexual abuse and women's substance abuse: national survey findings. *Journal of Studies on Alcohol*, **58**:264–271.

Wingood GM, DiClemente RJ (1997) Child sexual abuse, HIV sexual risk, and gender relations of African-American women. *American Journal of Preventive Medicine*, **13**:380–384.

Wise LA, Zierler S, Krieger N, Harlow BL (2001) Adult onset of major depressive disorder in relation to early life violent victimisation: a case–control study. *The Lancet*, **358**:881–887.

Wolf J (1992) *Adult reports of sexual abuse during childhood: results of a statewide telephone survey in Kentucky*. (Paper presented at the 47th Annual Conference of the American Association for Public Opinion Research.) St Petersburg Beach, FL.

Wonderlich SA, Brewerton TD, Jocic Z, Dansky BS, Abbott DW (1997) Relationship of childhood sexual abuse and eating disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **36**:1107–1115.

Wyatt GE (1985) The sexual abuse of Afro-American and white-American women in childhood. *Child Abuse and Neglect*, **9**:507–519.

Wyatt GE, Loeb TB, Solis B, Carmona JV (1999) The prevalence and circumstances of child sexual abuse: changes across a decade. *Child Abuse and Neglect*, **23**:45–60.

Wyatt GE, Newcomb M (1990) Internal and external mediators of women's sexual abuse in childhood. *Journal of Consulting and Clinical Psychology*, **58**:758–767.

Wyatt GE, Powell GJ, eds. (1988) *Lasting effects of child sexual abuse*. Sage Focus Editions, Vol. 100. Sage Publications, Newbury Park, CA.

Wynkoop TF, Capps SC, Priest BJ (1995) Incidence and prevalence of child sexual abuse: a critical review of data collection procedures. *Journal of Child Sexual Abuse*, **4**:49–66.

Yama MF, Tovey SL, Fogas BS, Morris J (1995) The relationship among childhood sexual abuse, parental alcoholism, family environment and suicidal behavior in female college students. *Journal of Child Sexual Abuse*, **4**:79–93.

Zuravin SJ, Fontanella C (1999) The relationship between child sexual abuse and major depression among low-income women: a function of growing up experiences? *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children*, **4**:3–12.

Chapter 24

Distribution of risk factors by poverty

Tony Blakely, Simon Hales, Charlotte Kieft, Nick Wilson and Alistair Woodward

Summary

Socioeconomic position is an important distal risk determinant for many health outcomes. While it was not possible (owing to limitations of data and other factors) to directly map socioeconomic position to the burden of disease, it was considered possible to map some risk factors by absolute poverty, which is one measure of socioeconomic position.

The proportions of the population living on <US$ 1, on US$ 1–2 and on >US$ 2 per day were estimated for each of the 14 subregions[1] using World Bank estimates of poverty by country. The counterfactual scenario was no absolute poverty in the world (no one living on <US$ 2 per day). The prevalences of risk factors for each subregion were obtained from the relevant risk factor chapters in this publication.

The associations of absolute poverty with eight risk factors (child and maternal underweight, unsafe water and sanitation, unsafe sex, indoor air pollution, outdoor air pollution, tobacco use, alcohol use and overweight [women only]) were determined by an indirect method, using asset scores calculated from demographic and health survey (DHS) data and income from living standards measurement surveys (LSMS) data. First, the joint association of the asset score or income variable with the risk factor was determined for each subregion. Second, the percentage estimates of poverty by subregion were overlaid on the ranked asset scores and income variables (e.g. if 20% of people in a subregion were estimated to be living on <US$ 1 per day, then the prevalence of each factor among these poor people was assumed to be that observed for the 20% of people with lowest asset scores). Third, the crude relative risks of each risk factor by level of poverty were estimated based on this overlay. We also undertook selective literature reviews for some risk factors.

Approximately one fifth of the world's population live on <US$ 1 per day and almost half on <US$ 2 per day. Of the 14 subregions, three

(EUR-A, AMR-A and WPR-A) had negligible levels of absolute poverty and were excluded from all subsequent analyses. We estimate that 9% of people in EMR-B were living on <US$2 per day (2% on <US$1 per day), but the estimates for this subregion were based on sparse data. The estimates for the remaining 10 subregions ranged from 18% (3%) for EUR-B to 85% (42%) for SEAR-D and 78% (56%) for AFR-D.

Childhood malnutrition, unimproved water and sanitation and indoor air pollution were strongly associated with absolute poverty. The associations of poverty with one indicator of unsafe sex (unprotected sex with a non-marital partner) and tobacco and alcohol consumption were weaker and variable across subregions. Our analyses and literature reviews were consistent with the proposition that tobacco and alcohol consumption, adverse lipid profiles, hypertension and overweight initially affect the non-poor in developing countries.

If the worldwide prevalence of childhood malnutrition among all children living on <US$2 per day were changed to that of children living on >US$2 per day (i.e. counterfactual scenario 1), 37% of the cases worldwide of underweight would be prevented (assuming a causal relationship). The equivalent percentage reduction from shifting all poor children to exactly US$2 per day (counterfactual scenario 2) was 23%. For inadequate water and sanitation these percentage reductions were 51% and 36%, respectively.

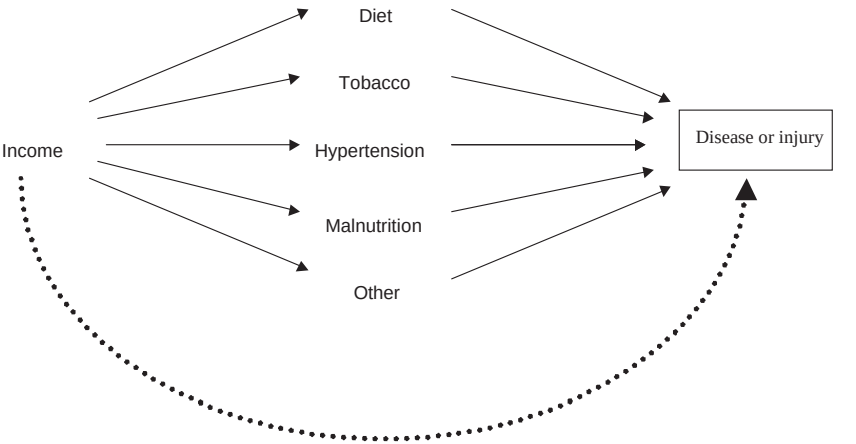
Both poverty and risk factor data were available or were used for only some countries within each subregion, and thus extrapolations had to be made from those countries with data. Second, some subregions had no data at all for some risk factors (e.g. unsafe sex), thus limiting the number of subregions for which we could conduct analyses. Third, all results were based on survey data with their associated errors. Another notable limitation with our analyses was the assumption that the ranking by asset score was comparable to the unobserved ranking by income poverty. Nevertheless, our findings confirm that there are currently severe inequalities in the distribution of childhood malnutrition, inadequate water and sanitation and indoor air pollution by income poverty worldwide, with these risks concentrated among the poorest sectors of society.

1. Introduction

Income may affect health outcomes directly or through exposure to other risk factors (Figure 24.1). In this work, we mapped the crude associations of poverty with some of the health risk factors detailed elsewhere in this book. The risk factors included were childhood underweight, unsafe water and sanitation, unsafe sex, alcohol use, tobacco use, overweight (women only), indoor air pollution and outdoor air pollution.

There were insufficient data at the time to present analyses for the other risk factors. Also, some of the analyses for the risk factors men-

Figure 24.1 Diagrammatic representation of poverty in relation to other risk factors and the global burden of disease



tioned above were more prone to bias because of unavailability of data, data being available or used for only a few countries within a given subregion, or proxy data being used (e.g. expenditure on tobacco and alcohol used as a proxy for actual consumption). Literature reviews were therefore conducted for both (i) the risk factors for which we conducted quantitative analyses but had concerns about the accuracy of our results; and (ii) those remaining risk factors for which there was published literature on the association with poverty or socioeconomic position. These reviews are presented for tobacco use, alcohol use, hypertension/blood pressure, cholesterol level, physical inactivity, overweight, lead and illicit drugs.

This chapter includes neither quantitative nor literature review findings for the following risk factors: child sexual abuse, contaminated injections in health care settings, fruit and vegetable consumption, injuries, micronutrient (vitamin A, iron and zinc) deficiencies, climate change, mental health risk factors and occupational risks.

1.1 Socioeconomic position, risk factors and health status

Socioeconomic position is a multifaceted construct that includes both resource-based and prestige-based measures:

> Resource-based measures refer to material and social resources and assets, including income, wealth and educational credentials; terms used to describe inadequate resources include "poverty" and "deprivation". Prestige-based measures refer to individuals' rank or status in

> a social hierarchy, typically evaluated with reference to people's access to and consumption of goods, services and knowledge, as linked to their occupational prestige, income and educational level (Krieger 2001).

Health status is strongly determined by socioeconomic position. A burgeoning research literature from developed countries demonstrates that all-cause mortality and most causes of death occur at greater rates among groups with lower socioeconomic position (Blakely 2002; Drever and Whitehead 1997; Howden-Chapman and Tobias 2000; Mackenbach et al. 1997; Pamuk et al. 1998; Sorlie et al. 1995). Emerging research in the developing world points to similar associations within countries (Evans et al. 2001; Leon and Walt 2001; Wagstaff 2000; Wagstaff and Watanabe 2001), although data are still limited for individual-level (as opposed to ecological-level) analyses. However, it is unlikely that the relationship of personal socioeconomic position with health is of similar magnitude in all countries. For example, the association of lung cancer with lower socioeconomic position in developed countries is (mainly) due to the greater prevalence of tobacco smoking among groups with a lower socioeconomic position. However, this picture has arisen over time as tobacco consumption has moved from a pattern of behaviour associated with the more advantaged groups to one associated with those of lower socioeconomic positions in the developed world. As the tobacco epidemic moves through the developing world, we would first expect to see higher tobacco consumption among groups with higher socioeconomic position, and correspondingly higher lung cancer rates among groups with higher socioeconomic position. Thus we would expect to see varying associations of socioeconomic position with health status by subregion, and variation over time. This means that caution should be applied when generalizing the association of socioeconomic position with health status from one country to another, and from one time period to another.

1.2 Poverty as a measure of socioeconomic position

The definition of socioeconomic position given above makes it clear that poverty is only one component of the broader construct of socioeconomic position. As with socioeconomic position, poverty too is a multifaceted construct.

> A complex construct, poverty is inherently a normative concept that can be defined—in both absolute and relative terms—in relation to: "need", "standard of living", "limited resources", "lack of basic security", "lack of entitlement", "multiple deprivation", "exclusion", "inequality", "class", "dependency" and "unacceptable hardship" (Krieger 2001).

The United Nations has identified two forms of poverty: "human poverty" and "income poverty" (UNDP 2000). Income poverty refers to

deprivation in a single construct—income. While income poverty is only one dimension of poverty, it is undoubtedly a core element. Human poverty, on the other hand, is characterized by impoverishment in multiple dimensions such as health, knowledge, standard of living and participation in society. The World Bank also accepts this view of poverty, which covers not only material deprivation but also low achievement in health and education (World Bank 2001a).

Sen (1999) views income poverty as one component of "capability deprivation", whereby the latter includes health and educational status that may offset income poverty, lead to income poverty, or be affected by income poverty. Also, the instrumental relationship between income and capability deprivation varies by circumstance. For example, someone in poor health may lose his or her source of income *and* incur extra expenditure.

A World Health Organization (WHO)-sponsored commission has recently underlined the importance at a macroeconomic level of investing in health care and public health services to enhance economic development (Commission on Macroeconomics and Health 2001). Such an approach is in contrast to the assumption that improved health status follows economic development. Thus, at both the individual and the country level, we are forced to think of a bi-directional association between poverty and health.

The advantages of income poverty as the indicator of socioeconomic status (SES) include:

- income poverty is more readily measured in a standardized way across countries and subregions than a more general and multifaceted measure of poverty;
- there has recently been a concerted attempt to measure the prevalence of absolute poverty by country (sponsored by the World Bank) in which the percentages of people in each country living on <US$1 and <US$2 per day have been estimated (Chen and Ravallion 2000; World Bank 2001a);
- while we fully recognize that a broader concept of poverty encompasses health and other capabilities, in mapping health risk factors by poverty we wanted to avoid the endogeneity that would be incurred by directly incorporating health status in our measure of poverty (put another way, we wanted to avoid "self-correlation" whereby we map health risk behaviour by a poverty variable that incorporates health status as one component);
- income poverty might be considered more directly amenable to policy intervention than measures of socioeconomic position incorporating status and prestige; and

- it is not implausible to imagine a world in which people do not live on <US$ 1 or <US$ 2 per day—it would take only a relatively small amount of the income of populations in developed countries to be redistributed to poor people to lift everyone out of absolute poverty.

Nevertheless, there are disadvantages in using absolute poverty as the measure of socioeconomic position. First, the use of the absolute limits of US$ 1 and US$ 2 per day means that the affluent regions of the world (Western Europe, North America and the Western Pacific) are not included in analyses, since only a negligible proportion of the population in these regions live in such depths of absolute income poverty. Poverty still exists in these regions, but given the higher standards of living and average income it is of a more "relative" than "absolute" nature. One solution to this problem is to use relative poverty cut-offs, e.g. 60% of the median income in each country. We did not pursue this alternative, although our methods could be adapted to do so and there are World Bank estimates of relative poverty that are consistent with the absolute poverty estimates used in this chapter (Chen and Ravallion 2000). Second, poverty is only one aspect of socioeconomic position; we could, for example, have mapped health risk factors by educational status. Difficulties with such an approach include data limitations and a lack of standardized measures of education between countries and subregions.

Studies in developing countries were examined in order to give an approximation of the degree of concordance between the different socioeconomic measures, e.g. education, income and occupation. Of the 79 papers retrieved, 29 reported on the prevalence of a risk factor by more than one socioeconomic measure. Table 24.1 shows the papers that included more than one socioeconomic measure and the number that then revealed similar trends between the measures used. Some risk factors and socioeconomic measures have distinctive gender or cultural patterns that may affect these results.

Based on these papers, the distribution of risk factors showed similar trends by education and by income. Indeed, 10 of the 13 papers that used both income and education as independent measures of socioeconomic position showed similar relationships with the risk factor in question during analysis. By contrast, four out of the six papers that used both occupation and income independently as measures of socioeconomic position and only five out of 10 papers with occupation and education showed similar trends.

1.3 Poverty–risk association

In addition to the advantages and disadvantages of our choice of absolute poverty as the measure of socioeconomic position, there are other limitations that must be borne in mind when interpreting the findings of this analysis.

Table 24.1 Overview of papers from the literature review that used more than one measure of socioeconomic position

Socioeconomic variable	*Reference*	*Proportion showing similar association*
Occupation, socioeconomic status	Agarwal et al. (1994)	1/1
Education, occupation, income	al-Mannai et al. (1996); Ekpo et al. (1992); Grol et al. (1997a)	2/3
Education, income	al-Nuaim et al. (1996, 1997); Delpeuch et al. (2000); Kikafunda et al. (1998); Monteiro et al. (1995); Obot (1990); Rossouw et al. (1990); Sakamoto et al. (2001)	6/8[a]
Education, occupation	Chaturvedi et al. (1998); Hodge et al. (1994); Hu and Tsai (2000); Jarallah et al. (1999); Narayan et al. (1996)	2/5
Water supply type, income, education	Chen (1996)	0/1
Education, insurance premium level	Chung et al. (1992)	0/1
Occupation, income	Dhurandhar and Kulkarni (1992); Ge et al. (1994); Naidu and Rao (1994)	2/3[b]
Education, socioeconomic measure	Delpeuch et al. (1994); Martorell et al. (1998)	1/2[c]
Nongovernmental assistance, education, land ownership, occupation, housing, media exposure	Hadi (2000)	
Education, occupation, housing type	Hoa et al. (1995)	0/1
Education, water, toilet type	Li et al. (1999)	1/1
Education, income, land ownership	Radebe et al. (1996)	1/1
Television/radio ownership, water source, toilet type, floor type, fuel type, education	Ricci and Becker (1996)	1/1

[a] In women, not men.

[b] In men, not women.

[c] Not rural.

Causal inference

There are two major problems with determining causality for poverty–risk factor relationships: endogeneity (i.e. the inseparability of poverty and health owing to dynamic, synergistic and bi-directional causal associations, as described above) and confounding (since poverty may be correlated with other determinants of health such as education) (Kaufman and Cooper 1999). We have taken the approach of simply

reporting ("mapping") crude associations of poverty and health risk factors. While this approach is limited by problems of endogeneity and confounding, it also avoids tenuous assumptions about controlling for confounders that are part of the constellation of factors that accompany poverty.

Time-lags

Any changes in health status or health risk factors caused by poverty will take time to occur. Some effects may be relatively rapid, such as an improvement in income that may immediately lead to the ability to buy food and avoid malnutrition. Other effects may take a long time. For example, it may take many years before political decisions utilize the collective wealth of the community to invest in a supply of clean water and adequate sewage disposal.

We do not account for time-lags, so that current distributions of risk factors may be a consequence of poverty levels in the past. This is consistent with the "mapping" nature of our project that determines the crude associations of poverty and risk factors.

Heterogeneity of income within poverty levels

People living on <US$ 1 per day in one part of the world will differ in many ways from those living on <US$ 1 per day in other parts of the world. Some of this variation will be due to the accompanying (and possibly confounding) factors. For example, two countries may differ greatly in average educational level, despite having similar levels of income poverty.

However, there is likely to be variation in the depth of income poverty between countries and subregions. Most of the people living on <US$ 1 per day in one country may be living on an amount just less than US$ 1 per day, whereas in another country the depth of poverty may be much greater. Having determined the prevalence of poverty by cut-offs such as US$ 1 and US$ 2 per day, economists then go on to characterize variation in depth of poverty by measures such as the poverty gap (and the squared poverty gap), which take account of the absolute shortfall (squared) in income beneath the poverty cut-off.

We have not accounted *directly* for varying levels of depth of poverty by subregion. However, by conducting analysis of people *ranked* by value of socioeconomic position (e.g. asset scores calculated at a subregional level), and using both a <US$ 1 and <US$ 2 per day cut-off, we have captured some of this variation.

Heterogeneity of the association of poverty with risk factors

In addition to the varying depths of poverty, there will be a range of reasons for heterogeneity in the association between poverty and health or health risk factors across the world, including:

- macroeconomic policies may mitigate against the expected health consequences of income poverty (for example, welfare system development and infrastructure may reduce the impact on risk factors of being poor);
- health care and public health services, if well organized and effective, should alter socioeconomic gradients in risk factors; and
- the effect of one socioeconomic factor (e.g. education and cultural factors) may interact with another (e.g. poverty), which if their joint distribution between countries/regions varies (as would the joint distribution of education and poverty), would manifest as heterogeneity of effect for either education or poverty considered in isolation.

For example, the cross-national comparison studies by Kunst and Mackenbach and colleagues demonstrate that the all-cause mortality differentials are surprisingly similar across European countries. However, the cause-specific gradients vary notably (Kunst et al. 1998a, 1998b; Mackenbach et al. 1997). For example, there are stronger cancer gradients in the south of Europe, and stronger cardiovascular disease gradients in the north. This variation by cause of death points to underlying variations between the north and south in the distribution of traditional lifestyle risk factors by socioeconomic position.

2. Methods

2.1 Estimating absolute poverty and risk factor levels

The proportion of total population in each country in absolute poverty was taken from the World Development Indicators 2001 CD-ROM (World Bank 2001b). These estimates were made by Chen and Ravallion (2000) using 265 national sample surveys for 83 countries, allowing for purchasing power parity (1993 base). In brief, these estimates of absolute poverty:

- extended previous work by Ravillion et al. (1991) for 1985 that found US$1 a day to be representative of poverty lines found among low-income countries;
- were based on consumption data where possible, although income data (with adjustment) were used when consumption data were missing;
- considered local as well as global estimates of poverty lines in order to arrive at one poverty level applicable to all countries;
- used data from different time periods, but adjusted all dollar estimates to one point in time (1993) and adjusted for purchasing power parity; and

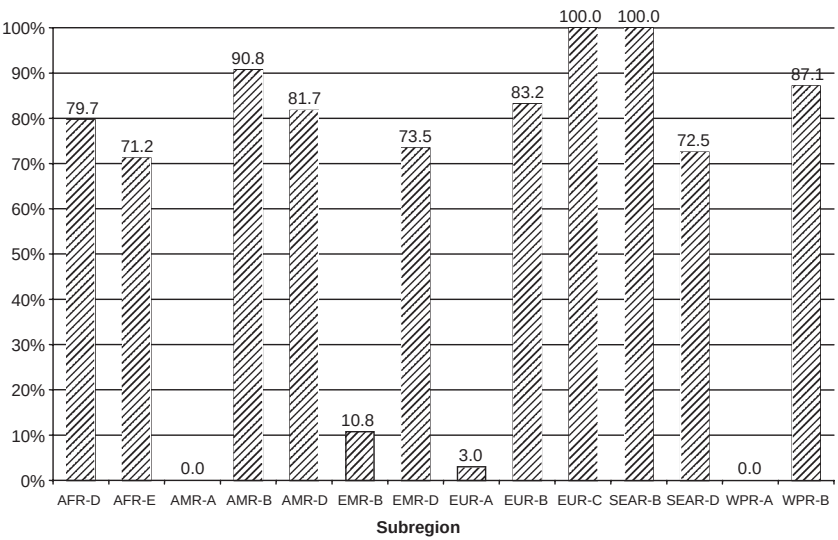
- applied to consumption per person per day (i.e. not household consumption).

Chen and Ravallion set the main poverty line at US$ 1.08 per person per day, using 1993 as the base year. These authors and the World Bank simply referred to this cut-off as US$ 1 per day, and we followed this method. We also included an additional income range of US$ 1–2 per day.

To allow comprehensiveness, we used 1993 estimates of poverty as provided on the World Development Indicators 2001 CD-ROM (World Bank 2001b) and extracted data on the percentage of the population with a daily income of <US $1 and <US$ 2 per day for all available countries. We used the estimates of country populations provided on the CD-ROM to calculate the distribution of the population by subregion within the trichotomous absolute poverty variable.

To estimate the percentage of people living in absolute poverty by subregion, we simply extrapolated the estimated percentages of people living in absolute poverty based on countries with data to the countries without data. Figure 24.2 shows the variation in the percentages of the total population of each subregion living in countries with World Bank estimates

Figure 24.2 Percentages of the population in each subregion living in countries with available World Bank income poverty estimates



Source: World Bank (2001a).

of absolute poverty. For example, in AFR-D the countries for which there were poverty data comprised 79.7% of the total population of that subregion.

The three subregions with <5% of their populations represented by countries with absolute poverty estimates were the three most affluent subregions: AMR-A, EUR-A and WPR-A. As these three subregions contain no (or very few) people living in absolute poverty, they were not included in the analyses.

Ten of the remaining 11 subregions had country-level poverty data covering >70% of the total population in the subregion. The subregion with the least data was EMR-B. For this subregion only 11% of the total population lived in countries for which there were absolute poverty estimates from the World Bank. Accordingly, all estimates presented in this chapter of an association between absolute poverty and risk factors in EMR-B, should be treated with considerable caution.

The country-level (World Bank) and subregional-level estimates (our extrapolations) of the percentages of people in absolute poverty are shown in Table 24.2. We estimated that 9.1% of people in EMR-B were living on <US$2 per day (2.1% on <US$1 per day), although the estimates for this subregion were based on sparse data, as described above. The poverty estimates for the remaining 10 subregions ranged from 17.6% of people living on <US$2 per day (3.0% on <US$1 per day) for EUR-B to 85.2% (42.4%) for SEAR-D and 77.9% (55.5%) for AFR-D.

Table 24.3 presents estimates of absolute poverty for the whole world and for developing countries. The first panel of estimates is that provided by Chen and Ravallion (2000) for the member 2 (i.e. developing, non-donor) countries of the World Bank. These estimates, covering the period 1987–1998, generally indicate that a quarter of the population of the developing world live on <US$1 per day, that over half live on <US$2 per day, and that there was some reduction in the prevalence of absolute poverty between 1993 and 1996. However, the estimates for 1998 were provisional.

The second panel of results in Table 24.3 shows our estimates of absolute poverty for the 11 "non-A" subregions, i.e. the pooled estimates based on the data in Table 24.2. The World Bank data used in Table 24.2 were mostly for the late 1990s. We have therefore placed our estimates under 1996–1998 and they can be seen to be very close to those for a similar time period for the non-donor countries of the World Bank.

The third panel of results in Table 24.3 shows our estimates of absolute poverty for the whole world, assuming nil absolute poverty in EUR-A, AMR-A and WPR-A. We estimated that 20.2% of the world's population live on <US$1 per day and 48.6% live on <US$2 per day, representing 1.2 billion and 2.9 billion people, respectively. These estimates agree well with those of 1.2 and 2.8 billion, respectively, reported in the *World development report 2000/2001* (World Bank 2001a).

Table 24.2 Populations, poverty estimates and availability of survey data by country and subregion[a]

Subregion	Country	Population in 1999 (000s)[b]	Percentage living on <US$ 1 per day[c]	Percentage living on US$ 1–2 per day[c]	Percentage living on >US$ 2 per day[c]	Demographic and health survey (DHS) data	Living standards measurement survey (LSMS) data
AFR-D	Algeria	29 950.0	2.0	13.1	84.9		
	Benin	6 114.0				Yes	
	Burkina Faso	10 995.7	61.2	24.6	14.2	Yes	
	Cameroon	14 691.0				Yes	
	Chad	7 486.0				Yes	
	Comoros	544.0				Yes	
	Gambia	1 251.0	53.7	30.4	16.0		
	Ghana	18 784.5	38.8	35.8	25.4	Yes	Yes
	Guinea	7 251.0				Yes	
	Liberia	3 044.0				Yes	
	Madagascar	15 050.5	63.4	25.7	11.0	Yes	
	Mali	10 583.7	72.8	17.8	9.5	Yes	
	Mauritania	2 598.3	28.6	40.0	31.3		
	Niger	10 495.6	61.4	23.9	14.7	Yes	
	Nigeria	123 896.5	70.2	20.6	9.2	Yes	
	Senegal	9 285.3	26.3	41.5	32.2	Yes	
	Sierra Leone	4 949.3	57.0	17.4	25.5		
	Togo	4 567.0				Yes	
	Total for subregion[d]	*286 129.7*	*55.5*	*22.4*	*22.1*	*NA*	*NA*
AFR-E	Botswana	1 588.1	33.3	28.1	38.7		
	Burundi	6 678.0				Yes	
	Central African Republic	3 539.8	66.6	17.4	16.0	Yes	
	Côte d'Ivoire	15 545.5	12.3	37.1	50.6	Yes	Yes
	Ethiopia	62 782.0	31.3	45.2	23.6	Yes	
	Kenya	29 410.0	26.5	35.8	37.7	Yes	
	Lesotho	2 105.0	43.1	22.6	34.3		
	Malawi	10 788.0				Yes	

	Mozambique	17 299.0	37.9	40.5	21.6	Yes	
	Namibia	1 701.3	34.9	20.9	44.2	Yes	
	Rwanda	8 310.0	35.7	48.8	15.5	Yes	
	South Africa	42 106.2	11.5	24.3	64.2		Yes
	United Republic of Tanzania	32 922.6	19.9	39.8	40.4	Yes	
	Zambia	9 881.2	63.7	23.8	12.6	Yes	
	Zimbabwe	11 903.7	36.0	28.3	35.8	Yes	
	Total for subregion[d]	*330 084.7*	*27.3*	*36.2*	*36.5*	*NA*	*NA*
AMR-B	Brazil	167 966.7	9.0	16.4	74.7	Yes	
	Chile	15 017.8	2.0	16.4	81.6		
	Colombia	41 539.0	11.0	17.7	71.3	Yes	
	Costa Rica	3 589.0	6.9	16.4	76.7		
	Dominican Republic	8 404.4	3.2	12.8	84.0	Yes	
	El Salvador	6 153.9	26.0	28.0	46.0		
	Honduras	6 317.7	40.5	28.3	31.2		
	Jamaica	2 598.0	3.2	22.1	74.8		
	Mexico	96 585.7	12.2	22.6	65.2	Yes	
	Panama	2 811.0	10.3	14.8	74.9		Yes
	Paraguay	5 358.8	19.5	29.8	50.7	Yes	
	Trinidad and Tobago	1 292.8	12.4	26.6	61.0	Yes	
	Uruguay	3 313.0	2.0	4.6	93.4		
	Venezuela	23 707.0	18.7	25.9	55.4		
	Total for subregion[d]	*424 396.0*	*11.0*	*19.1*	*69.8*	*NA*	*NA*
AMR-D	Bolivia	8 138.0	29.4	22.0	48.6	Yes	
	Ecuador	12 412.0	20.2	32.1	47.7	Yes	Yes
	Guatemala	11 088.4	10.0	23.8	66.2		
	Nicaragua	4 919.0				Yes	
	Peru	25 230.0	15.5	25.9	58.6	Yes	
	Total for subregion[d]	*69 897.5*	*17.4*	*26.3*	*56.3*	*NA*	*NA*

continued

Table 24.2 Populations, poverty estimates and availability of survey data by country and subregion[a] (*continued*)

Subregion	Country	Population in 1999 (000s)[b]	Percentage living on <US$ 1 per day[c]	Percentage living on US$ 1–2 per day[c]	Percentage living on >US$ 2 per day[c]	Demographic and health survey (DHS) data	Living standards measurement survey (LSMS) data
EMR-B	Jordan	4 739.9	2.0	5.4	92.6		
	Tunisia	9 456.7	2.0	8.0	90.0	Yes	
	Total for subregion[d]	*136 797.5*	*2.0*	*7.1*	*90.9*	*NA*	*NA*
EMR-D	Egypt	62 654.9	3.1	49.6	47.3	Yes	
	Morocco	28 238.0	2.0	5.5	92.5	Yes	
	Pakistan	134 790.0	31.0	53.7	15.4	Yes	Yes
	Sudan	28 993.0				Yes	
	Yemen	17 047.6	15.7	29.5	54.8	Yes	
	Total for subregion[d]	*348 468.4*	*19.3*	*45.3*	*35.3*	*NA*	*NA*
EUR-B	Armenia	3 808.9	7.8	26.2	66.0		
	Azerbaijan	7 983.0	2.0	7.6	90.4		Yes
	Bulgaria	8 208.0	2.0	19.9	78.1		Yes
	Georgia	5 452.0	2.0	0.0	98.0		
	Krygyzstan	4 865.0				Yes	
	Poland	38 654.0	2.0	0.0	98.0		
	Romania	22 458.0	2.8	24.7	72.5		
	Tajikistan	6 237.0					Yes
	Turkey	64 385.0	2.4	15.7	82.0		
	Turkmenistan	4 779.3	20.9	38.1	41.0		
	Uzbekistan	24 406.3	3.3	23.2	73.5	Yes	
	Total for subregion[d]	*215 275.9*	*3.0*	*14.6*	*82.3*	*NA*	*NA*
EUR-C	Belarus	10 032.0	2.0	0.0	98.0		
	Estonia	1 442.4	2.0	3.2	94.8		
	Hungary	10 068.0	2.0	5.3	92.7		
	Kazakhstan	14 927.0	1.5	13.8	84.7	Yes	Yes

	Latvia	2 431.1	2.0	6.3	91.7		
	Lithuania	3 699.0	2.0	5.8	92.2		
	Republic of Moldova	4 281.0	11.3	27.2	61.6		
	Russian Federation	146 200.0	7.1	18.0	74.9		Yes
	Ukraine	49 950.0	2.9	42.7	54.4		
	Total for subregion[d]	*246 335.9*	*5.4*	*21.3*	*73.3*	*NA*	*NA*
SEAR-B	Indonesia	207 021.6	7.7	47.7	44.7	Yes	
	Sri Lanka	18 985.0	6.6	38.8	54.7		
	Thailand	60 245.8	2.0	26.2	71.9	Yes	
	Total for subregion[d]	*288 750.3*	*6.4*	*42.5*	*51.0*	*NA*	*NA*
SEAR-D	Bangladesh	127 668.8	29.1	48.8	22.2	Yes	
	India	997 515.2	44.2	42.0	13.8	Yes	
	Nepal	23 384.2	37.7	44.8	17.5	Yes	
	Total for subregion[d]	*1 219 491.8*	*42.4*	*42.8*	*14.8*	*NA*	*NA*
WPR-B	China	1 253 595.0	18.5	35.2	46.3		Yes
	Lao People's Democratic Republic	5 096.7	26.3	46.8	26.9		
	Mongolia	2 378.3	13.9	36.0	50.0		
	Philippines	74 259.0				Yes	
	Republic of Korea	46 858.0	2.0	0.0	98.0		
	Total for subregion[d]	*1 520 272.9*	*17.9*	*34.0*	*48.1*	*NA*	*NA*
	Total for all subregions[e]	***5 085 900.6***	***23.7***	***33.4***	***42.9***	***NA***	***NA***

NA Not applicable.

[a] Countries are included in this table if they have any one of the following: poverty estimates, DHS data, LSMS data.

[b] Population counts for country-level from *World development indicators 2001* (World Bank 2001b) and for subregional-level from UN Population Division.

[c] Poverty estimates (World Bank 2001b).

[d] Subregional totals include countries not listed in the table. Poverty estimates are based on those countries with estimates in the subregion—see text for details.

[e] Subregional totals calculated using country-level poverty and population data from World Bank (2001b) and subregional-level population data from UN Population Division.

Table 24.3 Estimates of absolute poverty for the whole world and for developing countries

	Year				
Poverty category	*1987*	*1990*	*1993*	*1996*	*1998*
Member 2 countries of the World Bank					
<US$ 1.08 per day	28.3	29.0	28.2	24.5	24.0
<US$ 2.15 per day	61.0	61.7	60.1	56.1	56.0
Our estimates—11 poorest subregions only					
<US$ 1 per day	—	—	—	23.7	
<US$ 2 per day	—	—	—	57.1	
Our estimates—whole world					
<US$ 1 per day	—	—	—	20.2	
<US$ 2 per day	—	—	—	48.6	

— No data.

The country-level poverty estimates from the World Bank were not disaggregated by sex and age. We assumed a uniform distribution of absolute poverty by sex and age. (We also estimated the crude joint distribution of poverty and risk factors without considering variation by sex and age.) A future improvement to the analyses presented here, therefore, would be to conduct sensitivity analyses on the variation in the poverty distribution by sex and age. This improvement, however, would also require estimating sex- and age-specific joint associations of poverty and each risk factor, for which there are currently very few data.

We obtained the prevalence (or distribution) of risk factors by subregion from the respective chapters in this publication. Wherever possible we used a dichotomous specification of the risk factor for the age and sex strata combined (i.e. crude prevalence estimates by subregion for each risk factor) except for unsafe sex, for which the different sexes were analysed separately. Using a dichotomous variable resulted in the loss of information included in distributions or multiple exposure categories.

2.2 Estimating joint distribution of absolute poverty and risk factor relationship

Table 24.4 represents a stylized version of the data collection sheets required for each subregion. The marginal totals (in bold) show absolute poverty and risk factor prevalence. Our task was to estimate the cell values, a, b, c, d, e and f. If each of the values "a" to "f" represents the absolute percentage of the population of the total subregion in each cell, then:

$$a + b + c + d + e + f = 100\%.$$

Table 24.4 Data collection sheets by subregion: 2 by 3 tables of percentage risk factor by poverty

	Exposed to risk factor	*Not exposed to risk factor*	
<US$ 1 per day	a	b	$\mathbf{M_1}$
US$ 1–2 per day	c	d	$\mathbf{M_2}$
Non-poverty	e	f	$\mathbf{M_3}$
	P	**Q**	**100%**

It was possible to approximate the values for "a" to "f" via the results of the regression analyses and indirect method described above. However, these estimates would have been consistent with the prevalence of the risk factor in the survey data sets used (i.e. DHS or LSMS), but not necessarily consistent with the prevalence of the risk factors determined in the relevant chapters in this publication. Likewise, the fixed poverty percentages came from an external source, and thus all the marginal percentages (M_1, M_2, M_3, P and Q) were fixed. To solve the joint distribution within the table, we used prevalence ratios (referred to as "relative risks" hereafter) estimated using the regression and indirect method, such that:

$$\begin{aligned} RR_1 &= \text{relative risk of risk factor for } <\text{US\$1 compared to } >\text{US\$2} \\ &= (a/M_1)/(e/M_3) \end{aligned} \tag{1}$$

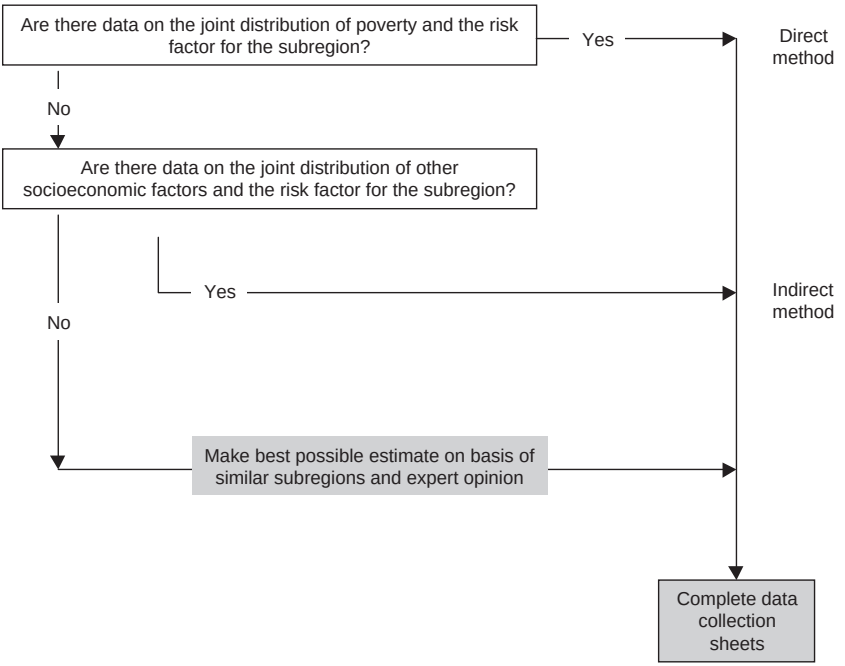
$$\begin{aligned} RR_2 &= \text{relative risk of risk factor for } >\text{US\$ 1 but } <\text{US\$ 2} \\ &\quad \text{compared to } >\text{US\$ 2} \\ &= (c/M_2)/(e/M_3) \end{aligned} \tag{2}$$

Given that c = (P−a−e), then Equation 2 can be rearranged to give:

$$\begin{aligned} RR_2 &= ((P-a-e)/M_2)/(e/M_3) \\ &= \frac{(P-a-e)M_3}{eM_2} \\ \rightarrow eRR_2eM_2 &= PM_3 - aM_3 - eM_3 \\ \rightarrow e(RR_2M_2 + M_3) &= PM_3 - aM_3 \\ \rightarrow e &= \frac{M_3(P-a)}{RR_2M_2 + M_3} \end{aligned} \tag{3}$$

Rearranging Equation 1 to solve for "e", and substituting in Equation 3 allows a solution in terms of "a":

Figure 24.3 Flow diagram of how the distribution of poverty and risk factors might be estimated



$$
\begin{aligned}
aM_3/RR_1M_1 &= M_3\,(P - a)/(RR_2M_2 + M_3) \\
\rightarrow a &= (RR_1M_1P - RR_1M_1a)/RR_2M_2 + M_3 \\
\rightarrow a\,(RR_2M_2 + M_3) &= RR_1M_1P - RR_1M_1a \qquad (1) \\
\rightarrow a\,((RR_2M_2 + M_3) + RR_1M_1) &= RR_1M_1P \\
\rightarrow a &= RR_1M_1P/((RR_2M_2 + M_3) + RR_1M_1)
\end{aligned}
$$

Having calculated "a" the remaining within-cell percentages in Table 24.4 are easily solved.

Figure 24.3 shows a flow diagram of three methods that might be used to estimate this joint distribution.

Ideally the direct method would have been used, but there were no appropriate survey data using a poverty definition equivalent to that of Chen and Ravallion (2000). We therefore had to employ an indirect estimation procedure, using survey data on the joint distribution of other socioeconomic and risk factors. A necessary and key assumption of this indirect method was that the association of risk factors with various

ranked socioeconomic factors was similar. Put another way, we had to assume that one socioeconomic factor could be substituted for another when measuring socioeconomic gradients in health-related behaviour. This was a strong assumption, but one with some support. We did not use the last option of extrapolation from other subregions or expert opinion.

2.3 Data sources and analysis

Table 24.5 provides a summary of the risk factors reported elsewhere in this book, and the extent of inclusion of risk factors in this work.

Indirect estimation of the joint distribution of poverty by risk factor required individual-level data from countries within all subregions. Selected DHS were the main source of these data. We also used LSMS data and survey data specific to China.

Demographic and health survey (DHS) data

DHS are nationally representative household surveys with large sample sizes of about 5000 households (see http://www.measuredhs.com). The DHS programme is conducted by Macro International, with support from the US Agency for International Development. A standard set of questionnaires, similar in all countries, is used to collect data at individual, household and community levels. Four rounds of surveys have been conducted in the past two decades, in which several thousand households were sampled at intervals in some 50 countries across Asia, Africa, the Middle East, Latin America and the area covered by the former Soviet Union.

The core DHS consists of questionnaires for the household and for women specifically. A sample of women aged 15–49 years are interviewed on the following topics: household characteristics, lifetime reproduction, contraceptive knowledge and use, maternity and breastfeeding, immunization, children's health, marriage and fertility preferences, husband's background and the woman's work status. In addition, other country-specific questions may be asked. The DHS contains data for five of the comparative risk assessment (CRA) risk factors:

- underweight in children (weight-for-age *z*-score, DHS variable code hw8);
- improved water and sanitation (DHS variable codes v113 and v116);
- unsafe sex (DHS variable codes v502, 525, 531, 761, 851, 852, 853, 872);
- mother's body mass index (DHS variable code v445); and
- cooking fuel (country-specific variables).

We constructed these variables, in some cases by using several of the original DHS variables. We obtained unit-level data for 53 countries,

Table 24.5 Summary of the CRA risk factors

Risk factor	*Continuous/categorical*	*Survey data available for socioeconomic position by risk factor*	*Inclusion in data analysis*	*Inclusion in systematic literature review*
Air pollution, indoor	Prevalence of solid fuel exposure	Yes—LSMS	Yes	No
Air pollution, urban	PM_{10} ($\mu g/m^3$; 1–10, 11–50, 51–100, ≥100)	Partial	Yes	No
Alcohol use	Amount weighted by pattern of drinking	Yes[a]	Yes	Yes
Child sexual abuse		No	No	No
Contaminated injections in health care settings	Proportion receiving injection contaminated with HIV, HCV, HBV	Insufficient	No	No
Global climate change		Insufficient	No	No
High blood pressure		Insufficient	No	Yes
High cholesterol		Insufficient	No	Yes
Illicit drug use		Insufficient	No	Yes
Lead exposure	Mean (SD), prevalence >0.016 μg/dl	Insufficient	No	Yes
Low fruit and vegetable consumption	Mean (g/day), prevalence could be derived	Insufficient	No	No
Malnutrition: protein–energy (0–4 years)	Prevalence underweight (weight-for-age z-score <–2)	Yes—DHS	Yes	No
Micronutrient deficiencies	(Vitamin A, haemoglobin, iodine)	Insufficient—UNICEF MICS2 not yet available	No	No
Occupational risks (selected)	Prevalence chemical, physical, biological, human factors, safety risk	Insufficient	No	No
Overweight and obesity	Prevalence of overweight and obesity	Yes (women only)—DHS	Yes (women only)	Yes
Physical inactivity		No	No	Yes
Tobacco use	Used indirect method	Yes[a]	Yes	Yes
Unsafe sex	United Nations variables	Yes—DHS	Yes	No
Unsafe water, sanitation and hygiene	Prevalence of five exposure scenarios	Yes—DHS	Yes	No

[a] Alcohol and tobacco data on LSMS were household expenditure data only. Further, it was not possible to "weight" alcohol data by pattern of drinking as in chapter 12.

taking the most recent survey if the country had been surveyed more than once in the period 1986–2000. Of these, an asset score could be calculated for 51 countries, as listed in Table 24.2.

Calculation of asset score

Income was not directly elicited by the standard DHS. However, the World Bank has undertaken extensive work to create asset scores using DHS data (Filmer and Pritchett 1988). Data from the first round of surveys were re-coded to ensure comparability with the subsequent three rounds of surveys. Core questions from all surveys were combined into a single data set. Four categorical variables were constructed as component variables for a DHS asset score: urban–rural status, housing construction material (usually floor material), educational status and availability of electricity. If these variables were missing for a particular country a substitute variable was used where possible, as follows:

- wall material was substituted for floor material in Pakistan;
- number of rooms was substituted for floor material in India; and
- possession of a radio was substituted for electricity supply in Burundi, the Dominican Republic, Liberia and Tunisia.

For two countries (Sri Lanka and Turkey) it was not possible to find suitable substitute variables and asset scores could not be calculated. We considered including more variables in the factor analysis in order to increase the resolution of the asset scores. However, no further variables were available for all countries; thus the addition of further variables would have meant that fewer countries could be included in the subsequent analyses. We considered it important that the asset score be applicable to as many countries as possible. Also, unlike some other asset scores, ours did not include access to safe water to avoid self-correlation in the analyses of water and sanitation by asset score.

Figures 24.4–24.7 show the distribution of DHS observations by the four variables used to calculate the asset scores. (The actual numbers of observations are shown in Appendix A, Table A.1.) There was a predominance of rural people in the DHS data sets (Figure 24.4). There was notable variation between subregions in floor type (Figure 24.5), but there was also some possible variation in coding patterns between countries with regard to the use of the intermediate "rudimentary" floor type. The education variable had four levels, with reasonable spread across respondents and variation in patterns between subregions (Figure 24.6). There was marked variation between subregions in the dichotomous variable for access to electricity.

The distributions of these variables varied markedly between different countries, even within subregions (Appendix A, Table A.1).

A global[2] asset score was calculated using factor analysis. The four variables used to make up the asset scores were coded within the DHS

Figure 24.4 Distribution of DHS asset score variables by country: urban–rural index

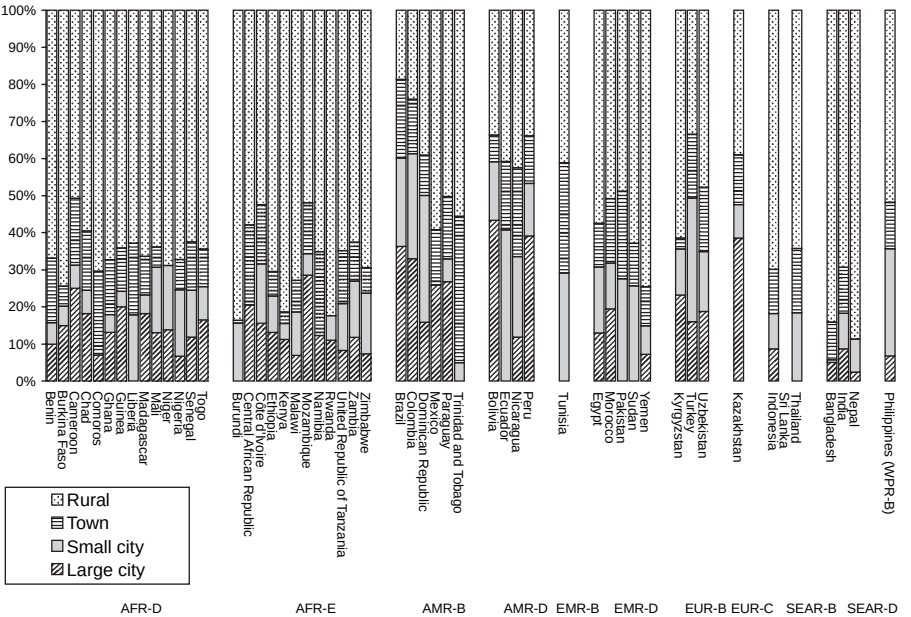


Figure 24.5 Distribution of DHS asset score variables by country: floor type

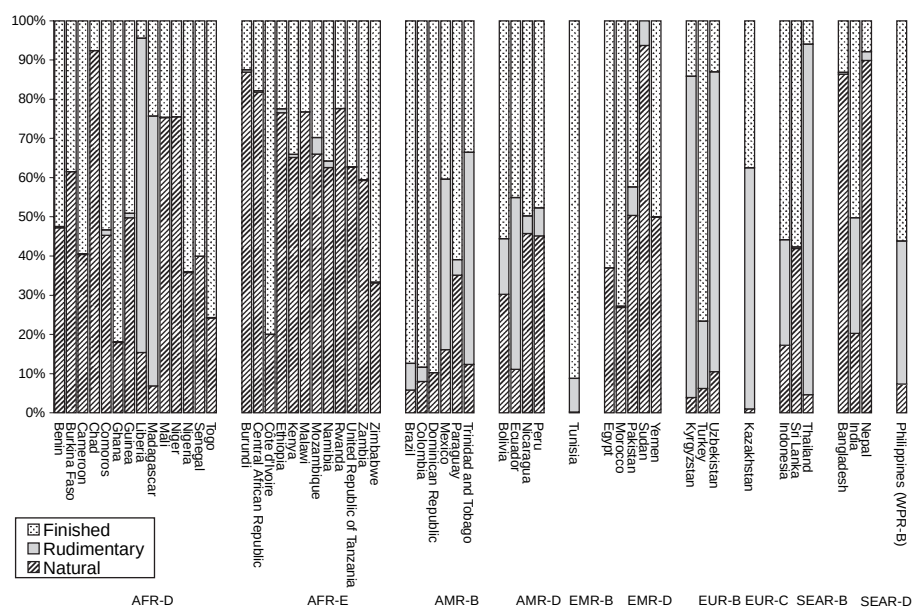


Figure 24.6 Distribution of DHS asset score variables by country: highest education level achieved

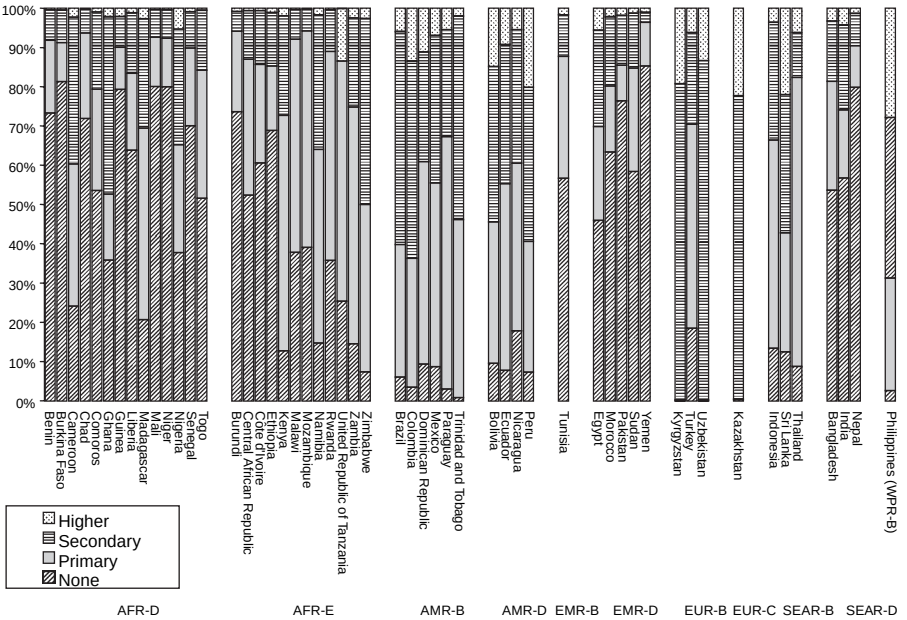


Figure 24.7 Distribution of DHS asset score variables by country: access to electricity

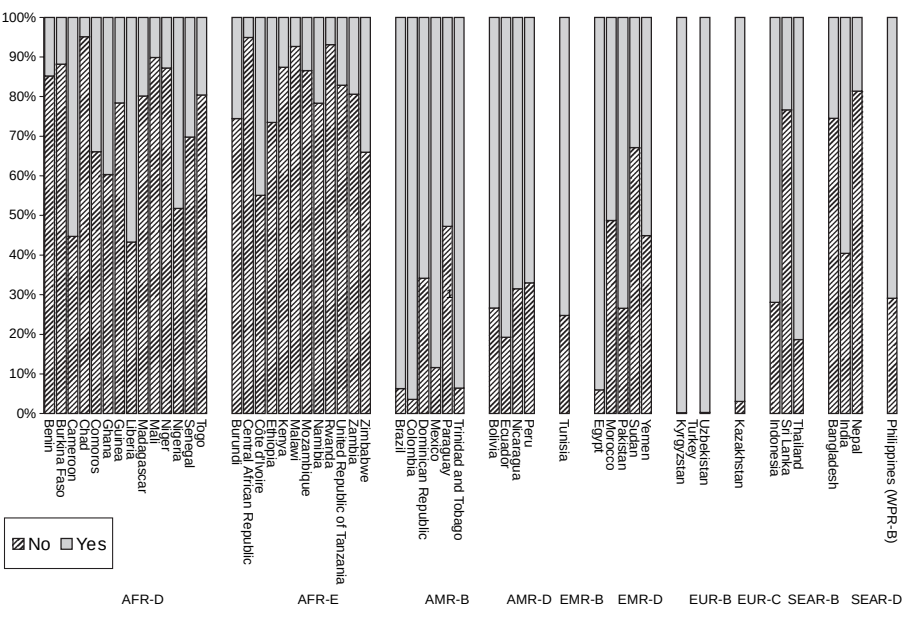


Table 24.6 Principal factors, one factor retained

Factor	*Eigenvalue*	*Difference*	*Proportion*	*Cumulative*
1	1.53227	1.61037	1.3339	1.3339
2	–0.07810	0.05096	–0.0680	1.2659
3	–0.12906	0.04731	–0.1124	1.1535
4	–0.17637		–0.1535	1.0000

Table 24.7 Loadings for factor 1 (from Table 24.6)

Variable	*Factor loading*	*Uniqueness*
Education	0.60216	0.63740
Urban–rural	–0.58536	0.65735
Floor type	0.59736	0.64316
Electricity	0.68570	0.52982

Table 24.8 Scoring coefficients for factor 1 (from Table 24.6)

Variable	*Scoring coefficient*
Education	0.25759
Urban–rural	–0.24446
Floor type	0.25310
Electricity	0.34373

data sets with numeric values corresponding to different categories or variable values:

- urban–rural status: 0 = large city, 1 = small city, 2 = town, 3 = rural area
- floor type: 1 = natural, 2 = rudimentary, 3 = finished
- highest education level achieved: 0 = none, 1 = primary, 2 = secondary
- access to electricity: 0 = no, 1 = yes.

Principal factor analysis was performed in Stata (an interactive data analysis program) for these four variables. The results are shown in Tables 24.6 and 24.7.

Given that only four variables (each with relatively few values) were available for the factor analysis, only 96 discrete asset score values were generated, ranging from –1.12 (more poor) to 1.60 (least poor). The scoring coefficients are shown in Table 24.8.

A full account of the resulting asset score values, and the number of DHS observations with each score by subregion, is given in Appendix A, Table A.2.

Figure 24.8 shows the number of DHS observations by asset score by subregion. The size of the circles represents the number of observations at that asset score value. The number of DHS observations with asset scores is fewer for EMR-B (n = 4184) and EUR-C (n = 4704) than for subregions such as AFR-D (n = 94 811) and SEAR-D (n = 107 044). For AFR-D and AFR-E the asset scores are skewed towards the lower end (as expected) and for AMR-B and AMR-D are skewed to the higher end.

We did not calculate subregional-level asset scores. As the eventual use of asset scores was to rank individuals by socioeconomic position, based on the assumption that this ranking was comparable to that by absolute poverty (also a global construct), we did not consider the construction of subregional-level asset scores justified. Moreover, the range of discrete values of the asset score and sample sizes within some subregions would have introduced instability in asset score indices between subregions that may not have been a function of genuinely varying "asset score constructs" but rather introduced greater random error.

Weighting of the DHS data sets

The objective of the weighting was to ensure that the analyses represented as closely as possible the distribution of the whole population in each subregion. A two-step weighting procedure was thus required.

Figure 24.8 Distribution of asset scores by subregion

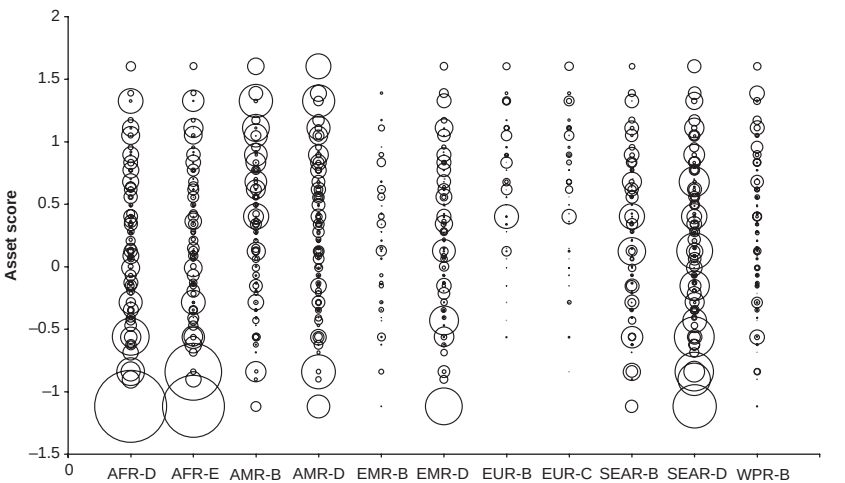


Table 24.9 Examples of weights for DHS analyses

Country in given subregion	*Number of people in DHS sample (n)*	*Number of people in country (N)*	*Weight (N / n)*
Country A	1 000	500 000	500
Country B	5 000	5 000 000	1 000
Country C	2 000	100 000	50
Total of countries sampled	8 000	5 600 000	

1. Owing to the sampling characteristics of the DHS, each observation in the DHS is assigned a country-level weight. The sum of these weights is the same as the total sample size for that particular country, multiplied by 1 million. We simply converted that weight back to a distribution with a mean of 1.0 by dividing by 1 million. We called this component weight w_i, being the individual-level DHS weight for the i^{th} observation in each country.
2. We then weighted the individual DHS observations to be representative of country population by dividing the total population of the j^{th} country (N_j) by $\sum^i w_1$–the country-level weight. (Note that $\sum^i w_i$ is simply the DHS sample size for that particular country.) The product of the individual-level and country-level weights, $w_{ij} = (N_j / \sum^i w_i) \times w_i$, was then assigned to the i^{th} observation in the j^{th} DHS country (within subregions). Note that within each country, the sum of w_{ij} is simply the total population of that country.

To illustrate the necessity for the second weighting step, Table 24.9 provides hypothetical data for a given subregion. If each observation were entered unweighted into any analysis, the results would be skewed towards the pattern of risk factor by poverty present for the individuals "most represented". In Table 24.9, that would be the individuals in country C, where each observation represents only 50 people. To rectify this, the weights in step 2 above (i.e. those in the final column of Table 24.9) were combined with the DHS weights.

The application of the weights for the analysis using the indirect method was not simply a matter of weighting each individual observation. Rather, the asset score was modelled as a cumulative proportion of the population, i.e. a ranking. Thus, we used the weights to make the ranking representative of the total subregional population's ranking, not just the ranking in the available data samples, as described below.

Table 24.10 shows a hypothetical data set for 10 DHS observations (record numbers i, for i = 1–10). Without considering weights, each of these 10 observations represents 10% (or a proportion of 0.1) of all

Table 24.10 Hypothetical example of 10 DHS observations and weighted "midpoint" of cumulative proportion

Record number (i)	*Midpoint of cumulative proportion distribution of i*	*Weight*	*Sum of weights*	*Cumulative proportion of weights*	*Midpoint of cumulative proportion of weights*
1	0.05	200	200	0.040	0.020
2	0.15	700	900	0.180	0.110
3	0.25	900	1 800	0.360	0.270
4	0.35	150	1 950	0.390	0.375
5	0.45	1 900	3 850	0.770	0.580
6	0.55	200	4 050	0.810	0.790
7	0.65	100	4 150	0.830	0.820
8	0.75	400	4 550	0.910	0.870
9	0.85	300	4 850	0.970	0.940
10	0.95	150	5 000	1.000	0.985

observations. The midpoint of the cumulative proportion distribution is therefore 0.05, 0.15, . . . , 0.95, as shown in the second column of Table 24.10. Assume, however, that each of these 10 observations represents varying numbers of the total population, as represented by the weights in third column. The cumulative sum, cumulative proportion and "midpoint" of the cumulative proportion of the weights are shown in the final three columns of Table 24.10. For example, the weight for record 2 was 700, and the sum of weights for records 1 and 2 was 200 + 700 = 900. The sum of all weights was 5000. Therefore, the cumulative proportion of weights up to record 2 was 900/5000 = 0.180. The midpoint of the cumulative proportion distribution was half way between the cumulative proportion of weights for records 1 and 2, i.e. 0.040 + ((0.180–0.040)/2) = 0.110.

Thus, weighting of the midpoints of the cumulative proportion "repositions" the rank value between 0 and 1. This repositioning aims to place each observation at about the position it would have been if the entire population of the subregion had been sampled.

Comparison of asset scores and income in Pakistan

The Pakistan Integrated Household Survey 1991, one of a series of LSMS conducted by the World Bank, included direct estimates of household income as well as variables required to generate an asset score. Thus, we were able to examine the distribution of asset scores by income at the individual level within one country. Note that the available income data

were not consumption data. Also, the asset score variables were not identical to those on the DHS data sets.

We calculated an asset score similar to the method described above. It was not possible, however, to include a variable for urban–rural status, as there were no comparable data readily available from the survey.

Figure 24.9 is a whisker plot of asset scores by rank of household income. Each box-whisker plot is for a given asset score. The boxes show the 25th and 75th percentiles of rank income for a given asset score, and the whiskers show the 5th and 95th percentile ranks of income for that asset score. It is evident that, while there is an association in the expected direction of rank asset score and rank household income, there is also considerable variation of household income ranks within a given asset score.

It is possible that the wide distribution of income data in Figure 24.9 is due to measurement error. Also, incomes tend to be volatile whereas assets are a more stable indicator of long-term income. Therefore, while income and asset scores are not as closely associated as one might wish, this does not render asset scores an unsuitable proxy for income poverty. First, in developed countries we find similar associations between a range of socioeconomic factors and health despite considerable imperfections

Figure 24.9 Box-whisker plot of the normalized rank of estimated household income (NRinc) by normalized rank of household asset score for Pakistan, using LSMS data

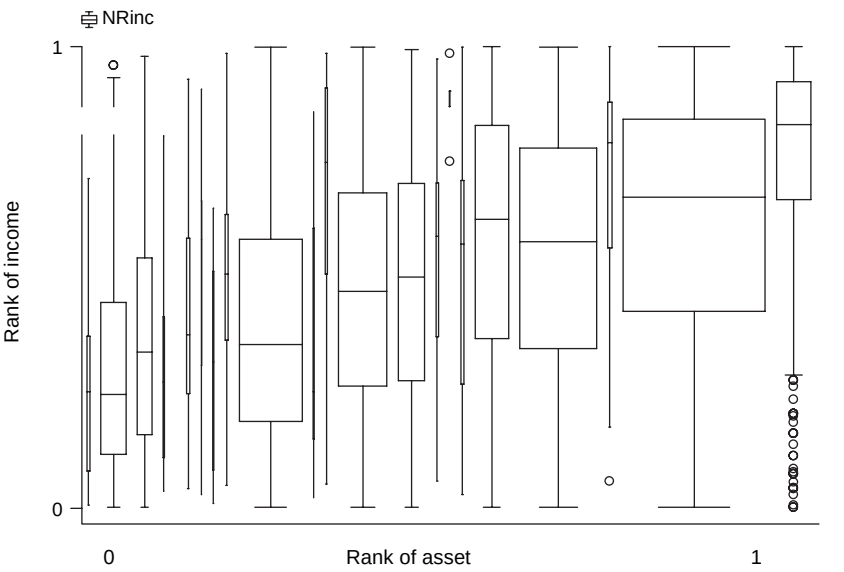


Table 24.11 Comparison of assigning income poverty level in Pakistan using asset score and household income score according to LSMS data

		Allocation to level of poverty by household asset score		
		<US$ 1	*US$ 1–2*	*>US$ 2*
Allocation to level of poverty by household income score	<US$ 1	17%	14%	1%
	US$ 1–2	16%	35%	3%
	>US$ 2	1%	10%	3%

in the correlation of these socioeconomic factors at individual level (Blakely and Pearce 2002; Blakely et al. 2002). Second, we cannot tell, from Figure 24.9, which is the better proxy for income poverty. Third, we treat income poverty here as a categorical variable. It may thus be more reasonable to compare household income score and asset score rank in assigning people to <US$ 1, US$ 1–2 and >US$ 2 per day using the World Bank estimates of 31%, 54% and 15%, respectively (Table 24.11). Fifty-five per cent of households are correctly assigned to an income category using asset score rankings.

Asset scores by gross national product (GNP)

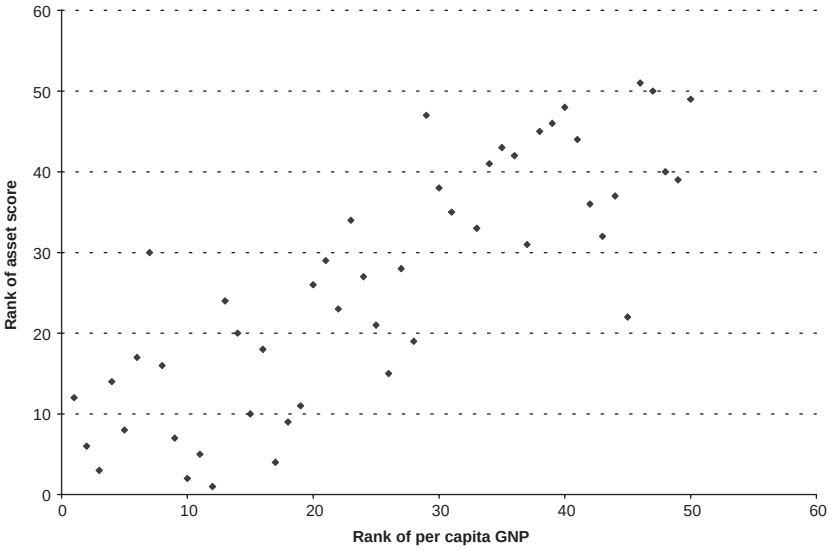
There was a moderately strong correlation between the rank of weighted average asset score for each country and the rank of GNP per capita by country (Spearman rank correlation coefficient 0.68), providing further (albeit highly aggregated) evidence that the asset score is a proxy measure of income (Figure 24.10).

Indirect method using DHS asset scores

We identified three possible pathways for estimating the joint distribution of poverty and risk factors at the subregional level.

1. Pool DHS data by subregion, directly estimate association of asset scores with risk factors and then derive relative risks of each risk factor by poverty categories.
2. Determine the association of asset scores and risk factors for each country, derive relative risks of each risk factor by poverty categories at the country level, then either: (a) pool (e.g. Maentel-Haenzel) the country-level relative risk estimates; or (b) estimate the number of people in each cell of the two by three table (risk factor by three-level poverty variable) for each country using population data, sum the numbers of people in each cell for all countries in the subregion with data, and then calculate subregional-level relative risks. (Note: 2b estimates the "crude" relative risks by poverty at the subregional

Figure 24.10 Scatter plot of rank of a country's per capita GNP against the rank of a country's average asset score



level, whereas 2a estimates the relative risks unconfounded by country.)

A priori we chose method 1 as our task was to estimate subregional-level associations. We were not confident that the data would support country-level calculations for each risk factor, and we were concerned about instability arising from country-level working. However, a strength of method 2 was that it did not assume that a given asset score equated to the same income poverty across countries within subregions.

The principle of the indirect method applied directly to subregional-level data is first illustrated using a simplified example of aggregate (quintile) data before the more detailed non-parametric regression techniques are considered.

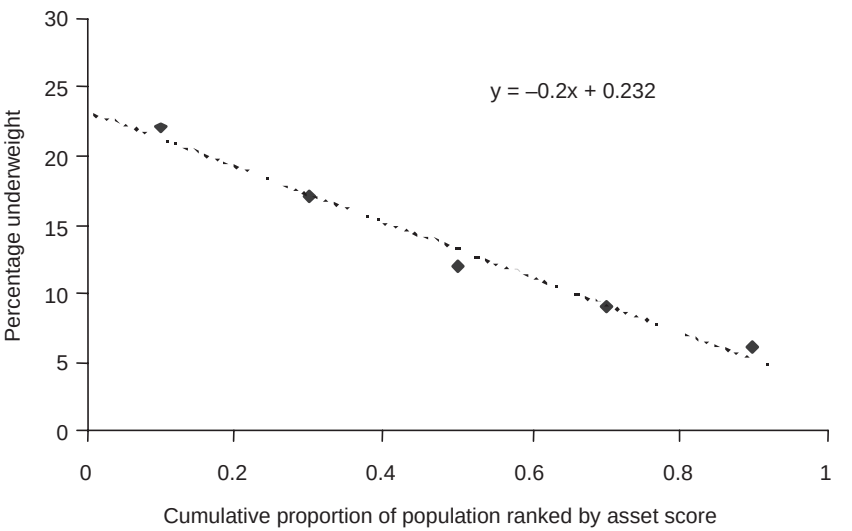
Table 24.12 shows some hypothetical data for the distribution of childhood malnutrition by quintile of asset score at the subregional level. The midpoints of each quintile on a cumulative proportional distribution are shown in the last column.

Figure 24.11 shows a scatter plot of the prevalence of childhood malnutrition by these midpoints. If the plot is nearly linear (as in this case), a regression line is fitted. This method is similar to that used to calculate the relative and slope indices of inequalities (Mackenbach and Kunst 1997).

Table 24.12 Hypothetical distribution of childhood malnutrition by asset score

Quintile of asset score	*Prevalence of childhood malnutrition (%)*	*Midpoint on cumulative proportion distribution*
1 (poorest)	22	0.1
2	17	0.3
3	12	0.5
4	9	0.7
5	6	0.9

Figure 24.11 Percentage childhood malnutrition plotted against cumulative proportion of population ranked by asset score



Assuming that this regression line is similar to the unobserved line for poverty by underweight, it can be used to estimate the prevalence of malnutrition by level of poverty. For example, if in a particular subregion the percentage of children living on <US$1 per day is 20%, the percentage living on US$1–2 per day is 30%, leaving 50% non-poor. The midpoints on the cumulative proportional distribution for these poverty groups would accordingly be 0.10, 0.35 and 0.75. Solving the regression equation (shown in Figure 24.11) for each of these midpoints, we estimate the prevalence of underweight to be 21.2%, 16.2% and 8.2%, respectively. Treating the non-poor as the reference group, these per-

centages equate to relative risks of 2.59 for children living on <US$ 1 per day (i.e. 21.2/8.2) and 1.98 for children living on US$ 1–2 per day (i.e. 16.2/8.2). Also, the relative risk of underweight for children living on US$ 2 per day compared to all those living on >US$ 2 can be estimated as 1.61.

Locally linear kernel regression. While illustrative, this example is an oversimplification for two reasons. First, we needed to allow for non-linear associations. Second, we wanted to use the fact that more than five values of the asset score were captured in the unit-level DHS data. These objectives can be met using a non-parametric regression technique: locally linear kernel regression. This method essentially fits separate linear regression curves for each unit increase in the independent variable(s). In so doing, observations closest to the midpoint are given more weight, and observations further away are given less weight. The outcome is a fitted curve comprising many linear segments that, together, give a smoothed regression line.

Given the large size of the data set, we were able to use the unit-level DHS data for this regression method only for the unsafe sex risk factor. For these unit-level analyses we modelled the logit in Stata. However, for underweight, and for water and sanitation, the observations were too many for efficient computation. Therefore, we calculated the average prevalence of each risk factor within categories of asset score. These aggregate data were then analysed with PROC LOESS in SAS software. The procedure for a single risk factor within one subregion was as follows.

1. The normalized rank of each asset score value was calculated based on the DHS and country-level weights as before (section 2.3).
2. The average prevalence of the risk factor was estimated for each asset score value.
3. The prevalence of the risk factor was regressed on the asset score using locally linear kernel regression.
4. This analysis was weighted according to the number of DHS observations represented by each data point ("weight" option of PROC LOESS in SAS—otherwise the default options were used).

This method was repeated for each subregion, and the whole process repeated for each of the risk factors.

Estimating the joint association of poverty and risk factors. To estimate the joint association of absolute poverty and each risk factor, we determined the average prevalence of the risk factor over the range of asset score rankings equivalent to the poverty estimate.

1. Dummy observations were appended to the data set, representing 0.1% increments of asset score rank between the poorest (rank 0.00) and the least poor (rank 1.00).
2. The results of the locally linear kernel regression were linearly interpolated to provide estimated risk factor prevalences for each of the 0.1% increments of asset score ranking.
3. Average risk factor prevalences were estimated for the appropriate range of the dummy observations, to the nearest 0.1%.

For example, if 20.5% of people in a given subregion lived on <US$1 per day, we estimated the percentage underweight among those living on <US$1 as the average prevalence under the interpolated curve representing asset score ranks of 0.000–0.205. Thus, for each subregion, we estimated the prevalence of the risk factor for:

- people living on <US$ 1 per day;
- people living on (equal to) US$ 1 per day;
- people living on US$ 1–2 per day;
- people living on (equal to) US$ 2 per day; and
- people living on >US$ 2 per day.

Using these prevalence estimates, we then estimated the following relative risks of each risk factor for each subregion:

- <US$ 1 per day compared to >US$ 2 per day;
- US$ 1–2 per day compared to >US$ 2 per day;
- (equal to) US$ 2 per day compared to >US$ 2 per day;
- <US$ 1 per day compared to >US$ 1 per day; and
- (equal to) US$ 1 per day compared to >US$ 1 per day.

Living standards measurement survey (LSMS) data

The LSMS protocol and questionnaires were developed by the World Bank in 1980 and have been carried out in a number of countries. The main purpose of the LSMS is to provide a tool for monitoring living standards and household behaviour in developing countries, with a focus on measuring and understanding poverty. There are three levels of questionnaires within the LSMS: household (household-level income, expenditure and possibly behavioural questions), community (information common to all members of the local community, such as availability of schools, water and electricity) and price (local market prices for basic commodities).

Table 24.13 LSMS countries available for analysis: total population of country and year surveyed

Country	*Subregion*	*Total population (000s)*	*Year surveyed*[a]
Ghana	AFR-D	18 785	1998/1999
Côte d'Ivoire	AFR-E	15 545	1988
South Africa	AFR-E	42 106	1994
Panama	AMR-B	2 811	1997
Ecuador	AMR-D	12 412	1995
Pakistan	EMR-D	134 790	1991
Azerbaijan	EUR-B	7 983	1995
Bulgaria	EUR-B	8 208	1995
Tajikistan	EUR-B	6 237	1999
Kazakhstan	EUR-C	14 927	1996
Russian Federation	EUR-C	146 200	1992

[a] Where the country was surveyed in more than one year only the latest year's data were taken.

The questionnaires are made up of modules (not all of which are used by every country) and are often modified to suit the needs and the situation of the country in question. However, all LSMS surveys are run according to similar protocols, with rigorous quality control measures. They are designed to be quickly and easily administered in the field and rapidly entered into data entry programs.

There are currently about 25 countries listed on the World Bank web site as having LSMS data, some of which have been surveyed more than once. There are different access policies for data from the different countries: some have open data policies, while others allow only restricted access to data. The data set collected varies from country to country, but may include information on income, expenditure, cooking fuel, tobacco use, alcohol consumption, height and weight and many other variables. We were restricted in our analyses to countries for which we could access LSMS data within the given time frame, and to those for which income or expenditure information as well as appropriate risk factor information was available (Table 24.13). Unless otherwise stated, it was assumed that every LSMS survey was representative of the country; hence no survey weights were used in the analyses.

Table 24.14 summarizes the risk factor information available by country. As information on hypertension and physical inactivity were available for only one country each (Bulgaria and the Russian Federation, respectively), these risk factors were not analysed.

For Bulgaria, Ghana, South Africa and Tajikistan there was no question in the surveys asking about individual tobacco or alcohol consumption. The prevalence of this risk factor thus had to be calculated

Table 24.14 LSMS countries and risk factor information available from their data sets

Country	*Subregion*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Hypertension*	*Physical inactivity*
Ghana	AFR-D	Yes	Yes*	Yes*	No	No
Côte d'Ivoire	AFR-E	Yes	No	No	No	No
South Africa	AFR-E	Yes	Yes*	Yes*	No	No
Panama	AMR-B	Yes	Yes	Yes	No	No
Ecuador	AMR-D	Yes	Yes	Yes	No	No
Pakistan	EMR-D	Yes	Yes	No	No	No
Azerbaijan	EUR-B	Yes	Yes**	Yes**	No	No
Bulgaria	EUR-B	Yes	Yes*	Yes*	Yes	No
Tajikistan	EUR-B	Yes	Yes*	Yes*	No	No
Kazakhstan	EUR-C	No	Yes	Yes	No	No
Russian Federation	EUR-C	Yes	Yes	Yes	No	Yes

Yes Variable is available.

No Variable is not available.

* Variable is available only in form of household spending on item.

** Alcohol and tobacco information available only in combined variable.

using household expenditure information for alcohol and for tobacco. A household was classified as using tobacco if any money was recorded as having been spent on cigarettes or tobacco within the expenditure section of the questionnaire. Likewise, a household was classified as using alcohol if any money had been spent on alcohol. While this is a rather crude measure, it should at least give us an approximate idea of the proportion of households with smokers and/or alcohol consumers.

For Azerbaijan the measure was even cruder, as there was only one variable to cover both alcohol and tobacco expenditure. Hence it could not be determined whether any money spent was on tobacco or alcohol or both.

Finally, it must be emphasized that the results based on LSMS data must be treated somewhat cautiously owing to the limited number of countries (and subregions) represented and the proxy nature of some of the risk factor variables.

Calculation of income

LSMS data sets included information on income. A pre-generated aggregate per capita income variable was used if one was provided with the data set; otherwise an income variable of this type was generated using the available data. All income variables were equivalized by dividing by the square root of the number of members of the household, if this had

not already been performed within the generated variable. For Ecuador, only aggregated per capita expenditure was used.

As these income variables were generated in different ways, and to allow for purchasing power parities between countries, each country's income data were treated separately and ranked on a continuous scale from 0 to 1 for further analysis. For example, the household with the median LSMS income in a given country was assigned a ranked income value of 0.5.

Indirect method using LSMS income data

Since we were concerned about purchasing power parities between subregions, we did not pool the data by subregion to estimate the association of poverty with risk factors. Rather, we estimated associations at country level and aggregated these to subregions. Unlike the DHS data, the LSMS analyses were unweighted. Also, the smaller data sets allowed regression analyses on unit-level data (rather than using aggregated data). The associations between income and the various risk factors were determined by categorizing each risk factor into a dichotomous variable and then using the indirect method at the country level to estimate relative risks by poverty (see section 2.3).

Having obtained country-level joint distributions of the total population for poverty by the risk factor, we then aggregated these to estimate the association at the subregional level. Table 24.15 shows, as an example, the distribution of indoor air pollution by countries within subregion EUR-B.

Thus, the estimated subregional prevalence of indoor air pollution in EUR-B for those living on <US$ 1 per day is $0.36 \times 64\% + 0.37 \times 20\% + 0.28 \times 74\% = 0.51$. The final subregional estimates are based only on those countries for which we had data (as for the DHS analyses).

Table 24.15 Distribution of indoor air pollution by countries in EUR-B for which data were available

Country	*Population (000s)*	*Proportion of population*	*Prevalence of indoor air pollution by poverty category (%)*		
			<US$ 1/day	*US$ 1–2/day*	*>US$ 2/day*
Azerbaijan	7 983	0.36	64	62	43
Bulgaria	8 208	0.37	20	22	17
Tajikistan	6 237	0.28	74	76	74
Total	22 428	1.00			

Data for China

We used data from the 1993 China Health and Nutrition Survey (CHNS). The information on this survey that follows is adapted from the web site of the Carolina Population Center at the University of North Carolina at Chapel Hill (http://www.cpc.unc.edu/china, accessed 23 June 2003).

The study population is drawn from the provinces of Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning and Shandong. This sample is diverse, with variation in a wide range of socioeconomic factors (income, employment, education, modernization) and other related health, nutritional and demographic measures.

A multistage, random cluster process was used to draw the sample surveyed in each of the provinces. Counties in the nine provinces were stratified by income (low, middle, high) and a weighted sampling scheme was used to randomly select four counties in each province. In addition, the provincial capital and a lower-income city were selected. Villages and townships within the counties, and urban and suburban neighbourhoods within the cities, were selected randomly. The 190 primary sampling units consisted of 32 urban neighbourhoods, 30 suburban neighbourhoods, 32 towns and 96 villages. In 1989–1993 there were 190 primary sampling units and about 3800 households in the overall survey, covering some 16 000 individuals.

All household members in 1993 provided individual data on dietary intake, body fat distribution, blood pressure, medical history and health-related behaviour (e.g. smoking, beverage consumption, medication, key chronic diseases). The following risk factor variables were available.

- *Underweight* (n = 702 children)

Weight-for-age z-scores, calculated using the EpiNut program, ranged from –5.08 to 9.98 (mean –0.313). Using a cut-off z-score of ≤–2.0, 12.4% of children were malnourished.

- *Water and sanitation* (n = 3422)

Data on how households obtained water and type of toilet facilities were used to assign 63.9% of households as having "improved water and sanitation".

- *Indoor air pollution* (n = 3412)

Data on fuel used for cooking was used to assign 81.4% of households as using "smoky" fuels (coal, wood, sticks, straw and charcoal).

- *Tobacco* (n = 8617)

A total of 31.9% of subjects were assigned as smokers.

- *Alcohol* (n = 8659)

A total of 34.5% of subjects were assigned as alcohol drinkers.

- *Body weight* ($n = 4465$)

Calculation of income

Questions on income and time allocation look for any possible activity that each person might have engaged in during the previous year, both inside and outside the formal market. Full income from market and non-market activities can be imputed. The variable we used for income was the aggregated deflated total per capita income variable. We considered this a better option than creating a new equivalized household income variable. Also, it was not exactly clear how many people there were in each household to allow for equivalization. Some exploratory analyses (not reported) suggested that the results for an equivalized household income variable were very similar to the deflated per capita income variable. Estimated deflated per capita household income was used to generate income rankings.

The association between income and the various risk factors was determined by categorizing each risk factor into a dichotomous variable. The indirect method, using locally weighted sum of squares regression in the same manner as described below for DHS asset scores and LSMS income data, was then used to estimate the relationship between the normalized rank of income and the dichotomous risk factor variable. The income-poverty cut-offs were then superimposed to estimate the prevalence of each risk factor within each income-poverty category. A total of 18.5% of the Chinese population were assumed to be living on <US$ 1 per day, 35.2% on US$ 1–2 per day and 46.3% on >US$ 2 per day.

Other methods adopted for specific risk factors

Ambient air pollution

Asset scores were used to estimate income distribution in urban and rural populations. For this purpose, the urban–rural variable in the DHS data was reclassified as follows: 0 or 1 = urban; 2 or 3 = rural.

We assumed that the distribution of asset scores in the DHS data was representative of the distribution in the relevant subregion. The approximate income distribution of urban and rural populations was estimated based on subregional asset-score cut-offs as before. Because the estimates of urban and rural income were approximate, we used only two income categories: <US$ 1 per day and >US$ 1 per day.

We assumed that rural populations (proportions c and d in Table 24.16) were uniformly exposed to a particulate level of 5 $\mu g/m^3$ in all subregions. We further assumed that, within subregions, the exposure of urban populations (proportions a and b in Table 24.16) was independent of income. Average urban air pollution exposures by subregion were obtained from chapter 17. The exposure estimates were weighted accord-

Table 24.16 Proportion of population within urban/rural category

	Income	
	<US$1 per day	*>US$1 per day*
Urban	a	b
Rural	c	d

ing to the proportions in Table 24.16. For example, the average exposure in populations living on <US$1 per day = (a x exposure in urban areas) + (c x exposure in rural areas).

Body weight

The only body weight (body mass index) data we had available were for women in the DHS data set with children aged <5 years. Thus our analyses were restricted to this group.

Systematic literature reviews

We conducted eight literature reviews focusing on the developing world for tobacco, alcohol, illicit drugs, hypertension, cholesterol, body weight, physical inactivity and lead (see section 4). We also conducted some less rigorous literature reviews for malnutrition, water and sanitation and unsafe sex (see Appendix C).

An initial literature search was conducted using Medline. Papers were limited to those published in English between 1990 and 2001. Four search domains were specified, as shown in Table 24.17. Search domains 1–3 were common to all, while domain 4 was risk-factor-specific.

There was substantial heterogeneity among the retrieved papers with regard to:

- subregions included;
- socioeconomic factor(s) used;
- measurement of the risk factor (e.g. not many of the protein malnutrition papers used weight-for-age [underweight]; they used rather height-for-age [stunting] or weight-for-height [wasting]);
- time period (a particular problem if the prevalence of poverty or the patterning of risk factor [e.g. tobacco] by socioeconomic position was rapidly changing during the 1980s and 1990s);
- study population sampling (e.g. many studies were conducted among poor communities, not as random population samples);
- study population demographics (e.g. limited to certain age groups); and

Table 24.17 Search domains and Medline search strategy

Search domain	*Medline search strategy*
1. Socioeconomic position or poverty	Poverty.af[a] or socioeconomic:.af or income:.af MeSH headings Socioeconomic factors (exploded to include subheadings: career mobility, poverty, poverty areas, social class, social mobility)
2. Survey or quantitative data	Prevalence.af or survey:.af or review.af MeSH headings Data collection, health surveys, health status indicators, nutrition surveys, diet surveys, population surveillance, sentinel surveillance, nutrition assessment, questionnaires, records (exploded), registries, vital statistics, interviews and meta-analysis
3. Developing countries	Developing countr:.af or third-world.af MeSH headings Africa (exploded), Caribbean region (exploded), Central America (exploded), Latin America (exploded), South America (exploded), Asia (exploded), Pacific Islands, Melanesia (exploded), Micronesia (exploded) or Polynesia (exploded)
4. Risk factor	Where possible, we used the same Medline search strategy as that used for other chapters of this publication

[a] The suffix ".af" in the Medline search strategy refers to "all fields".

- reporting of results (e.g. only odds ratios reported with no prevalence data, only multivariate results reported, or only *P* values reported with no effect estimate).

For the more rigorous literature reviews, additional searches were performed as follows.

The Cochrane Library was searched for relevant systematic reviews. While Cochrane reviews have been shown to be of higher quality and to be less biased on average than other systematic reviews, they nevertheless have limitations. Other systematic reviews were searched for in the Database of Abstracts of Reviews of Effectiveness and by the recommended methods for review identification (Glanville and Lefebvre 2000).

Medline searches were conducted for the period 1980 to February 2002. The key search terms for the topics included: smoking, tobacco, alcohol, illicit, cannabis, marijuana, opiate, amphetamine, cocaine, inhalant, hypertension, blood pressure, cholesterol, obesity, body weight, inactivity, exercise, lead and poisoning, and blood lead. The key search terms for socioeconomic status included: poverty, inequality and socioeconomic factors. Where more detailed searches were undertaken, additional search terms included: social class, education, occupation and income.

Other references were identified from the reference lists of key review articles and from a WHO key informant.

3. Quantitative results

3.1 Underweight

Underweight in children was defined as a weight-for-age *z*-score <–2, using the NCHS or Harvard reference populations (see chapter 2). For the developing world, excluding WPR-B, underweight information was ascertained using DHS data sets. For WPR-B, for which we had no applicable DHS data set, underweight prevalence information was obtained using the CHNS data set.

As shown in Table 24.18, there were reasonable numbers of children available for analysis in all subregions except EUR-C and WPR-B. Thus, results for those two subregions should be treated with caution.

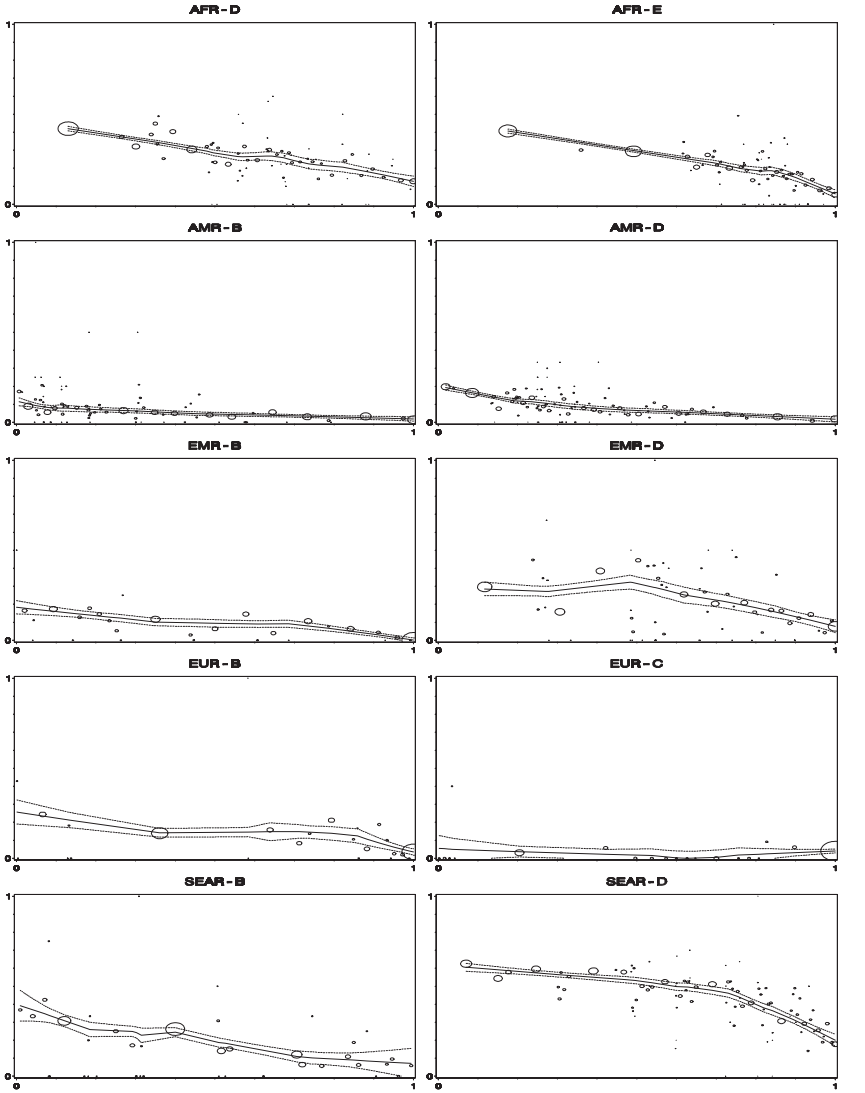
Locally linear kernel regression smooth plots ("Loess plots") of underweight according to weighted asset score ranking at the subregional level are shown in Figure 24.12. Each sub-figure plots the proportion of malnourished children (y-axis ranging from 0 to 1) by asset score rank for the subregion (x-axis ranging from 0 [poorest] to 1 [least poor in subregion]). The centre point of each plotted circle depicts the prevalence of underweight children at that given asset score rank, and the size of each circle is proportional to the number of observations. The middle line is the fitted curve and the upper and lower lines are the 95% confidence bands. (More weight is given to those observations with larger numbers as indicated by a larger circle size.)

From Figure 24.12 it is clear that children living in households with low asset scores (and by inference also living in absolute income poverty) were substantially more likely to be underweight in all subregions. This is consistent with expectation and the findings of Wagstaff and

Table 24.18 Sample sizes for child underweight by subregion

	Low weight-for-age (z-score <–2)		*Weight-for-age (z-score >–2)*	
Subregion	*N*	*Percentage*	*N*	*Percentage*
AFR-D	10 142	34	19 723	66
AFR-E	9 417	29	23 551	71
AMR-B	683	6	11 333	94
AMR-D	1 900	10	17 499	90
EMR-B	175	10	1 498	90
EMR-D	1 707	24	5 391	76
EUR-B	238	14	1 467	86
EUR-C	13	3	427	97
SEAR-B	365	21	1 333	79
SEAR-D	17 323	49	17 905	51
WPR-B	321	14.2	1 943	85.8

Figure 24.12 Loess plots of the prevalence of underweight children (y-axis) by normalized rank of asset score (scale 0–1, x-axis)



Watanabe (2001). Other than the overall differences in prevalence of underweight by subregion, the only other notable difference between subregions was a prevalence of underweight that did not vary greatly among those with lowest ranked asset scores for EMR-D and SEAR-D, but then dropped off more rapidly among those with higher ranked asset scores (i.e. a "shoulder effect"). The EMR-D analyses were based on only two countries—Morocco and Pakistan. With nearly five times the population of Morocco, Pakistan dominates the empirical estimates for EMR-D. An inspection of the Pakistan-only Loess plot (not shown) demonstrates a shoulder effect similar to that for EMR-D in Figure 24.12, although not as marked. Morocco, with only 2% of its population living on <US$ 1 per day, had much lower levels of underweight children than Pakistan, further exaggerating the apparent shoulder effect. Using only the Pakistan data, we would have estimated relative risks of 2.3 and 1.8 for those living on <US$ 1 per day and on US$ 1–2 per day, respectively, compared to >US$ 2 per day. The overall relative risk estimates for EMR-D were both 1.7 (Table 24.19). The sensitivity analysis presented below also suggests that the results for EMR-D set out in Table 24.19 should be treated with considerable caution.

The SEAR-D analyses were based on three countries—Bangladesh, India and Nepal. India makes up about 87% of the population of these three countries combined, and therefore dominates the SEAR-D estimates. The Loess plot for India (not shown) also demonstrates a shoulder effect. Some 86% of the Indian population lives on <US$ 2 per day (44% on <US$ 1 per day). It appears that the steeply declining levels of underweight at the right-hand end of the SEAR-D plot in Figure 24.12 are due to the non-poor (and often financially quite well-off) Indian population living on >US$ 2 per day.

Inspection of the remaining country-level Loess plots (not shown) did not disclose any other countries with a distinctly different pattern of underweight by asset score rank. Consequently, we conclude that the remaining subregional-level plots in Figure 24.12 provide an approximate summary of the country-level associations.

Based on the smoothed curves in Figure 24.12 and a subregional-level analysis of asset scores by underweight, we estimated the relative risks of underweight by level of absolute poverty, as shown in the final column of Table 24.19. For example, we estimated that children in AFR-D living on <US$ 1 per day were 2.3 times more likely to be underweight than children living on >US$ 2 per day. Likewise, children living on US$ 1–2 per day were estimated to be 1.4 times more likely to be underweight, and children living on exactly US$ 2 per day were estimated to be 1.2 times more likely to be underweight. Using these first two relative risk estimates and the fixed marginal prevalences of absolute poverty and underweight (shown in italics to the right of the two by three tables), we calculated the values for each of the cells within the two by three tables. Finally, the middle column of Table 24.19 simply presents the prevalence

Table 24.19 Cell prevalence of weight-for-age by poverty, prevalence of underweight by poverty, and relative risk of underweight by poverty, by subregion, for children aged 0–5 years for a US$ 2 per day cut-off

Subregion	*Weight-for-age*			*Prevalence of underweight by poverty (%)*		*Relative risk of underweight*	
	<–2 SD	*>–2 SD*					
AFR-D	<US$ 1/day	22.5	33.0	*55.5*	<US$ 1/day	40.5	<US$ 1/day 2.3
	US$ 1–2/day	5.8	16.6	*22.4*	US$ 1–2/day	25.7	US$ 1–2/day 1.4
	>US$ 2/day	4.0	18.1	*22.1*	>US$ 2/day	17.9	US$ 2/day 1.2
	32.2	*67.8*	***100.0***				>US$ 2/day 1
AFR-E	<US$ 1/day	12.6	14.7	27.3	<US$ 1/day	46.1	<US$ 1/day 2.6
	US$ 1–2/day	11.9	24.3	*36.2*	US$ 1–2/day	32.8	US$ 1–2/day 1.8
	>US$ 2/day	6.5	30.0	*36.5*	>US$ 2/day	17.9	US$ 2/day 1.4
	31.0	*69.0*	***100.0***				>US$ 2/day 1
AMR-B	<US$ 1/day	1.0	10.0	11.0	<US$ 1/day	9.3	<US$ 1/day 2.4
	US$ 1–2/day	1.3	17.8	19.1	US$ 1–2/day	7.0	US$ 1–2/day 1.8
	>US$ 2/day	2.6	67.2	69.8	>US$ 2/day	3.8	US$ 2/day 1.6
	5.0	*95.0*	***100.0***				>US$ 2/day 1
AMR-D	<US$ 1/day	4.5	12.9	*17.4*	<US$ 1/day	25.8	<US$ 1/day 3.7
	US$ 1–2/day	3.9	22.4	*26.3*	US$ 1–2/day	15.0	US$ 1–2/day 2.1
	>US$ 2/day	4.0	52.3	*56.3*	>US$ 2/day	7.0	US$ 2/day 1.6
	12.4	*87.6*	***100.0***				>US$ 2/day 1
EMR-B	<US$ 1/day	0.3	1.7	*2.0*	<US$ 1/day	15.3	<US$ 1/day 2.1
	US$ 1–2/day	1.0	6.1	*7.1*	US$ 1–2/day	14.3	US$ 1–2/day 1.9
	>US$ 2/day	6.8	84.1	*90.9*	>US$ 2/day	7.5	US$ 2/day 1.8
	8.1	*91.9*	***100.0***				>US$ 2/day 1
EMR-D[a]	<US$ 1/day	5.7	13.6	*19.3*	<US$ 1/day	29.6	<US$ 1/day 1.7
	US$ 1–2/day	13.4	31.9	*45.3*	US$ 1–2/day	29.5	US$ 1–2/day 1.7
	>US$ 2/day	6.0	29.3	*35.3*	>US$ 2/day	17.0	US$ 2/day 1.5
	25.1	*74.9*	***100.0***				>US$ 2/day 1
EUR-B	<US$ 1/day	0.4	2.6	*3.0*	<US$ 1/day	13.1	<US$ 1/day 1.9
	US$ 1–2/day	1.6	13.0	*14.6*	US$ 1–2/day	10.8	US$ 1–2/day 1.6
	>US$ 2/day	5.6	76.7	*82.3*	>US$ 2/day	6.8	US$ 2/day 1.3
	7.6	*92.4*	***100.0***				>US$ 2/day 1
EUR-C	<US$ 1/day	0.3	5.1	*5.4*	<US$ 1/day	4.8	<US$ 1/day 2.4
	US$ 1–2/day	0.9	20.4	*21.3*	US$ 1–2/day	4.2	US$ 1–2/day 2.1
	>US$ 2/day	1.5	71.8	*73.3*	>US$ 2/day	2.0	US$ 2/day 1.8
	2.6	*97.4*	***100.0***				>US$ 2/day 1
SEAR-B	<US$ 1/day	3.2	3.2	*6.4*	<US$ 1/day	50.4	<US$ 1/day 3.3
	US$ 1–2/day	14.7	27.8	*42.5*	US$ 1–2/day	34.5	US$ 1–2/day 2.2
	>US$ 2/day	7.9	43.1	*51.0*	>US$ 2/day	15.5	US$ 2/day 1.7
	25.8	*74.2*	***100.0***				>US$ 2/day 1
SEAR-D	<US$ 1/day	22.9	19.5	*42.4*	<US$ 1/day	54.1	<US$ 1/day 2.1
	US$ 1–2/day	19.1	23.7	*42.8*	US$ 1–2/day	44.7	US$ 1–2/day 1.7
	>US$ 2/day	3.9	10.9	*14.8*	>US$ 2/day	26.1	US$ 2/day 1.3
	45.9	*54.1*	***100.0***				>US$ 2/day 1

Table 24.19 Cell prevalence of weight-for-age by poverty, prevalence of underweight by poverty, and relative risk of underweight by poverty, by subregion, for children aged 0–5 years for a US$ 2 per day cut-off (*continued*)

	Weight-for-age				*Prevalence of underweight by poverty (%)*		*Relative risk of underweight*	
Subregion		*<–2 SD*	*>–2 SD*					
WPR-B	<US$ 1/day	2.7	15.2	*17.9*	<US$ 1/day	15.1	<US$ 1/day	1.1
	US$ 1–2/day	6.6	27.4	*34.0*	US$ 1–2/day	19.4	US$ 1–2/day	1.4
	>US$ 2/day	6.7	41.4	*48.1*	>US$ 2/day	13.9	US$ 2/day	1.3
		16.0	*84.0*	***100.0***			>US$ 2/day	1
Total[b]	<US$ 1/day	10.6	15.5	*26.1*	<US$ 1/day	40.5	<US$ 1/day	3.1
	US$ 1–2/day	10.4	23.5	*33.9*	US$ 1–2/day	30.6	US$ 1–2/day	2.4
	>US$ 2/day	5.2	34.8	*40.0*	>US$ 2/day	13.0	US$ 2/day	—
		26.1	*73.9*	***100.0***			>US$ 2/day	1

[a] The results for EMR-D should be interpreted with caution.

[b] Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

of underweight children within strata (or "rows") of poverty level. For example, the prevalence of underweight children among those living on <US$ 1 per day in AFR-D was 22.5/55.5 = 40.5%.

The results in Table 24.19 demonstrate a reasonably consistent association of underweight with absolute poverty across subregions. Aggregating the 11 subregions to give a total developing world summary, we estimated that children living on <US$ 1 per day were 3.1 times more likely to be malnourished than children living on >US$ 2 per day. Table 24.20 shows the calculations for a cut-off of US$ 1 per day.

Sensitivity analysis

An alternative way of estimating the association of poverty and DHS risk factors for subregions was to conduct analyses *for each country* (i.e. to apply the indirect method at the country level to estimate the joint distribution of income poverty and risk factors), then aggregate up to subregions. The advantage of this method is that it allows for instances in which a given asset score does not equate to a similar income poverty level in different countries. The main disadvantages are that for many risk factors country-level estimates of the association of asset score and risk factors are unstable, and that countries without a World Bank estimate of income poverty could not be included. In terms of the latter, we had to exclude six countries from AFR-D (Benin, Cameroon, Chad, the Comoros, Guinea and Togo), Malawi from AFR-E, Nicaragua from AMR-D and Kyrgyzstan from EUR-B from country-level analyses. (See section 2 for discussion on the relative advantages of the country-level and subregional-level alternatives.) The country-level estimates, together

Table 24.20 Cell prevalence of weight-for-age by poverty, prevalence of underweight by poverty, and relative risk of underweight by poverty, by subregion, for children aged 0–5 years for a US$ 1 per day cut-off

	Weight-for-age			*Prevalence of underweight by poverty (%)*		*Relative risk of underweight*	
Subregion		*<–2 SD*	*>–2 SD*				
AFR-D	<US$ 1/day	22.2	33.3	*55.5*	<US$ 1/day	39.9	<US$ 1/day 1.8
	>US$ 1/day	10.0	34.5	*44.5*	>US$ 1/day	22.6	US$ 1/day 1.3
		32.2	*67.8*	***100.0***			>US$ 1/day 1
AFR-E	<US$ 1/day	12.3	15.0	*27.3*	<US$ 1/day	44.9	<US$ 1/day 1.7
	>US$ 1/day	18.7	54.0	*72.7*	>US$ 1/day	25.8	US$ 1/day 1.5
		31.0	*69.0*	***100.0***			>US$ 1/day 1
AMR-B	<US$ 1/day	1.0	10.0	*11.0*	<US$ 1/day	9.3	<US$ 1/day 2.1
	>US$ 1/day	4.0	85.0	*89.0*	>US$ 1/day	4.5	US$ 1/day 1.7
		5.0	*95.0*	***100.0***			>US$ 1/day 1
AMR-D	<US$ 1/day	4.5	12.9	*17.4*	<US$ 1/day	25.6	<US$ 1/day 2.7
	>US$ 1/day	7.9	74.7	*82.6*	>US$ 1/day	9.6	US$ 1/day 2.0
		12.4	*87.6*	***100.0***			>US$ 1/day 1
EMR-B	<US$ 1/day	0.3	1.7	*2.0*	<US$ 1/day	15.2	<US$ 1/day 1.9
	>$1/day	7.8	90.2	*98.0*	>$1/day	8.0	US$ 1/day 1.9
		8.1	*91.9*	***100.0***			>US$ 1/day 1
EMR-D[a]	<US$ 1/day	5.7	13.6	*19.3*	<US$ 1/day	29.5	<US$ 1/day 1.2
	>US$ 1/day	19.4	61.3	*80.7*	>US$ 1/day	24.0	US$ 1/day 1.2
		25.1	*74.9*	***100.0***			>US$ 1/day 1
EUR-B	<US$ 1/day	0.4	2.6	*3.0*	<US$ 1/day	12.5	<US$ 1/day 1.7
	>US$ 1/day	7.2	89.8	*97.0*	>US$ 1/day	7.4	US$ 1/day 1.6
		7.6	*92.4*	***100.0***			>US$ 1/day 1
EUR-C	<US$ 1/day	0.3	5.1	*5.4*	<US$ 1/day	5.1	<US$ 1/day 2.1
	>US$ 1/day	2.3	92.3	*94.6*	>US$ 1/day	2.5	US$ 1/day 2.1
		2.6	*97.4*	***100.0***			>US$ 1/day 1
SEAR-B	<US$ 1/day	3.2	3.2	*6.4*	<US$ 1/day	50.2	<US$ 1/day 2.1
	>US$ 1/day	22.6	71.0	*93.6*	>US$ 1/day	24.1	US$ 1/day 2.0
		25.8	*74.2*	***100.0***			>US$ 1/day 1
SEAR-D	<US$ 1/day	22.9	19.5	*42.4*	<US$ 1/day	54.1	<US$ 1/day 1.4
	>US$ 1/day	23.0	34.6	*57.6*	>US$ 1/day	39.9	US$ 1/day 1.3
		45.9	*54.1*	***100.0***			>US$ 1/day 1
WPR-B	<US$ 1/day	2.7	15.2	*17.9*	<US$ 1/day	15.0	<US$ 1/day 0.9
	>US$ 1/day	13.3	68.8	*82.1*	>US$ 1/day	16.2	US$ 1/day 1.0
		16.0	*84.0*	***100.0***			>US$ 1/day 1
Total[b]	<US$ 1/day	10.5	15.6	*26.1*	<US$ 1/day	40.3	<US$ 1/day 1.9
	>US$ 1/day	15.6	58.3	*73.9*	>US$ 1/day	21.1	US$ 1/day —
		26.1	*73.9*	***100.0***			>US$ 1/day 1

[a] The results for EMR-D should be interpreted with caution.

[b] Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

Table 24.21 Country- and subregional-level relative risk estimates

Subregion	*<US$ 1 cf. >US$ 2*	*US$ 1–2 cf. >US$ 2*
AFR-D	1.3 (1.4)	1.7 (2.3)
AFR-E	1.8 (1.8)	2.3 (2.6)
AMR-B	3.0 (2.4)	2.1 (1.8)
AMR-D	4.2 (3.7)	2.6 (2.1)
EMR-D	3.6 (1.7)	2.8 (1.7)
EUR-B	2.0 (1.9)	1.7 (1.6)
EUR-C	2.3 (2.4)	2.1 (1.9)
SEAR-D	2.0 (2.1)	1.6 (1.7)

with the subregional-level estimates from Table 24.20 (in parentheses), are shown in Table 24.21.

The country and subregional-level estimates vary notably for AFR-D, but any conclusion as to which method is more valid is difficult owing to the absence of six countries from the country-level sensitivity analysis. The other notable disagreement was for EMR-D (Morocco and Pakistan only). Closer scrutiny revealed a low estimate of absolute poverty in Morocco compared to Pakistan (7.5% and 85% <US$2 per day, respectively), *yet* the asset scores in Morocco and Pakistan largely overlapped, *and* the prevalence of underweight children was much less in Morocco than in Pakistan. Correspondingly, the subregional-level relative risk estimates of 1.7 and 1.7 should be treated with some caution. More generally, the subregional-level estimates for EMR-D for water and sanitation, and overweight should also be treated with caution. For other subregions there was insufficient reason to prefer the country-level to the subregional-level estimates presented in sections 3.1, 3.2, 3.3 and 3.9 using DHS data. However, the findings for EMR-D illustrate the importance of scale. Relationships between income, asset score and risk factors may vary, not only between countries but also within national populations.

3.2 Unsafe water and poor sanitation

Water and sanitation availability was defined as a two-level categorical variable:

- no water supply and/or no sanitation (categories[3] Va, Vb and VI in Prüss et al. 2001); and
- improved water and improved sanitation (category IV in Prüss et al. 2001).

Table 24.22 shows the actual samples sizes for the water and sanitation analyses. CHNS data were used in WPR-B (see Appendix B).

Table 24.22 Sample sizes for water and sanitation by subregion

	No water and/or no sanitation		*Improved water and sanitation*	
Subregion	n	*Percentage*	n	*Percentage*
AFR-D	49 323	55	39 968	45
AFR-E	43 815	53	38 781	47
AMR-B	6 554	14	40 314	86
AMR-D	16 422	32	35 647	68
EMR-B	1 154	28	3 030	72
EMR-D	13 217	33	26 650	67
EUR-B	826	10	7 336	90
EUR-C	74	2	4 472	98
SEAR-B	9 346	74	3 339	26
SEAR-D	67 293	64	38 647	36
WPR-B	2185	64	1237	36

Smoothed scatter-plots of water and sanitation according to household asset score rankings showed increasing availability of water and sanitation with increasing asset score, despite considerable variation in the shape of the curves (Figure 24.13). Conversely, households living in absolute poverty were more likely to lack water and/or sanitation.

Table 24.23 presents the relative risks derived from the smoothed plots in Figure 24.13 and the two by three tables. Note that the data in Table 24.23 are expressed for *un*improved water *and* sanitation, i.e. 1 – category IV, where category IV is according to Prüss et al. (2001). The strength of the association was even stronger than that for childhood underweight, with relative risks for people living on <US$ 1 per day compared to people living on >US$ 2 per day ranging from 2.0 to 15.1. Given this wide range of relative risks, the aggregated global results should be treated with caution. Full tabular results for a US$ 1 per day cut-off are presented in Table 24.24.

3.3 Unsafe sex

To derive estimates of unsafe sex, we estimated the prevalence of sex with a non-marital partner (variable UN1 on DHS data) and, of those having sex with a non-marital partner, the proportion using a condom (variable UN2 on DHS data).

UN1 Higher-risk sex in the last year	All who had sex in the last year (denominator)	Sex with non-marital partner in the last year (numerator)
UN2 Condom use at latest higher-risk sex	All who had higher-risk sex in the last year and were asked about condom use at the latest time (denominator)	People who used condom at latest higher-risk sex (numerator)

Figure 24.13 Loess plots of the prevalence of improved water and sanitation (y-axis) by normalized rank of asset score (x-axis)

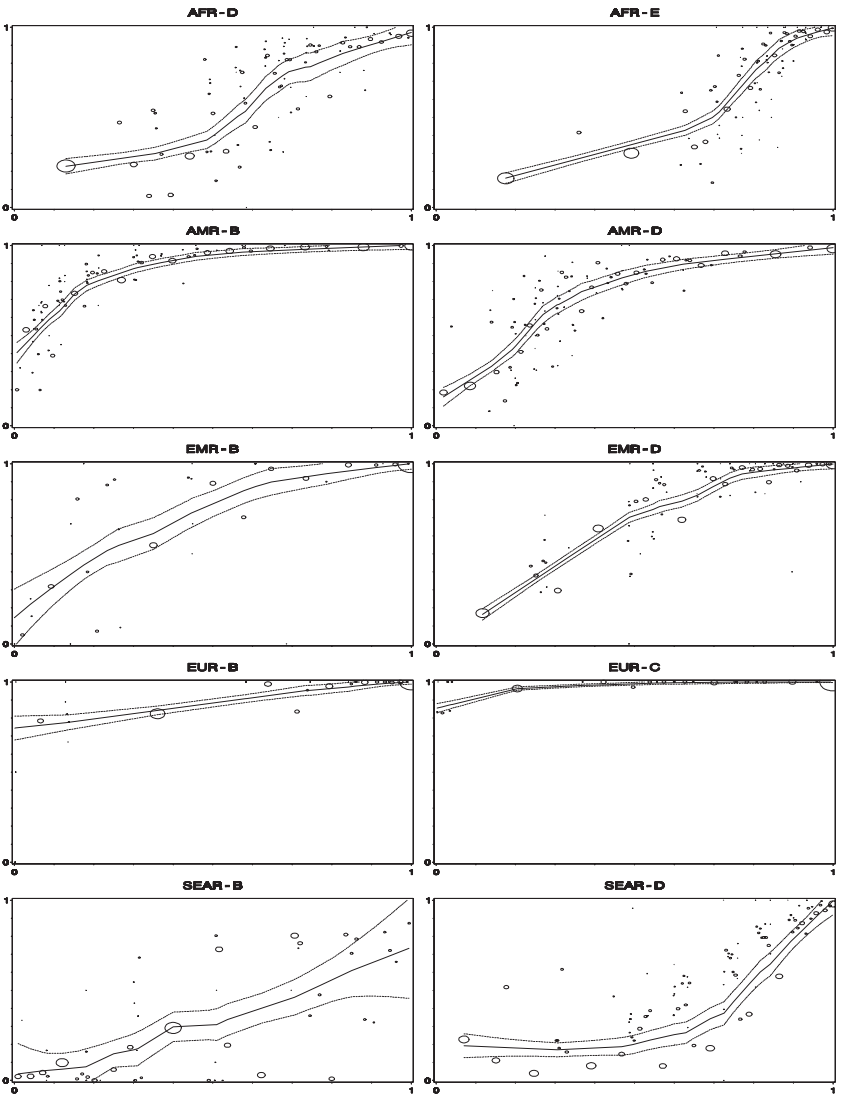


Table 24.23 Cell prevalence of unimproved water supply by poverty, prevalence of unimproved water supply by poverty, and relative risk of unimproved water supply by poverty, by subregion, for a US$ 2 per day cut-off

	Unimproved water supply				*Prevalence of unimproved water supply (%)*		*Relative risk of unimproved water supply*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	39.3	16.2	*55.5*	<US$ 1/day	70.8	<US$ 1/day	9.4
	US$ 1–2/day	4.9	17.5	*22.4*	US$ 1–2/day	21.8	US$ 1–2/day	2.9
	>US$ 2/day	1.7	20.4	*22.1*	>US$ 2/day	7.6	US$ 2/day	1.6
		45.8	*54.2*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	25.1	2.2	*27.3*	<US$ 1/day	92.1	<US$ 1/day	4.6
	US$ 1–2/day	25.2	11.0	*36.2*	US$ 1–2/day	69.6	US$ 1–2/day	3.5
	>US$ 2/day	7.3	29.2	*36.5*	>US$ 2/day	20.0	US$ 2/day	2.7
		57.6	*42.4*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	10.4	0.6	*11.0*	<US$ 1/day	94.8	<US$ 1/day	12.3
	US$ 1–2/day	8.4	10.7	*19.1*	US$ 1–2/day	44.2	US$ 1–2/day	5.7
	>US$ 2/day	5.4	64.4	*69.8*	>US$ 2/day	7.7	US$ 2/day	3.4
		24.2	*75.8*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	15.6	1.8	*17.4*	<US$ 1/day	89.9	<US$ 1/day	8.9
	US$ 1–2/day	10.7	15.6	*26.3*	US$ 1–2/day	40.8	US$ 1–2/day	4.0
	>US$ 2/day	5.7	50.6	*56.3*	>US$ 2/day	10.1	US$ 2/day	2.3
		32.1	*67.9*	***100.0***			>US$ 2/day	1
EMR-B	<US$ 1/day	1.0	1.0	*2.0*	<US$ 1/day	48.8	<US$ 1/day	3.6
	US$ 1–2/day	3.1	4.0	*7.1*	US$ 1–2/day	43.6	US$ 1–2/day	3.2
	>US$ 2/day	12.5	78.4	*90.9*	>US$ 2/day	13.7	US$ 2/day	2.9
		16.6	*83.4*	***100.0***			>US$ 2/day	1
EMR-D[a]	<US$ 1/day	15.0	4.3	*19.3*	<US$ 1/day	77.9	<US$ 1/day	15.1
	US$ 1–2/day	16.9	28.4	*45.3*	US$ 1–2/day	37.2	US$ 1–2/day	7.2
	>US$ 2/day	1.8	33.5	*35.3*	>US$ 2/day	5.2	US$ 2/day	3.2
		33.7	*66.3*	***100.0***			>US$ 2/day	1
EUR-B	<US$ 1/day	1.4	1.6	*3.0*	<US$ 1/day	47.8	<US$ 1/day	3.1
	US$ 1–2/day	6.0	8.6	*14.6*	US$ 1–2/day	41.3	US$ 1–2/day	2.7
	>US$ 2/day	12.7	69.6	*82.3*	>US$ 2/day	15.4	US$ 2/day	2.3
		20.2	*79.8*	***100.0***			>US$ 2/day	1
EUR-C	<US$ 1/day	1.2	4.2	*5.4*	<US$ 1/day	22.0	<US$ 1/day	11.8
	US$ 1–2/day	3.4	17.9	*21.3*	US$ 1–2/day	15.8	US$ 1–2/day	8.4
	>US$ 2/day	1.4	71.9	*73.3*	>US$ 2/day	1.9	US$ 2/day	5.4
		5.9	*94.1*	***100.0***			>US$ 2/day	1
SEAR-B	<US$ 1/day	2.8	3.6	*6.4*	<US$ 1/day	43.3	<US$ 1/day	2.0
	US$ 1–2/day	15.9	26.6	*42.5*	US$ 1–2/day	37.5	US$ 1–2/day	1.7
	>US$ 2/day	11.3	39.7	*51.0*	>US$ 2/day	22.1	US$ 2/day	1.4
		30.0	*70.0*	***100.0***			>US$ 2/day	1
SEAR-D	<US$ 1/day	34.2	8.2	*42.4*	<US$ 1/day	80.7	<US$ 1/day	5.0
	US$ 1–2/day	28.3	14.5	*42.8*	US$ 1–2/day	66.0	US$ 1–2/day	4.1
	>US$ 2/day	2.4	12.4	*14.8*	>US$ 2/day	16.0	US$ 2/day	2.1
		64.9	*35.2*	***100.0***			>US$ 2/day	1

Table 24.23 Cell prevalence of unimproved water supply by poverty, prevalence of unimproved water supply by poverty, and relative risk of unimproved water supply by poverty, by subregion, for a US$ 2 per day cut-off (*continued*)

	Unimproved water supply				*Prevalence of unimproved water supply (%)*		*Relative risk of unimproved water supply*	
Subregion		*Yes*	*No*					
WPR-B	<US$ 1/day	14.0	3.9	*17.9*	<US$ 1/day	78.1	<US$ 1/day	1.7
	US$ 1–2/day	21.5	12.5	*34.0*	US$ 1–2/day	63.2	US$ 1–2/day	1.3
	>US$ 2/day	22.6	25.5	*48.1*	>US$ 2/day	47.1	US$ 2/day	1.0
		58.1	*41.9*	***100.0***			>US$ 2/day	1
Total[b]	<US$ 1/day	18.6	5.1	*23.7*	<US$ 1/day	78.7	<US$ 1/day	3.3
	US$ 1–2/day	18.5	14.8	*33.4*	US$ 1–2/day	55.5	US$ 1–2/day	2.4
	>US$ 2/day	10.1	32.8	*42.9*	>US$ 2/day	23.6	US$ 2/day	—
		47.3	*52.7*	***100.0***			>US$ 2/day	1

[a] The results for EMR-D should be interpreted with caution.

[b] Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

For males, the relatively low sample sizes meant that it was possible to perform locally linear kernel regression directly on the unit-level (binary) data using logistic regression (see Figure 24.14). Thus, in Figure 24.14 (unlike the other figures of smooth plots) there are no "bubbles", since there was no need to use aggregate data. For females, the same aggregated procedure was used as for the other risk factors in this chapter, using the prevalence of each risk factor at the discrete values of the asset score. From the smoothed curves, we estimated the distribution of UN1 and UN2 within categories of income and by subregion, as before. Table 24.25 gives sample sizes by subregion.

Sample sizes were inadequate for several subregions. In EUR-B and EUR-C, adequate data were available only for women and detailed results for these subregions are not presented. Sex with a non-marital partner and condom use among those having sex with a non-marital partner were each less common among poor males and females within subregions (see Figures 24.14–24.16).

The prevalence of higher-risk (non-marital) sex by poverty for a US$2 per day cut-off is shown in Tables 24.26 (males) and 24.27 (females). The prevalence of condom use during higher-risk sex by poverty for a US$2 per day cut-off is shown in Tables 24.28 (males) and 24.29 (females). The corresponding values for a US$1 per day cut-off are given in Tables 24.30–24.33.

Table 24.24 Cell prevalence of unimproved water supply by poverty, prevalence of unimproved water supply by poverty, and relative risk of unimproved water supply by poverty, by subregion, for a US$ 1 per day cut-off

		Unimproved water supply			*Prevalence of unimproved water supply (%)*		*Relative risk of unimproved water supply*	
		Yes	*No*					
AFR-D	<US$ 1/day	38.5	17.0	*55.5*	<US$ 1/day	69.5	<US$ 1/day	4.2
	>US$ 1/day	7.3	37.2	*44.5*	>US$ 1/day	16.4	US$ 1/day	2.1
		45.8	*54.2*	***100.0***			>US$ 1/day	1
AFR-E	<US$ 1/day	24.0	3.3	*27.3*	<US$ 1/day	88.0	<US$ 1/day	1.9
	>US$ 1/day	33.6	39.1	*72.7*	>US$ 1/day	46.2	US$ 1/day	1.7
		57.6	*42.4*	***100.0***			>US$ 1/day	1
AMR-B	<US$ 1/day	10.5	0.5	*11.0*	<US$ 1/day	95.1	<US$ 1/day	6.2
	>US$ 1/day	13.8	75.2	*89.0*	>US$ 1/day	15.5	US$ 1/day	4.8
		24.2	*75.8*	***100.0***			>US$ 1/day	1
AMR-D	<US$ 1/day	15.4	2.0	*17.4*	<US$ 1/day	88.6	<US$ 1/day	4.4
	>US$ 1/day	16.6	66.0	*82.6*	>US$ 1/day	20.2	US$ 1/day	3.5
		32.1	*67.9*	***100.0***			>US$ 1/day	1
EMR-B	<US$ 1/day	1.0	1.0	*2.0*	<US$ 1/day	47.9	<US$ 1/day	3.0
	>US$ 1/day	15.6	82.4	*98.0*	>US$ 1/day	15.9	US$ 1/day	2.9
		16.6	*83.4*	***100.0***			>US$ 1/day	1
EMR-D[a]	<US$ 1/day	14.9	4.4	*19.3*	<US$ 1/day	77.4	<US$ 1/day	3.3
	>US$ 1/day	18.8	61.9	*80.7*	>US$ 1/day	23.3	US$ 1/day	2.8
		33.7	*66.3*	***100.0***			>US$ 1/day	1
EUR-B	<US$ 1/day	1.3	1.7	*3.0*	<US$ 1/day	42.6	<US$ 1/day	2.2
	>US$ 1/day	18.9	78.1	*97.0*	>US$ 1/day	19.5	US$ 1/day	2.2
		20.2	*79.8*	***100.0***			>US$ 1/day	1
EUR-C	<US$ 1/day	1.5	3.9	*5.4*	<US$ 1/day	27.0	<US$ 1/day	5.7
	>US$ 1/day	4.5	90.1	*94.6*	>US$ 1/day	4.7	US$ 1/day	5.6
		5.9	*94.1*	***100.0***			>US$ 1/day	1
SEAR-B	<US$ 1/day	2.8	3.6	*6.4*	<US$ 1/day	43.2	<US$ 1/day	1.5
	>US$ 1/day	27.2	66.4	*93.6*	>US$ 1/day	29.1	US$ 1/day	1.5
		30.0	*70.0*	***100.0***			>US$ 1/day	1
SEAR-D	<US$ 1/day	34.3	8.1	*42.4*	<US$ 1/day	80.8	<US$ 1/day	1.5
	>US$ 1/day	30.6	27.0	*57.6*	>US$ 1/day	53.1	US$ 1/day	1.5
		64.9	*35.2*	***100.0***			>US$ 1/day	1
WPR-B	<US$ 1/day	14.0	3.9	*17.9*	<US$ 1/day	78.1	<US$ 1/day	1.5
	>US$ 1/day	44.1	38	*82.1*	>US$ 1/day	46.9	US$ 1/day	1.5
		58.1	*41.9*	***100.0***			>US$ 1/day	1
Total[b]	<US$ 1/day	20.2	3.5	*23.7*	<US$ 1/day	85.3	<US$ 1/day	2.4
	>US$ 1/day	27.1	49.2	*76.3*	>US$ 1/day	35.5	US$ 1/day	—
		47.3	*52.7*	***100.0***			>US$ 1/day	1

[a] The results for EMR-D should be interpreted with caution.

[b] Summary of the subregions in the table, including the subregional-level estimates for EMR-D.

Figure 24.14 Loess plots of the prevalence of higher-risk sex (y-axis) by normalized rank of asset score (x-axis): men

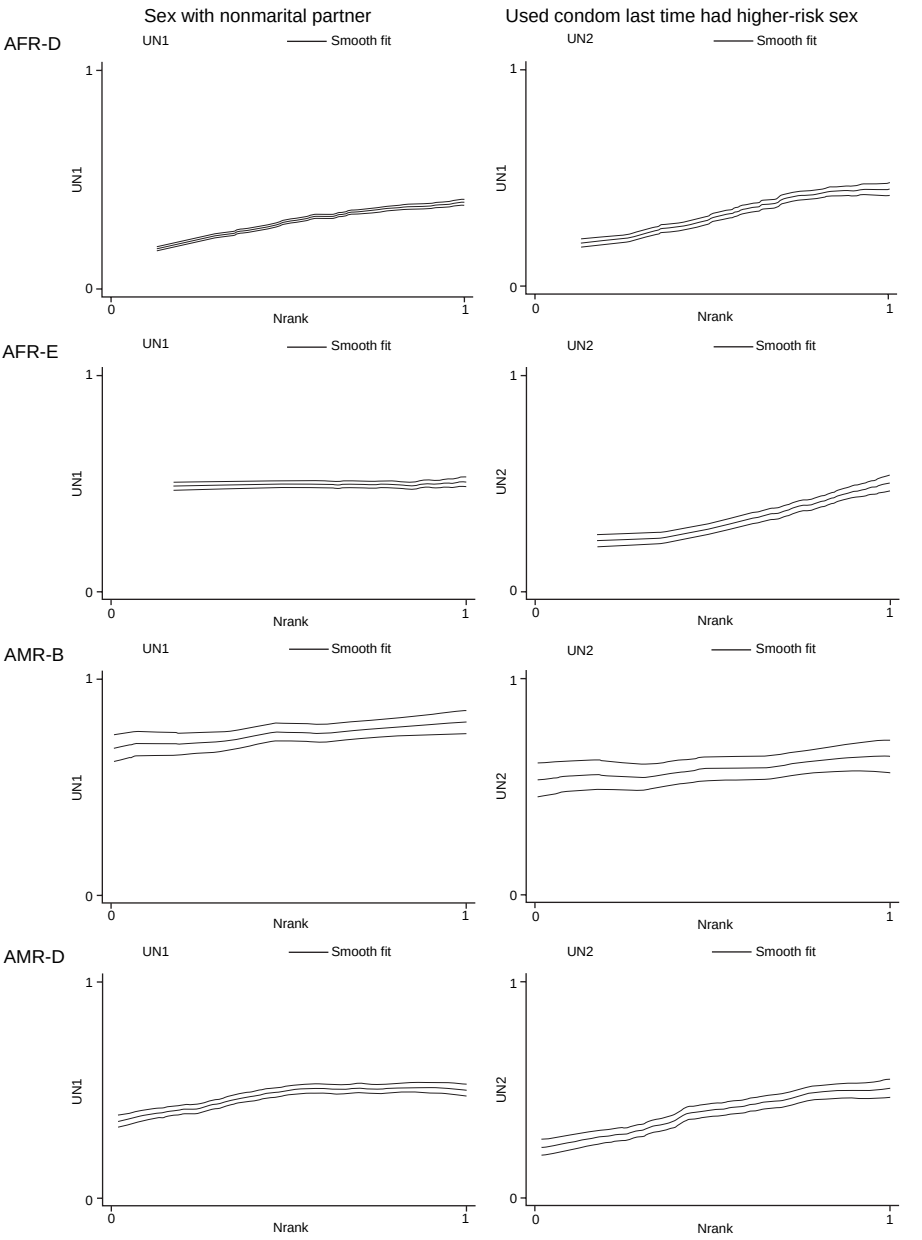


Figure 24.15 Loess plots of the prevalence of higher-risk sex (y-axis) by normalized rank of asset score (x-axis): women

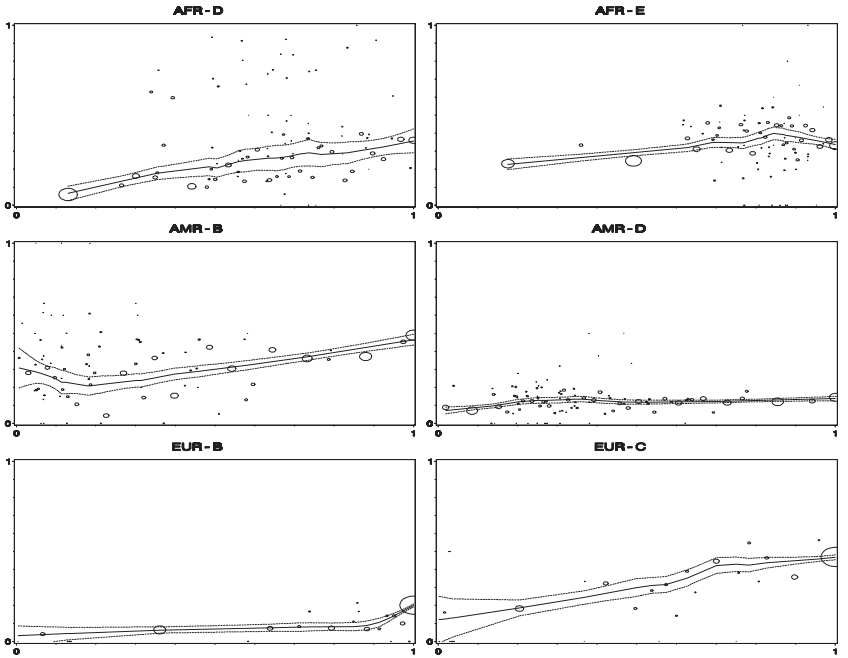


Table 24.25 Sample sizes for unsafe sex analyses by subregion

					Those having non-marital sex who:			
	No non-marital sex		*Non-marital sex*		*did not use a condom*		*used a condom*	
Subregion	*n*	*Percentage*	*n*	*Percentage*	*n*	*Percentage*	*n*	*Percentage*
Males								
AFR-D	13 036	73	4 820	27	2 995	67	1 487	33
AFR-E	4 043	51	3 897	49	1 883	64	1 045	36
AMR-B	220	26	638	74	265	42	373	58
AMR-D	2 360	54	1 999	46	1 193	60	792	40
EUR-C	147	48	159	52	60	38	99	62
Females								
AFR-D	48 735	82	11 050	18	7 250	84	1 350	16
AFR-E	27 056	69	12 131	31	6 873	83	1 401	17
AMR-B	13 720	68	6 385	32	3 950	77	1 154	23
AMR-D	33 841	88	4 472	12	3 605	90	382	10
EUR-B	2 297	92	197	8				
EUR-C	1 205	67	598	33	408	79	107	21

Figure 24.16 Loess plots of the prevalence of higher-risk sex (with condom use) (y-axis) by normalized rank of asset score (x-axis): women

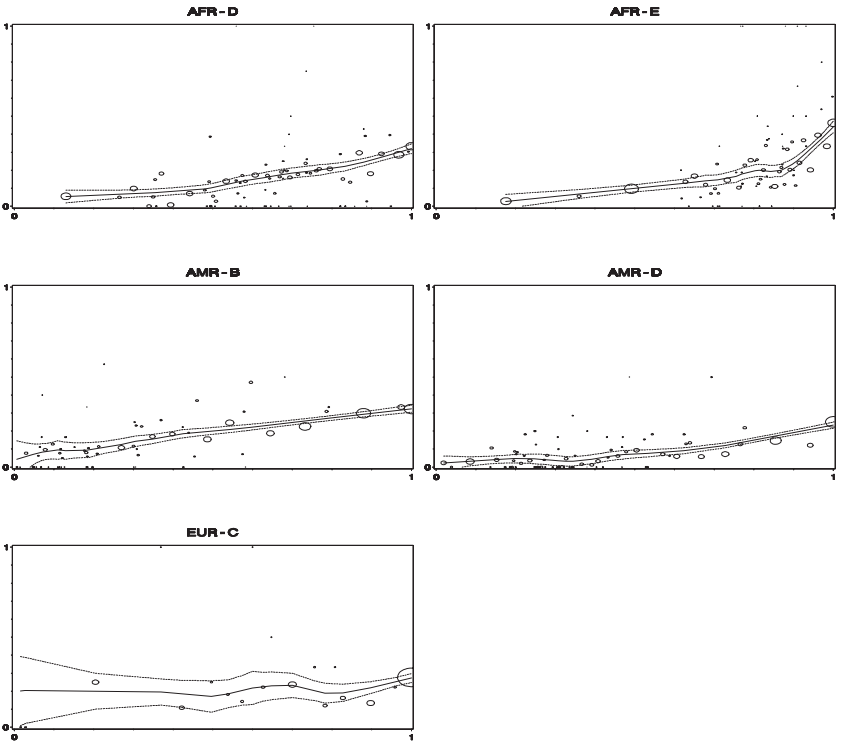


Table 24.26 Cell prevalence of non-marital sex in the last year (UN1) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among males, by subregion, for a US$ 2 per day cut-off

Subregion	*Non-marital sex*	*Yes*	*No*		*Prevalence of non-marital sex by poverty (%)*		*Relative risk of non-marital sex*	
AFR-D	<US$ 1/day	16.6	38.9	*55.5*	<US$ 1/day	29.9	<US$ 1/day	0.6
	US$ 1–2/day	10.3	12.1	*22.4*	US$ 1–2/day	45.8	US$ 1–2/day	0.9
	>US$ 2/day	11.8	10.3	*22.1*	>US$ 2/day	53.6	US$ 2/day	0.9
		38.7	*61.3*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	8.6	18.7	*27.3*	<US$ 1/day	31.4	<US$ 1/day	0.9
	US$ 1–2/day	14.1	22.1	*36.2*	US$ 1–2/day	38.9	US$ 1–2/day	1.1
	>US$ 2/day	13.2	23.3	*36.5*	>US$ 2/day	36.0	US$ 2/day	1.0
		35.8	*64.2*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	3.7	7.3	*11.0*	<US$ 1/day	33.8	<US$ 1/day	0.7
	US$ 1–2/day	7.8	11.3	*19.1*	US$ 1–2/day	41.0	US$ 1–2/day	0.9
	>US$ 2/day	32.9	36.9	*69.8*	>US$ 2/day	47.1	US$ 2/day	1.0
		44.4	*55.6*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	5.5	11.9	*17.4*	<US$ 1/day	31.6	<US$ 1/day	0.7
	US$ 1–2/day	11.1	15.2	*26.3*	US$ 1–2/day	42.3	US$ 1–2/day	0.9
	>US$ 2/day	26.6	29.7	*56.3*	>US$ 2/day	47.3	US$ 2/day	1.0
		43.3	*56.7*	***100.0***			>US$ 2/day	1
Total[a]	<US$ 1/day	8.2	18.2	*26.4*	<US$ 1/day	31.1	<US$ 1/day	0.7
	US$ 1–2/day	10.3	14.7	*25.0*	US$ 1–2/day	41.3	US$ 1–2/day	0.9
	>US$ 2/day	22.1	26.4	*48.6*	>US$ 2/day	45.6	US$ 2/day	—
		40.7	*59.3*	***100.0***			>US$ 2/day	1

[a] Summary of the subregions in the table.

Table 24.27 Cell prevalence of non-marital sex in the last year (UN1) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among females, by subregion, for a US$2 per day cut-off

	Non-marital sex				*Prevalence of non-marital sex by poverty (%)*		*Relative risk of non-marital sex*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	5.5	50.0	*55.5*	<US$ 1/day	9.9	<US$ 1/day	0.4
	US$ 1–2/day	4.3	18.1	*22.4*	US$ 1–2/day	19.0	US$ 1–2/day	0.9
	>US$ 2/day	4.9	17.2	*22.1*	>US$ 2/day	22.2	US$ 2/day	0.9
		14.6	*85.4*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	3.2	24.1	*27.3*	<US$ 1/day	11.6	<US$ 1/day	0.6
	US$ 1–2/day	5.6	30.6	*36.2*	US$ 1–2/day	15.3	US$ 1–2/day	0.8
	>US$ 2/day	6.9	29.6	*36.5*	>US$ 2/day	19.0	US$ 2/day	1.0
		15.7	*84.3*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	1.8	9.2	*11.0*	<US$ 1/day	16.1	<US$ 1/day	0.8
	US$ 1–2/day	2.4	16.7	*19.1*	US$ 1–2/day	12.6	US$ 1–2/day	0.6
	>US$ 2/day	13.8	56.0	*69.8*	>US$ 2/day	19.7	US$ 2/day	0.7
		18.0	*82.0*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	0.5	16.9	*17.4*	<US$ 1/day	2.9	<US$ 1/day	0.7
	US$ 1–2/day	1.1	25.2	*26.3*	US$ 1–2/day	4.2	US$ 1–2/day	1.0
	>US$ 2/day	2.3	54.0	*56.3*	>US$ 2/day	4.1	US$ 2/day	1.0
		3.9	*96.1*	***100.0***			>US$ 2/day	1
EUR-C	<US$ 1/day	0.3	5.1	*5.4*	<US$ 1/day	5.7	<US$ 1/day	0.4
	US$ 1–2/day	1.4	19.9	*21.3*	US$ 1–2/day	6.7	US$ 1–2/day	0.4
	>US$ 2/day	11.3	62.0	*73.3*	>US$ 2/day	15.4	US$ 2/day	0.5
		13.0	*87.0*	***100.0***			>US$ 2/day	11
Total[a]	<US$ 1/day	2.3	19.0	*21.4*	<US$ 1/day	10.8	<US$ 1/day	0.6
	US$ 1–2/day	3.1	21.0	*24.1*	US$ 1–2/day	12.8	US$ 1–2/day	0.7
	>US$ 2/day	9.6	44.9	*54.6*	>US$ 2/day	17.6	US$ 2/day	—
		15.0	*85.0*	***100.0***			>US$ 2/day	1

[a] Summary of the subregions in the table.

Table 24.28 Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among males, by subregion, for a US$ 2 per day cut-off

	Condom use				*Prevalence of condom use by poverty (%)*		*Relative risk of condom use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	14.2	41.3	*55.5*	<US$ 1/day	25.5	<US$ 1/day	0.5
	US$ 1–2/day	9.6	12.8	*22.4*	US$ 1–2/day	42.9	US$ 1–2/day	0.9
	>US$ 2/day	10.4	11.7	*22.1*	>US$ 2/day	46.9	US$ 2/day	1.0
		34.1	*65.9*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	4.6	22.7	*27.3*	<US$ 1/day	17.0	<US$ 1/day	0.3
	US$ 1–2/day	13.1	23.1	*36.2*	US$ 1–2/day	36.1	US$ 1–2/day	0.7
	>US$ 2/day	19.3	17.2	*36.5*	>US$ 2/day	52.9	US$ 2/day	0.8
		37.0	*63.0*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	4.3	6.7	*11.0*	<US$ 1/day	38.9	<US$ 1/day	0.6
	US$ 1–2/day	9.4	9.7	*19.1*	US$ 1–2/day	49.5	US$ 1–2/day	0.8
	>US$ 2/day	41.8	28.0	*69.8*	>US$ 2/day	59.9	US$ 2/day	0.9
		55.6	*44.4*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	3.4	14.0	*17.4*	<US$ 1/day	19.3	<US$ 1/day	0.4
	US$ 1–2/day	9.5	16.8	*26.3*	US$ 1–2/day	36.3	US$ 1–2/day	0.7
	>US$ 2/day	28.5	27.8	*56.3*	>US$ 2/day	50.7	US$ 2/day	0.8
		41.5	*58.5*	***100.0***			>US$ 2/day	1
Total[a]	<US$ 1/day	6.7	19.8	*26.4*	<US$ 1/day	25.2	<US$ 1/day	0.4
	US$ 1–2/day	10.5	14.5	*25.0*	US$ 1–2/day	41.9	US$ 1–2/day	0.7
	>US$ 2/day	27.4	21.2	*48.6*	>US$ 2/day	56.4	US$ 2/day	—
		44.5	*55.5*	***100.0***			>US$ 2/day	1

[a] Summary of the subregions in the table.

Table 24.29 Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among females, by subregion, for a US$ 2 per day cut-off

	Condom use				*Prevalence of condom use by poverty (%)*		*Relative risk of condom use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	5.1	50.4	*55.5*	<US$ 1/day	9.3	<US$ 1/day	0.3
	US$ 1–2/day	5.0	17.4	*22.4*	US$ 1–2/day	22.4	US$ 1–2/day	0.7
	>US$ 2/day	7.0	15.1	*22.1*	>US$ 2/day	31.5	US$ 2/day	0.9
		17.1	*82.9*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	0.8	26.5	*27.3*	<US$ 1/day	3.1	<US$ 1/day	0.1
	US$ 1–2/day	5.3	30.9	*36.2*	US$ 1–2/day	14.7	US$ 1–2/day	0.4
	>US$ 2/day	12.7	23.8	*36.5*	>US$ 2/day	34.8	US$ 2/day	0.6
		18.8	*81.2*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	1.4	9.6	*11.0*	<US$ 1/day	12.6	<US$ 1/day	0.3
	US$ 1–2/day	3.6	15.5	*19.1*	US$ 1–2/day	18.9	US$ 1–2/day	0.4
	>US$ 2/day	29.7	40.1	*69.8*	>US$ 2/day	42.5	US$ 2/day	0.6
		34.7	*65.3*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	2.6	14.8	*17.4*	<US$ 1/day	15.1	<US$ 1/day	0.3
	US$ 1–2/day	5.2	21.1	*26.3*	US$ 1–2/day	19.8	US$ 1–2/day	0.3
	>US$ 2/day	33.9	22.4	*56.3*	>US$ 2/day	60.1	US$ 2/day	0.5
		41.7	*58.3*	***100.0***			>US$ 2/day	1
EUR-C	<US$ 1/day	1.2	4.2	*5.4*	<US$ 1/day	21.7	<US$ 1/day	1.0
	US$ 1–2/day	4.7	16.6	*21.3*	US$ 1–2/day	22.2	US$ 1–2/day	1.0
	>US$ 2/day	16.6	56.7	*73.3*	>US$ 2/day	22.6	US$ 2/day	1.0
		22.5	*77.5*	***100.0***			>US$ 2/day	1
Total[a]	<US$ 1/day	2.0	19.4	*21.4*	<US$ 1/day	9.2	<US$ 1/day	0.3
	US$ 1–2/day	4.6	19.5	*24.1*	US$ 1–2/day	18.9	US$ 1–2/day	0.5
	>US$ 2/day	19.2	35.4	*54.6*	>US$ 2/day	35.2	US$ 2/day	—
		25.7	*74.3*	***100.0***			>US$ 2/day	1

[a] Summary of the subregions in the table.

Table 24.30 Cell prevalence of non-marital sex in the last year (UN1) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among males, by subregion, for a US$ 1 per day cut-off

	Non-marital sex			*Prevalence of non-marital sex by poverty (%)*		*Relative risk of non-marital sex*	
Subregion		*Yes*	*No*				
AFR-D	<US$ 1/day	16.8	38.7	*55.5*	<US$ 1/day	30.3	<US$ 1/day 0.6
	>US$ 1/day	21.9	22.6	*44.5*	>US$ 1/day	49.2	US$ 1/day 0.9
		38.7	*61.3*	***100.0***			>US$ 1/day 1
AFR-E	<US$ 1/day	8.5	18.8	*27.3*	<US$ 1/day	31.3	<US$ 1/day 0.8
	>US$ 1/day	27.3	45.4	*72.7*	>US$ 1/day	37.5	US$ 1/day 1.1
		35.8	*64.2*	***100.0***			>US$ 1/day 1
AMR-B	<US$ 1/day	3.7	7.3	*11.0*	<US$ 1/day	33.8	<US$ 1/day 0.7
	>US$ 1/day	40.7	48.3	*89.0*	>US$ 1/day	45.7	US$ 1/day 0.8
		44.4	*55.6*	***100.0***			>US$ 1/day 1
AMR-D	<US$ 1/day	5.5	11.9	*17.4*	<US$ 1/day	31.6	<US$ 1/day 0.7
	>US$ 1/day	37.8	44.8	*82.6*	>US$ 1/day	45.7	US$ 1/day 0.8
		43.3	*56.7*	***100.0***			>US$ 1/day 1
Total[a]	<US$ 1/day	8.3	18.2	*26.4*	<US$ 1/day	31.3	<US$ 1/day 0.7
	>US$ 1/day	32.4	41.2	*73.6*	>US$ 1/day	44.1	US$ 1/day —
		40.7	*59.3*	***100.0***			>US$ 1/day 1

[a] Summary of the subregions in the table.

Table 24.31 Cell prevalence of non-marital sex in the last year (UN1) by poverty, prevalence of non-marital sex by level of income poverty and relative risk of non-marital sex among females, by subregion, for a US$ 1 per day cut-off

	Non-marital sex			*Prevalence of non-marital sex by poverty (%)*		*Relative risk of non-marital sex*	
Subregion		*Yes*	*No*				
AFR-D	<US$ 1/day	5.5	50.0	*55.5*	<US$ 1/day	10.0	<US$ 1/day 0.5
	>US$ 1/day	9.1	35.4	*44.5*	>US$ 1/day	20.4	US$ 1/day 0.9
		14.6	*85.4*	***100.0***			>US$ 1/day 1
AFR-E	<US$ 1/day	3.2	24.1	*27.3*	<US$ 1/day	11.7	<US$ 1/day 0.7
	>US$ 1/day	12.5	60.2	*72.7*	>US$ 1/day	17.1	US$ 1/day 0.8
		15.7	*84.3*	***100.0***			>US$ 1/day 1
AMR-B	<US$ 1/day	1.8	9.2	*11.0*	<US$ 1/day	16.1	<US$ 1/day 0.9
	>US$ 1/day	16.2	72.8	*89.0*	>US$ 1/day	18.2	US$ 1/day 0.8
		18.0	*82.0*	***100.0***			>US$ 1/day 1
AMR-D	<US$ 1/day	0.5	16.9	*17.4*	<US$ 1/day	2.9	<US$ 1/day 0.7
	>US$ 1/day	3.4	79.2	*82.6*	>US$ 1/day	4.1	US$ 1/day 0.9
		3.9	*96.1*	***100.0***			>US$ 1/day 1
EUR-C	<US$ 1/day	0.3	5.1	*5.4*	<US$ 1/day	5.4	<US$ 1/day 0.4
	>US$ 1/day	12.7	81.9	*94.6*	>US$ 1/day	13.4	US$ 1/day 0.4
		13.0	*87.0*	***100.0***			>US$ 1/day 1
Total[a]	<US$ 1/day	2.3	19.0	*21.4*	<US$ 1/day	10.9	<US$ 1/day 0.7
	>US$ 1/day	12.7	65.9	*78.6*	>US$ 1/day	16.1	US$ 1/day —
		15.0	*85.0*	***100.0***			>US$ 1/day 1

[a] Summary of the subregions in the table.

Table 24.32 Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among males, by subregion, for a US$ 1 per day cut-off

		Condom use			*Prevalence of condom use by poverty (%)*		*Relative risk of condom use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	14.3	41.2	*55.5*	<US$ 1/day	25.7	<US$ 1/day	0.6
	>US$ 1/day	19.9	24.6	*44.5*	>US$ 1/day	44.7	US$ 1/day	0.9
		34.1	*65.9*	***100.0***			>US$ 1/day	1
AFR-E	<US$ 1/day	4.8	22.5	*27.3*	<US$ 1/day	17.4	<US$ 1/day	0.4
	>US$ 1/day	32.3	40.4	*72.7*	>US$ 1/day	44.4	US$ 1/day	0.7
		37.0	*63.0*	***100.0***			>US$ 1/day	1
AMR-B	<US$ 1/day	4.3	6.7	*11.0*	<US$ 1/day	38.9	<US$ 1/day	0.7
	>US$ 1/day	51.3	37.7	*89.0*	>US$ 1/day	57.6	US$ 1/day	0.7
		55.6	*44.4*	***100.0***			>US$ 1/day	1
AMR-D	<US$ 1/day	3.4	14.0	*17.4*	<US$ 1/day	19.4	<US$ 1/day	0.4
	>US$ 1/day	38.1	44.5	*82.6*	>US$ 1/day	46.1	US$ 1/day	0.6
		41.5	*58.5*	***100.0***			>US$ 1/day	1
Total[a]	<US$ 1/day	6.7	19.7	*26.4*	<US$ 1/day	25.5	<US$ 1/day	0.5
	>US$ 1/day	37.8	35.8	*73.6*	>US$ 1/day	51.4	US$ 1/day	—
		44.5	*55.5*	***100.0***			>US$ 1/day	1

[a] Summary of the subregions in the table.

Table 24.33 Cell prevalence of condom use with higher-risk sex (UN2) by poverty, prevalence of condom use with higher-risk sex by level of income poverty and relative risk of condom use with higher-risk sex among females, by subregion, for a US$ 1 per day cut-off

		Condom use			*Prevalence of condom use by poverty (%)*		*Relative risk of condom use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	5.3	50.2	*55.5*	<US$ 1/day	9.6	<US$ 1/day	0.4
	>US$ 1/day	11.8	32.7	*44.5*	>US$ 1/day	26.5	US$ 1/day	0.7
		17.1	*82.9*	***100.0***			>US$ 1/day	1
AFR-E	<US$ 1/day	0.9	26.4	*27.3*	<US$ 1/day	3.3	<US$ 1/day	0.1
	>US$ 1/day	17.9	54.8	*72.7*	>US$ 1/day	24.7	US$ 1/day	0.3
		18.8	*81.2*	***100.0***			>US$ 1/day	1
AMR-B	<US$ 1/day	1.4	9.6	*11.0*	<US$ 1/day	12.6	<US$ 1/day	0.3
	>US$ 1/day	33.3	55.7	*89.0*	>US$ 1/day	37.4	US$ 1/day	0.4
		34.7	*65.3*	***100.0***			>US$ 1/day	1
AMR-D	<US$ 1/day	2.7	14.7	*17.4*	<US$ 1/day	15.3	<US$ 1/day	0.3
	>US$ 1/day	39.0	43.6	*82.6*	>US$ 1/day	47.3	US$ 1/day	0.4
		41.7	*58.3*	***100.0***			>US$ 1/day	1
EUR-C	<US$ 1/day	1.2	4.2	*5.4*	<US$ 1/day	21.7	<US$ 1/day	1.0
	>US$ 1/day	21.3	73.3	*94.6*	>US$ 1/day	22.5	US$ 1/day	1.0
		22.5	*77.5*	***100.0***			>US$ 1/day	1
Total[a]	<US$ 1/day	2.0	19.4	*21.4*	<US$ 1/day	9.4	<US$ 1/day	0.3
	>US$ 1/day	23.7	54.9	*78.6*	>US$ 1/day	30.1	US$ 1/day	—
		25.7	*74.3*	***100.0***			>US$ 1/day	1

[a] Summary of the subregions in the table.

3.4 Indoor air pollution

LSMS data for 10 countries were used for the indoor air pollution analyses (see Table 24.14). Supplementary data were also used for three DHS countries—Colombia (AMR-B), Indonesia (SEAR-B) and India (SEAR-D). Two other demographic and health surveys (Ethiopia plus Zimbabwe combined and Panama) had some indoor air pollution data, but could not be used owing to inadequate sample sizes among the non-poor. The combined sample sizes and distribution are shown in Table 24.34.

We used estimates of indoor air pollution unadjusted for ventilation. The survey data questions we used to determine the joint association were compatible with these unadjusted estimates.

Locally linear kernel regression smooth plots of smoke-producing cooking fuel use according to normalized equivalized income ranking are shown in Figure 24.17. Each subfigure plots the proportion of households using smoke-producing cooking fuels (y-axis ranging from 0 to 1) by normalized equivalized income rank for the country (x-axis ranging from 0 [poorest] to 1 [richest in the country]).

Table 24.35 shows the estimated association of absolute poverty and indoor air pollution for a cut-off of US$ 2 per day, using the area under the curves in Figure 24.17. There is considerable variation between subregions in the average level of exposure to indoor air pollution and in the relative differences within subregions by poverty. In both African subregions, and to a lesser extent in both South-East Asian subregions, there is both an extraordinarily high prevalence of exposure to indoor air

Table 24.34 Sample sizes and distribution of household use of smoke-producing cooking fuels by subregion

	Presence of smoke-producing cooking fuels		*Absence of smoke-producing cooking fuels*	
Subregion	*n*	*Percentage*	*n*	*Percentage*
AFR-D	5 585	94	332	6
AFR-E	4 287	42	5 933	58
AMR-B	1 282	26	3 647	74
AMR-D	812	14	4 805	86
EMR-B	—	—	—	—
EMR-D	13 349	71	5 561	29
EUR-B	2 724	43	3 634	57
EUR-C	276	8	3 337	92
SEAR-B	12 705	44	16 105	56
SEAR-D	67 631	77	20 078	23
WPR-B	2 778	81	634	19

— No data.

Table 24.35 Cell prevalence of indoor air pollution by poverty, prevalence of indoor air pollution by poverty and relative risk of indoor air pollution by poverty, by subregion, for a US$ 2 per day cut-off

	Exposure			*Prevalence of indoor air pollution by poverty (%)*		*Relative risk of indoor air pollution*	
Subregion		*Yes*	*No*				
AFR-D	<US$ 1/day	42.1	13.4	*55.5*	<US$ 1/day	75.9	<US$ 1/day 1.1
	US$ 1–2/day	16.6	5.8	*22.4*	US$ 1–2/day	74.0	US$ 1–2/day 1.1
	>US$ 2/day	14.7	7.4	*22.1*	>US$ 2/day	66.4	US$ 2/day 1.1
		73.4	26.6	***100.0***			>US$ 2/day 1.0
AFR-E[a]	<US$ 1/day	27.3	0.0	*27.3*	<US$ 1/day	100.0	<US$ 1/day 2.0
	US$ 1–2/day	36.2	0.0	*36.2*	US$ 1–2/day	100.0	US$ 1–2/day 1.9
	>US$ 2/day	19.4	17.1	*36.5*	>US$ 2/day	53.3	US$ 2/day 1.7
		85.8	14.2	***100.0***			>US$ 2/day 1
AMR-B	<US$ 1/day	8.2	2.8	*11.0*	<US$ 1/day	74.6	<US$ 1/day 7.2
	US$ 1–2/day	9.2	9.9	*19.1*	US$ 1–2/day	48.1	US$ 1–2/day 4.7
	>US$ 2/day	7.2	62.6	*69.8*	>US$ 2/day	10.3	US$ 2/day 3.3
		24.6	75.4	***100.0***			>US$ 2/day 1
AMR-D[a]	<US$ 1/day	17.4	0.0	*17.4*	<US$ 1/day	100.0	<US$ 1/day 14.6
	US$ 1–2/day	15.0	11.3	*26.3*	US$ 1–2/day	56.9	US$ 1–2/day 4.7
	>US$ 2/day	6.9	49.4	*56.3*	>US$ 2/day	12.2	US$ 2/day 2.3
		52.9	47.1	***100.0***			>US$ 2/day 1
EMR-D	<US$ 1/day	17.0	2.3	*19.3*	<US$ 1/day	88.1	<US$ 1/day 4.0
	US$ 1–2/day	30.5	14.8	*45.3*	US$ 1–2/day	67.2	US$ 1–2/day 3.1
	>US$ 2/day	7.7	27.6	*35.3*	>US$ 2/day	21.9	US$ 2/day 1.7
		55.2	44.8	***100.0***			>US$ 2/day 1
EUR-B	<US$ 1/day	1.5	1.5	*3.0*	<US$ 1/day	49.1	<US$ 1/day 1.2
	US$ 1–2/day	7.2	7.4	*14.6*	US$ 1–2/day	49.5	US$ 1–2/day 1.2
	>US$ 2/day	32.8	49.5	*82.3*	>US$ 2/day	39.9	US$ 2/day 1.2
		41.5	58.5	***100.0***			>US$ 2/day 1
EUR-C	<US$ 1/day	1.5	3.9	*5.4*	<US$ 1/day	26.9	<US$ 1/day 1.3
	US$ 1–2/day	6.2	15.1	*21.3*	US$ 1–2/day	29.0	US$ 1–2/day 1.4
	>US$ 2/day	15.2	58.1	*73.3*	>US$ 2/day	20.7	US$ 2/day 1.3
		22.8	77.2	***100.0***			>US$ 2/day 1
SEAR-D[a]	<US$ 1/day	42.4	0.0	*42.4*	<US$ 1/day	100.0	<US$ 1/day 3.5
	US$ 1–2/day	35.6	7.2	*42.8*	US$ 1–2/day	83.3	US$ 1–2/day 2.9
	>US$ 2/day	4.3	10.5	*14.8*	>US$ 2/day	29.2	US$ 2/day 1.8
		83.5	16.5	***100.0***			>US$ 2/day 1
WPR-B	<US$ 1/day	16.6	1.3	*17.9*	<US$ 1/day	92.8	<US$ 1/day 1.4
	US$ 1–2/day	30.6	3.4	*34.0*	US$ 1–2/day	89.9	US$ 1–2/day 1.4
	>US$ 2/day	30.9	17.2	*48.1*	>US$ 2/day	64.3	US$ 2/day 1.3
		78.1	21.9	***100.0***			>US$ 2/day 1
Total[b]	<US$ 1/day	23.5	1.9	*25.4*	<US$ 1/day	92.3	<US$ 1/day 2.2
	US$ 1–2/day	26.9	6.7	*33.6*	US$ 1–2/day	80.1	US$ 1–2/day 1.9
	>US$ 2/day	17.2	23.9	*41.0*	>US$ 2/day	41.8	US$ 2/day —
		68.2	*31.8*	***100.0***			>US$ 2/day 1

[a] For these three subregions, the combination of the marginal prevalence of indoor air pollution and our relative risk estimates meant that the algebraic solutions gave prevalence estimates above 100% and below 0%. For example, the within US$ 1 per day stratum algebraic solution for AMR-D was a cell percentage of 31.1% for those exposed to pollution and a cell percentage of –13.7% for those unexposed. In these instances, the prevalence of exposure to indoor air pollution was set to 100%.

[b] Summary of the subregions in the table.

Figure 24.17 Loess plots of the prevalence of use of household smoke-producing cooking fuel (y-axis) by normalized income rank (x-axis)

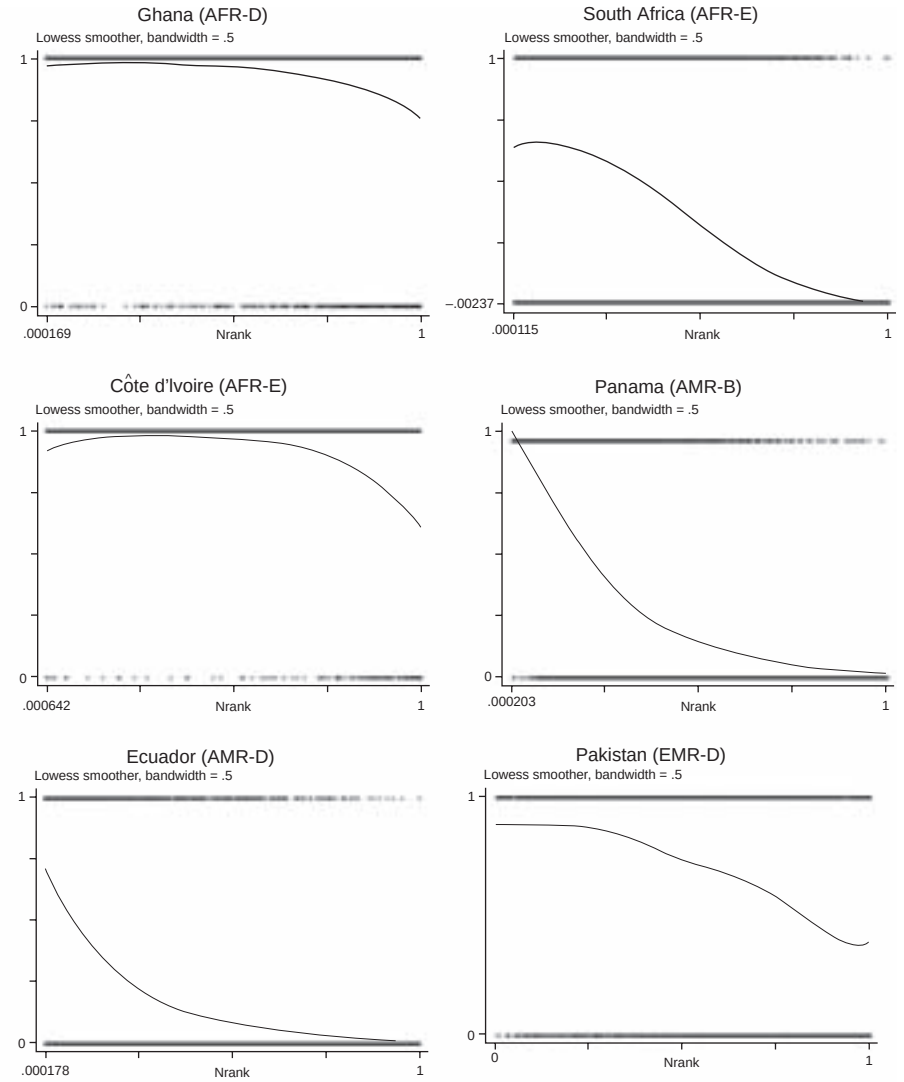
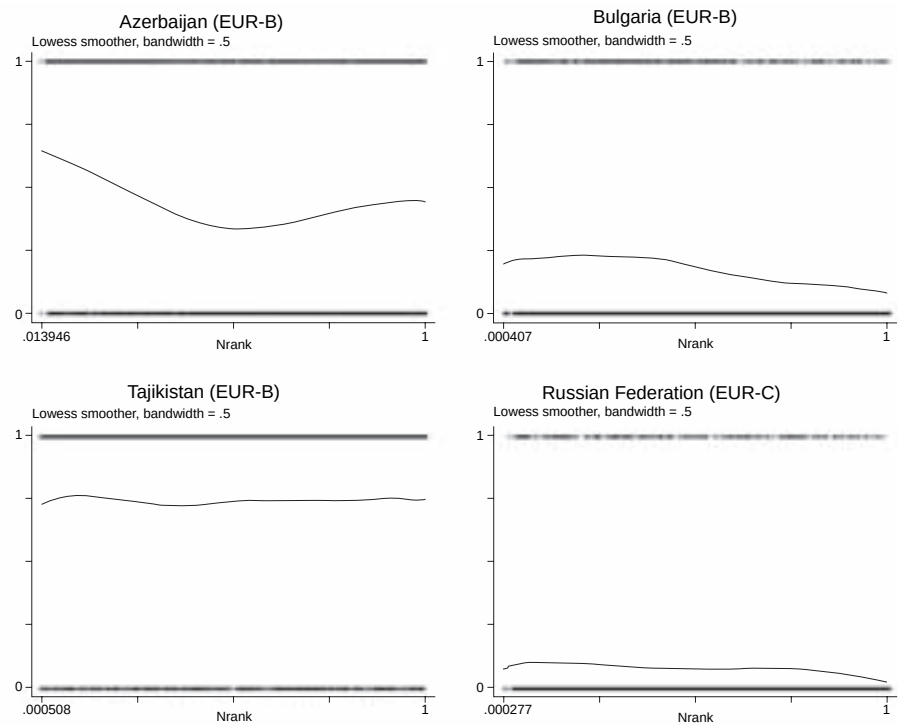


Figure 24.17 Loess plots of the prevalence of use of household smoke-producing cooking fuel (y-axis) by normalized income rank (x-axis) (*continued*)



pollution and little *relative* difference between the poor and non-poor. Analyses for SEAR-B are presented only for a cut-off of US$ 1 per day (Table 24.36), as the data were insufficient for analyses based on a US$2 per day cut-off.

For the 11 subregions, it was estimated that indoor air pollution was 2.2 times more likely among those living on <US$ 1 per day than among those living on >US$ 2 per day. We stress that this estimate obscures considerable heterogeneity between subregions.

3.5 Outdoor air pollution

The estimated exposures to ambient particulate air pollution (PM_{10}) by income category and subregion are shown in Table 24.37. The method used assumed that people living in urban areas are all exposed to a constant level of outdoor air pollution, and likewise those living in rural areas. However, the uneven distribution of income poverty between urban and rural areas allowed us to make (crude) estimates of varying

Table 24.36 Cell prevalence of indoor air pollution by poverty, prevalence of indoor air pollution by poverty and relative risk of indoor air pollution by poverty, by subregion, for a US$ 1 per day cut-off

	Indoor air pollution				*Prevalence of indoor air pollution by poverty (%)*		*Relative risk of indoor air pollution*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	42.0	13.5	*55.5*	<US$ 1/day	75.7	<US$ 1/day	1.1
	>US$ 1/day	31.4	13.1	*44.5*	>US$ 1/day	70.6	US$ 1/day	1.1
		73.4	*26.6*	***100.0***			>US$ 1/day	1
AFR-E[a]	<US$ 1/day	27.3	0.0	*27.3*	<US$ 1/day	100.0	<US$ 1/day	1.6
	>US$ 1/day	58.5	14.2	*72.7*	>US$ 1/day	80.5	US$ 1/day	1.6
		85.8	*14.2*	***100.0***			>US$ 1/day	1
AMR-B	<US$ 1/day	8.8	2.2	*11.0*	<US$ 1/day	79.8	<US$ 1/day	4.5
	>US$ 1/day	15.8	73.2	*89.0*	>US$ 1/day	17.8	US$ 1/day	3.8
		24.6	*75.4*	***100.0***			>US$ 1/day	1
AMR-D[a]	<US$ 1/day	17.4	0.0	*17.4*	<US$ 1/day	100.0	<US$ 1/day	5.9
	>US$ 1/day	35.5	47.1	*82.6*	>US$ 1/day	43.0	US$ 1/day	3.5
		52.9	*47.1*	***100.0***			>US$ 1/day	1
EMR-D	<US$ 1/day	14.9	4.4	*19.3*	<US$ 1/day	77.0	<US$ 1/day	1.5
	>US$ 1/day	40.3	40.4	*80.7*	>US$ 1/day	50.0	US$ 1/day	1.5
		55.2	*44.8*	***100.0***			>US$ 1/day	1
EUR-B	<US$ 1/day	1.5	1.5	*3.0*	<US$ 1/day	49.6	<US$ 1/day	1.2
	>US$ 1/day	40.0	57.0	*97.0*	>US$ 1/day	41.3	US$ 1/day	1.2
		41.5	*58.5*	***100.0***			>US$ 1/day	1
EUR-C	<US$ 1/day	1.5	3.9	*5.4*	<US$ 1/day	27.2	<US$ 1/day	1.2
	>US$ 1/day	21.3	73.3	*94.6*	>US$ 1/day	22.5	US$ 1/day	1.3
		22.8	*77.2*	***100.0***			>US$ 1/day	1
SEAR-B[a]	<US$ 1/day	6.4	0.0	*6.4*	<US$ 1/day	100.0	<US$ 1/day	2.6
	>US$ 1/day	60.1	33.5	*93.6*	>US$ 1/day	64.2	US$ 1/day	2.1
		66.5	*33.5*	***100.0***			>US$ 1/day	1
SEAR-D[a]	<US$ 1/day	42.4	0.0	*42.4*	<US$ 1/day	100.0	<US$ 1/day	1.5
	>US$ 1/day	41.1	16.5	*57.6*	>US$ 1/day	71.4	US$ 1/day	1.4
		83.5	*16.5*	***100.0***			>US$ 1/day	1
WPR-B	<US$ 1/day	16.5	1.4	*17.9*	<US$ 1/day	92.4	<US$ 1/day	1.2
	>US$ 1/day	61.6	20.5	*82.1*	>US$ 1/day	75.0	US$ 1/day	1.2
		78.1	*21.9*	***100.0***			>US$ 1/day	1
Total[b]	<US$ 1/day	23.3	2.1	*25.4*	<US$ 1/day	91.8	<US$ 1/day	1.5
	>US$ 1/day	44.9	29.7	*74.6*	>US$ 1/day	60.2	US$ 1/day	—
		68.2	*31.8*	***100.0***			>US$ 1/day	1

[a] For these four subregions, the combination of the marginal prevalence of indoor air pollution and our relative risk estimates meant that the algebraic solutions gave prevalence estimates above 100% and below 0%. In these instances, the prevalence of exposure to indoor air pollution was set to 100%.

[b] Summary of the subregions in the table.

Table 24.37 Exposure to ambient air pollution by poverty, by subregion, for a US$ 1 per day cut-off

Subregion	*Average exposure ($\mu g/m^3$)*		*Average exposure for people living on <US$ 1 as a percentage of average exposure for people living on >US$ 1*
AFR-D	<US$ 1/day	9	28
	>US$ 1/day	32	
AFR-E	<US$ 1/day	5	27
	>US$ 1/day	18	
AMR-B	<US$ 1/day	6	13
	>US$ 1/day	47	
AMR-D	<US$ 1/day	6	10
	>US$ 1/day	55	
EMR-B	<US$ 1/day	5	13
	>US$ 1/day	40	
EMR-D	<US$ 1/day	5	11
	>US$ 1/day	40	
EUR-B	<US$ 1/day	5	27
	>US$ 1/day	19	
EUR-C	<US$ 1/day	5	21
	>US$ 1/day	24	
SEAR-B	<US$ 1/day	5	13
	>US$ 1/day	38	
SEAR-D	<US$ 1/day	7	14
	>US$ 1/day	48	
WPR-B	<US$ 1/day	6	8
	>US$ 1/day	78	

exposure to air pollution by income poverty levels. (See section 2 for further details.)

Exposure to outdoor air pollution was estimated to be substantially higher in the higher-income category (those living on >US$ 1 per day) in all subregions. This is due to a greater proportion of poor people living in rural areas, where there is less outdoor air pollution. These summary estimates inevitably obscure considerable heterogeneity within countries and even within cities, where particular patterns of air pollution emissions and meteorological conditions may combine to produce local variations in air pollution.

3.6 Tobacco use

Chapter 11 uses an indirect method based on lung cancer mortality to estimate the contribution of tobacco use to the global burden of disease, and the prevalence of tobacco use by subregion was not estimated. We took estimates of subregional prevalence of tobacco use

from the web site of the WHO Tobacco or Health Programme (http://www.cdc.gov/tobacco/who/whofirst.htm).

Analyses for tobacco were based on LSMS data for 10 countries in seven subregions—AFR-D (Ghana), AFR-E (South Africa), AMR-B (Panama), AMR-D (Ecuador), EMR-D (Pakistan), EUR-B (Azerbaijan, Bulgaria, Tajikistan) and EUR-C (Kazakhstan, the Russian Federation). Data on tobacco for Bulgaria, Ghana, South Africa and Tajikistan were available only in the form of household expenditure data, and for Azerbaijan as a composite variable of combined alcohol and tobacco expenditure. A household was classified as a smoking household if any money was spent on cigarettes or tobacco. While this is a rather crude measure, it should at least give an approximation of the proportion of households containing smokers, assuming that all purchases made by the household were for household use and that these items were not received through other means. For the remaining countries, questions were asked in the surveys on whether the individual was a smoker or not. For these countries, analyses were confined to individuals aged >15 years. Because of the incorporation of proxy expenditure data, the quantitative results in this section should be treated cautiously and interpreted in the light of the literature review in section 4.

The sample sizes for tobacco use by subregion are shown in Table 24.38.

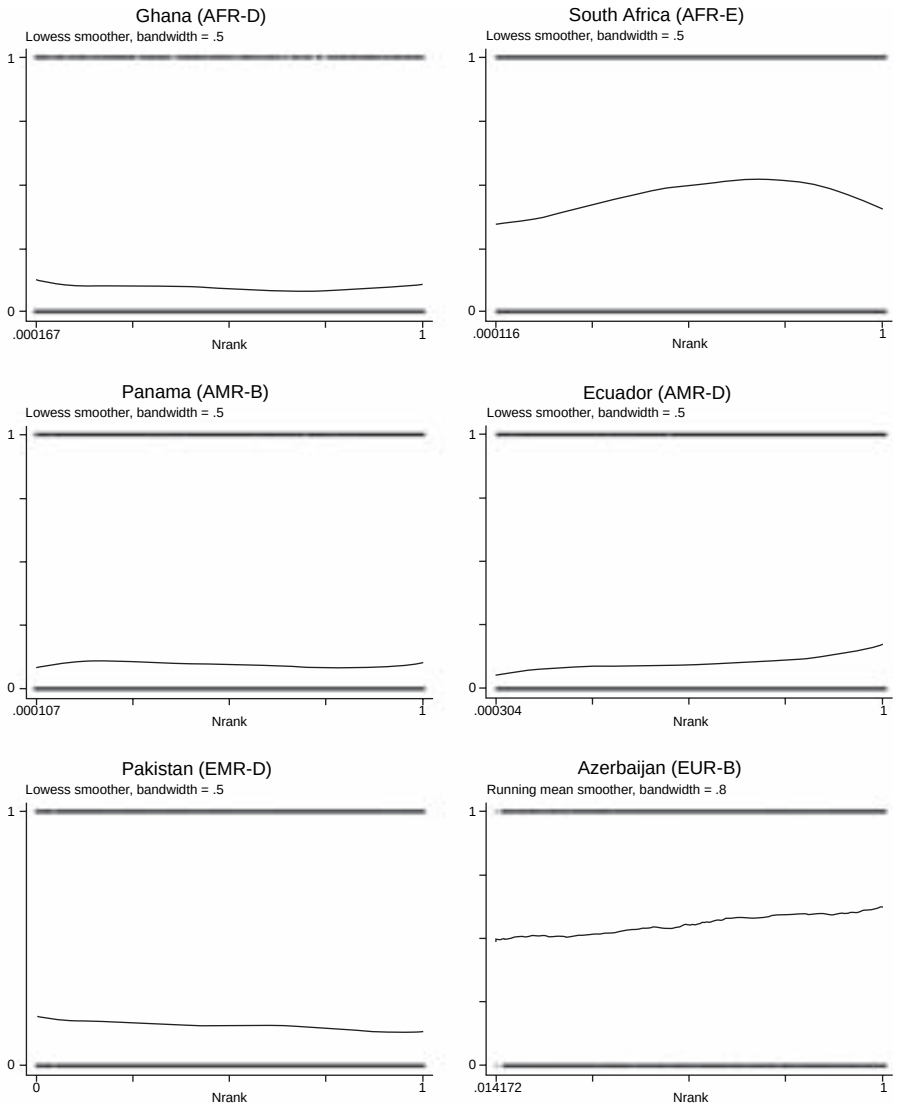
Locally linear kernel regression smooth plots of tobacco use according to normalized equivalized income ranking are shown in Figure 24.18. Each subfigure plots the proportion of tobacco use (y-axis ranging from 0 to 1) by normalized equivalized income rank for the country (x-axis ranging from 0 [poorest] to 1 [richest in the country]).

Table 24.38 Sample sizes for tobacco use by subregion

	Persons using tobacco		*Persons not using tobacco*	
Subregion	*n*	*Percentage*	*n*	*Percentage*
AFR-D	557	9	5441	91
AFR-E	3961	46	4644	54
AMR-B	1310	9	12669	91
AMR-D	1642	10	14823	90
EMR-B	—	—	—	—
EMR-D	2954	16	15956	84
EUR-B	2899	45	3580	55
EUR-C	3679	36	6653	64
SEAR-B	—	—	—	—
SEAR-D	—	—	—	—
WPR-B	2747	32	5870	68

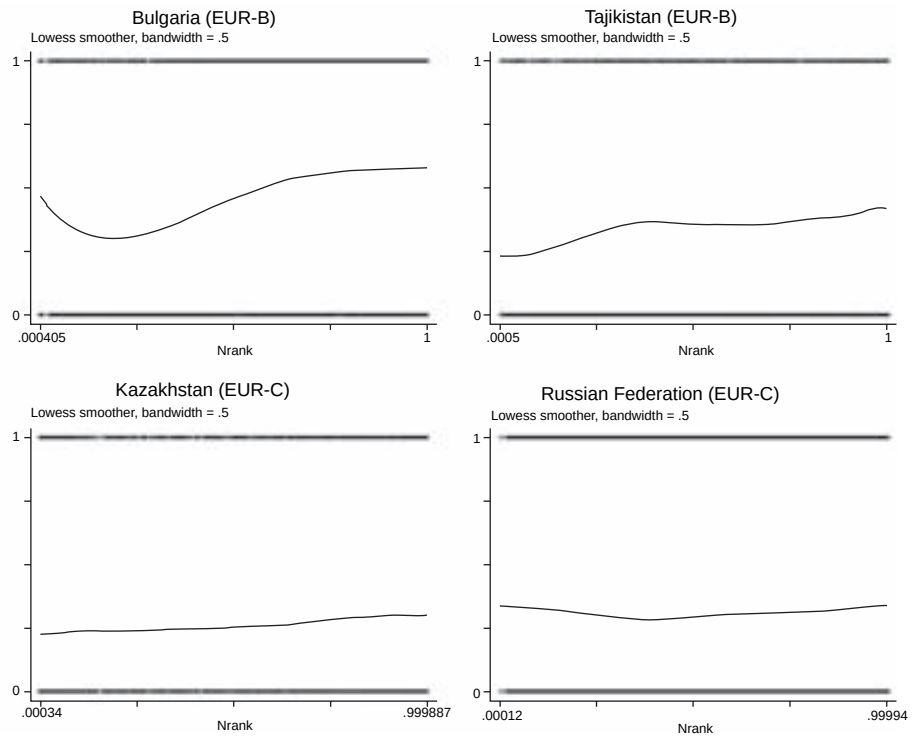
— No data.

Figure 24.18 Loess plots of the prevalence of tobacco use (*y*-axis) by normalized income rank (*x*-axis)



continued

Figure 24.18 Loess plots of the prevalence of tobacco use (y-axis) by normalized income rank (x-axis) (*continued*)



Full tabular results for US$2 and US$1 cut-offs are given in Tables 24.39 and 24.40, respectively, using the area under the curves in Figure 24.18.

Perhaps the most important result shown in Table 24.39 is the considerable variation between subregions in overall average tobacco use, but the relatively weak association of tobacco use within subregions by individual-level poverty. Thus the average prevalence of tobacco use in each subregion is a more useful predictor of individual tobacco use than absolute level of poverty.

Variation in the prevalence of tobacco use by poverty level within subregions was of secondary importance, and variable in direction. In AFR-E (South Africa data only) and AMR-D (Ecuador data only) there was a suggestion of lower prevalence of tobacco use among the poorest, and in EMR-B (Pakistan only) the converse. No reliable socioeconomic patterning of tobacco use was evident for other subregions. We emphasize that these are crude measures of tobacco use, possibly obscuring

Table 24.39 Cell prevalence of tobacco use by poverty, prevalence of tobacco use by poverty and relative risk of tobacco use by poverty, by subregion, for a US$ 2 per day cut-off

	Tobacco use				*Prevalence of tobacco use by poverty (%)*		*Relative risk of tobacco use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	9.2	46.3	*55.5*	<US$ 1/day	16.5	<US$ 1/day	1.1
	US$ 1–2/day	3.1	19.3	*22.4*	US$ 1–2/day	13.7	US$ 1–2/day	0.9
	>US$ 2/day	3.3	18.8	*22.1*	>US$ 2/day	14.7	US$ 2/day	0.9
		15.5	*84.5*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	5.2	22.1	*27.3*	<US$ 1/day	19.2	<US$ 1/day	0.7
	US$ 1–2/day	8.1	28.1	*36.2*	US$ 1–2/day	22.4	US$ 1–2/day	0.8
	>US$ 2/day	9.7	26.8	*36.5*	>US$ 2/day	26.5	US$ 2/day	0.9
		23.0	*77.0*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	3.4	7.6	*11.0*	<US$ 1/day	31.2	<US$ 1/day	1.1
	US$ 1–2/day	6.7	12.4	*19.1*	US$ 1–2/day	34.8	US$ 1–2/day	1.2
	>US$ 2/day	20.6	49.2	*69.8*	>US$ 2/day	29.5	US$ 2/day	1.2
		30.7	*69.3*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	3.3	14.1	*17.4*	<US$ 1/day	19.2	<US$ 1/day	0.6
	US$ 1–2/day	6.2	20.1	*26.3*	US$ 1–2/day	23.6	US$ 1–2/day	0.7
	>US$ 2/day	18.0	38.3	*56.3*	>US$ 2/day	31.9	US$ 2/day	0.8
		27.5	*72.5*	***100.0***			>US$ 2/day	1
EMR-D	<US$ 1/day	5.6	13.7	*19.3*	<US$ 1/day	29.2	<US$ 1/day	1.7
	US$ 1–2/day	11.2	34.1	*45.3*	US$ 1–2/day	24.8	US$ 1–2/day	1.4
	>US$ 2/day	6.2	29.1	*35.3*	>US$ 2/day	17.6	US$ 2/day	1.1
		23.1	*76.9*	***100.0***			>US$ 2/day	1
EUR-B	<US$ 1/day	0.9	2.1	*3.0*	<US$ 1/day	31.1	<US$ 1/day	0.8
	US$ 1–2/day	4.2	10.4	*14.6*	US$ 1–2/day	28.6	US$ 1–2/day	0.8
	>US$ 2/day	30.2	52.1	*82.3*	>US$ 2/day	36.7	US$ 2/day	0.8
		35.3	*64.7*	***100.0***			>US$ 2/day	1
EUR-C	<US$ 1/day	2.0	3.4	*5.4*	<US$ 1/day	37.1	<US$ 1/day	1.1
	US$ 1–2/day	7.5	13.8	*21.3*	US$ 1–2/day	35.4	US$ 1–2/day	1.0
	>US$ 2/day	25.2	48.1	*73.3*	>US$ 2/day	34.4	US$ 2/day	1.0
		34.8	*65.2*	***100.0***			>US$ 2/day	1
WPR-B	<US$ 1/day	6.4	11.5	*17.9*	<US$ 1/day	36.0	<US$ 1/day	1.0
	US$ 1–2/day	11.7	22.3	*34.0*	US$ 1–2/day	34.3	US$ 1–2/day	1.0
	>US$ 2/day	16.7	31.4	*48.1*	>US$ 2/day	34.6	US$ 2/day	1.0
		34.8	*65.2*	***100.0***			>US$ 2/day	1
Total[a]	<US$ 1/day	5.3	13.1	*18.4*	<US$ 1/day	28.7	<US$ 1/day	0.9
	US$ 1–2/day	9.2	20.8	*30.0*	US$ 1–2/day	30.6	US$ 1–2/day	1.0
	>US$ 2/day	16.4	35.1	*51.6*	>US$ 2/day	31.9	US$ 2/day	—
		30.9	*69.1*	***100.0***			>US$ 2/day	1

[a] Summary of the subregions in the table.

Table 24.40 Cell prevalence of tobacco use by poverty, prevalence of tobacco use by poverty and relative risk of tobacco use by poverty, by subregion, for a US$ 1 per day cut-off

		Tobacco use			*Prevalence of tobacco use by poverty (%)*		*Relative risk of tobacco use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	9.2	46.3	*55.5*	<US$ 1/day	16.5	<US$ 1/day	1.2
	>US$ 1/day	6.3	38.2	*44.5*	>US$ 1/day	14.2	US$ 1/day	1.1
		15.5	*84.5*	***100.0***			>US$ 1/day	1
AFR-E	<US$ 1/day	5.1	22.2	*27.3*	<US$ 1/day	18.6	<US$ 1/day	0.8
	>US$ 1/day	17.9	54.8	*72.7*	>US$ 1/day	24.6	US$ 1/day	0.8
		23.0	*77.0*	***100.0***			>US$ 1/day	1
AMR-B	<US$ 1/day	3.5	7.5	*11.0*	<US$ 1/day	31.4	<US$ 1/day	1.0
	>US$ 1/day	27.2	61.8	*89.0*	>US$ 1/day	30.6	US$ 1/day	1.1
		30.7	*69.3*	***100.0***			>US$ 1/day	1
AMR-D	<US$ 1/day	3.4	14.0	*17.4*	<US$ 1/day	19.6	<US$ 1/day	0.7
	>US$ 1/day	24.1	58.5	*82.6*	>US$ 1/day	29.2	US$ 1/day	0.8
		27.5	*72.5*	***100.0***			>US$ 1/day	1
EMR-D	<US$ 1/day	5.3	14.0	*19.3*	<US$ 1/day	27.6	<US$ 1/day	1.3
	>US$ 1/day	17.8	62.9	*80.7*	>US$ 1/day	22.0	US$ 1/day	1.2
		23.1	*76.9*	***100.0***			>US$ 1/day	1
EUR-B	<US$ 1/day	0.9	2.1	*3.0*	<US$ 1/day	31.2	<US$ 1/day	0.9
	>US$ 1/day	34.4	62.6	*97.0*	>US$ 1/day	35.5	US$ 1/day	0.9
		35.3	*64.7*	***100.0***			>US$ 1/day	1
EUR-C	<US$ 1/day	2.0	3.4	*5.4*	<US$ 1/day	37.1	<US$ 1/day	1.1
	>US$ 1/day	32.8	61.8	*94.6*	>US$ 1/day	34.7	US$ 1/day	1.1
		34.8	*65.2*	***100.0***			>US$ 1/day	1
WPR-B	<US$ 1/day	6.4	11.5	*17.9*	<US$ 1/day	36.0	<US$ 1/day	1.0
	>US$ 1/day	28.3	53.8	*82.1*	>US$ 1/day	34.5	US$ 1/day	1.0
		34.8	*65.2*	***100.0***			>US$ 1/day	1
Total[a]	<US$ 1/day	5.3	13.1	*18.4*	<US$ 1/day	28.6	<US$ 1/day	0.9
	>US$ 1/day	25.7	55.9	*81.6*	>US$ 1/day	31.4	US$ 1/day	—
		30.9	*69.1*	***100.0***			>US$ 1/day	1

[a] Summary of the subregions in the table.

important differences, for example between men and women or between age groups.

3.7 Alcohol use

Analyses for alcohol were based on LSMS data for nine countries in six subregions—AFR-D (Ghana), AFR-E (South Africa), AMR-B (Panama), AMR-D (Ecuador), EUR-B (Azerbaijan, Bulgaria, Tajikistan) and EUR-C (Kazakhstan, the Russian Federation). As for the data on tobacco (section 3.6), data for Bulgaria, Ghana, South Africa and Tajikistan were available only in the form of household expenditure data, and for Azerbaijan as a composite variable of combined expenditure on alcohol and tobacco. A household was classified as an alcohol-consuming household if any money was spent on alcohol. While this is a rather crude measure, it should at least give an approximation of the proportion of households containing consumers of alcohol, assuming that all purchases made by the household were for household use and that these items were not received through other means. For the remaining countries, questions were asked in the surveys on whether the individual consumed alcohol or not. For these countries, analyses were confined to individuals aged >15 years. As with the tobacco analyses, these quantitative results for alcohol should be treated cautiously and interpreted in the light of the literature review presented in section 4.

The sample sizes for alcohol use by subregion are shown in Table 24.41.

Locally linear kernel regression smooth plots of alcohol use according to normalized equivalized income ranking are shown in Figure 24.19.

Table 24.41 Sample sizes for alcohol use by subregion

	Persons using alcohol		*Persons not using alcohol*	
Subregion	*n*	*Percentage*	*n*	*Percentage*
AFR-D	3583	60	2415	40
AFR-E	3314	39	5282	61
AMR-B	3559	25	10414	75
AMR-D	4118	25	12345	75
EMR-B	—	—	—	—
EMR-D	—	—	—	—
EUR-B	2865	44	3614	56
EUR-C	4250	51	4071	49
SEAR-B	—	—	—	—
SEAR-D	—	—	—	—
WPR-B	2985	34	5674	66

— No data.

Figure 24.19 Loess plots of the prevalence of alcohol use (y-axis) by normalized income rank (y-axis)

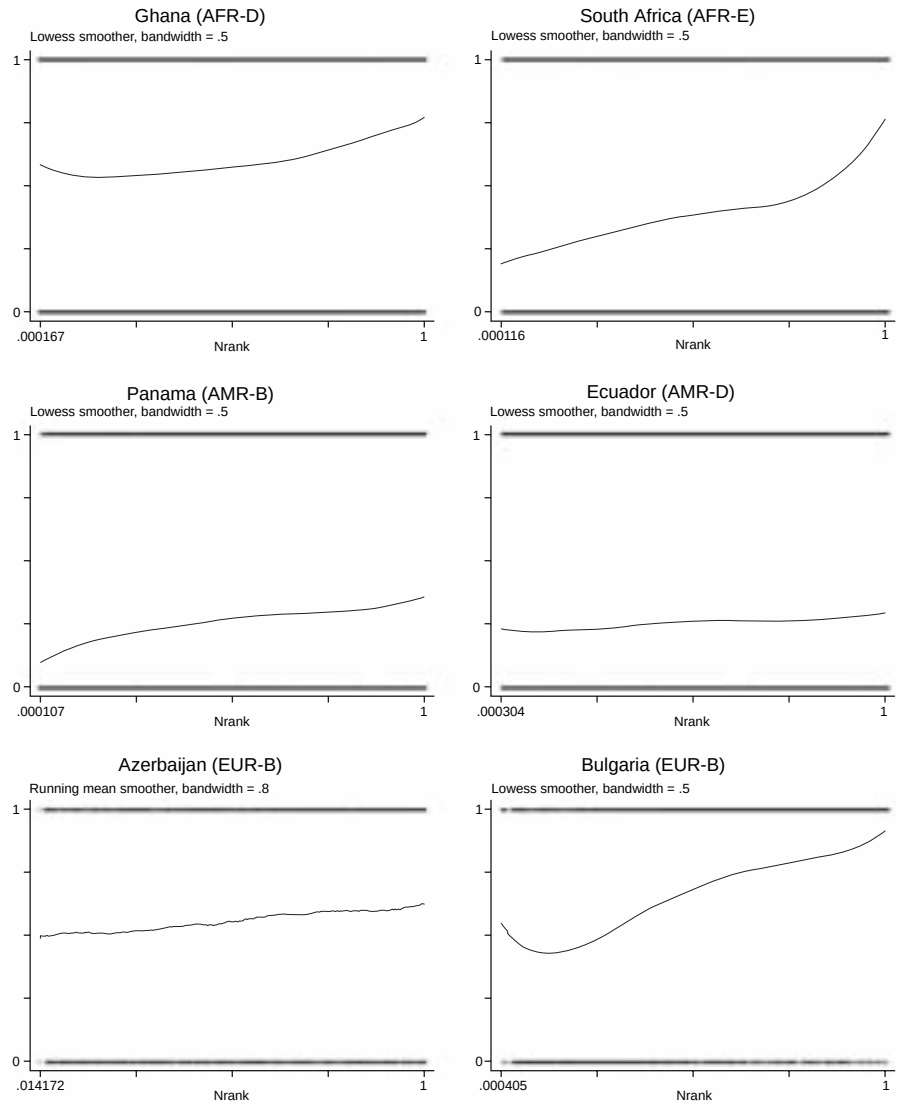
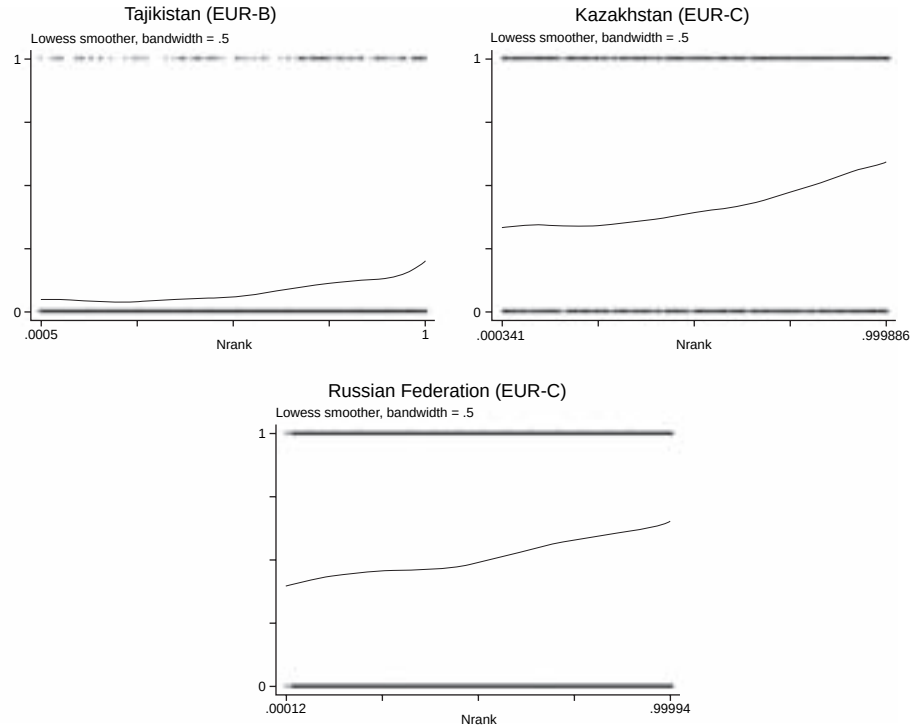


Figure 24.19 Loess plots of the prevalence of alcohol use (y-axis) by normalized income rank (y-axis) (*continued*)



Each subfigure plots the proportion of alcohol use (y-axis ranging from 0 to 1) by normalized equivalized income rank for the country (x-axis ranging from 0 [poorest] to 1 [richest in the country]).

Full tabular results for US$2 and US$1 cut-offs are given in Tables 24.42 and 24.43, respectively, based on the area under the curves in Figure 24.19. The results presented in Tables 24.42 and 24.43 demonstrate a more marked variation in overall average prevalence of alcohol use between subregions than within subregions by individual-level absolute poverty. In none of the subregions analysed was there a suggestion of increasing prevalence of alcohol use among the poorest. Also, in two subregions—AFR-E (South Africa data only) and AMR-B (Panama only)—poor people had approximately half the prevalence of alcohol use of non-poor people.

3.8 Overweight (women only)

For developing countries (excluding those in WPR-B) we used DHS data for maternal weight, and the body weight results reported here are

Table 24.42 Cell prevalence of alcohol use by poverty, prevalence of alcohol use by poverty and relative risk of alcohol use by poverty, by subregion, for a US$ 2 per day cut-off

	Alcohol use				*Prevalence of alcohol use by poverty (%)*		*Relative risk of alcohol use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	19.7	35.8	*55.5*	<US$ 1/day	35.5	<US$ 1/day	0.8
	US$ 1–2/day	8.6	13.8	*22.4*	US$ 1–2/day	38.3	US$ 1–2/day	0.8
	>US$ 2/day	10.1	12.0	*22.1*	>US$ 2/day	45.5	US$ 2/day	0.9
		38.3	*61.7*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	7.9	19.4	*27.3*	<US$ 1/day	28.9	<US$ 1/day	0.5
	US$ 1–2/day	14.0	22.2	*36.2*	US$ 1–2/day	38.7	US$ 1–2/day	0.6
	>US$ 2/day	22.1	14.4	*36.5*	>US$ 2/day	60.6	US$ 2/day	0.7
		44.0	*56.0*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	3.9	7.1	*11.0*	<US$ 1/day	35.7	<US$ 1/day	0.5
	US$ 1–2/day	9.9	9.2	*19.1*	US$ 1–2/day	51.6	US$ 1–2/day	0.7
	>US$ 2/day	52.4	17.4	*69.8*	>US$ 2/day	75.0	US$ 2/day	0.8
		66.2	*33.8*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	9.5	7.9	*17.4*	<US$ 1/day	54.8	<US$ 1/day	0.8
	US$ 1–2/day	15.8	10.5	*26.3*	US$ 1–2/day	60.2	US$ 1–2/day	0.9
	>US$ 2/day	37.1	19.2	*56.3*	>US$ 2/day	65.9	US$ 2/day	1.0
		62.5	*37.5*	***100.0***			>US$ 2/day	1
EUR-B	<US$ 1/day	1.5	1.5	*3.0*	<US$ 1/day	51.3	<US$ 1/day	0.8
	US$ 1–2/day	6.9	7.7	*14.6*	US$ 1–2/day	47.6	US$ 1–2/day	0.7
	>US$ 2/day	53.4	28.9	*82.3*	>US$ 2/day	64.9	US$ 2/day	0.7
		61.9	*38.1*	***100.0***			>US$ 2/day	1
EUR-C	<US$ 1/day	3.7	1.7	*5.4*	<US$ 1/day	67.9	<US$ 1/day	0.8
	US$ 1–2/day	15.6	5.7	*21.3*	US$ 1–2/day	73.2	US$ 1–2/day	0.8
	>US$ 2/day	65.3	8.0	*73.3*	>US$ 2/day	89.1	US$ 2/day	0.8
		84.5	*15.5*	***100.0***			>US$ 2/day	1
WPR-B	<US$ 1/day	9.2	8.7	*17.9*	<US$ 1/day	51.4	<US$ 1/day	0.8
	US$ 1–2/day	18.5	15.5	*34.0*	US$ 1–2/day	54.5	US$ 1–2/day	0.9
	>US$ 2/day	29.8	18.3	*48.1*	>US$ 2/day	62.0	US$ 2/day	0.9
		57.6	*42.4*	***100.0***			>US$ 2/day	1
Total[a]	<US$ 1/day	8.1	10.2	*18.3*	<US$ 1/day	44.2	<US$ 1/day	0.7
	US$ 1–2/day	15.0	13.5	*28.5*	US$ 1–2/day	52.7	US$ 1–2/day	0.8
	>US$ 2/day	35.9	17.2	*53.1*	>US$ 2/day	67.6	US$ 2/day	—
		59.0	*41.0*	***100.0***			>US$ 2/day	1

[a] Summary of the subregions in the table.

Table 24.43 Cell prevalence of alcohol use by poverty, prevalence of alcohol use by poverty and relative risk of alcohol use by poverty, by subregion, for a US$ 1 per day cut-off

		Alcohol use			*Prevalence of alcohol use by poverty (%)*		*Relative risk of alcohol use*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	19.8	35.7	*55.5*	<US$ 1/day	35.7	<US$ 1/day	0.9
	>US$ 1/day	18.5	26.0	*44.5*	>US$ 1/day	41.6	US$ 1/day	0.9
		38.3	*61.7*	***100.0***			>US$ 1/day	1
AFR-E	<US$ 1/day	7.3	20.0	*27.3*	<US$ 1/day	26.8	<US$ 1/day	0.5
	>US$ 1/day	36.7	36.0	*72.7*	>US$ 1/day	50.5	US$ 1/day	0.6
		44.0	*56.0*	***100.0***			>US$ 1/day	1
AMR-B	<US$ 1/day	3.9	7.1	*11.0*	<US$ 1/day	35.2	<US$ 1/day	0.5
	>US$ 1/day	62.3	26.7	*89.0*	>US$ 1/day	70.0	US$ 1/day	0.6
		66.2	*33.8*	***100.0***			>US$ 1/day	1
AMR-D	<US$ 1/day	9.6	7.8	*17.4*	<US$ 1/day	55.1	<US$ 1/day	0.9
	>US$ 1/day	52.9	29.7	*82.6*	>US$ 1/day	64.1	US$ 1/day	0.9
		62.5	37.5	***100.0***			>US$ 1/day	1
EUR-B	<US$ 1/day	1.5	1.5	*3.0*	<US$ 1/day	51.6	<US$ 1/day	0.8
	>US$ 1/day	60.3	36.7	*97.0*	>US$ 1/day	62.2	US$ 1/day	0.8
		61.9	*38.1*	***100.0***			>US$ 1/day	1
EUR-C	<US$ 1/day	3.6	1.8	*5.4*	<US$ 1/day	67.5	<US$ 1/day	0.8
	>US$ 1/day	80.9	13.7	*94.6*	>US$ 1/day	85.5	US$ 1/day	0.8
		84.5	*15.5*	***100.0***			>US$ 1/day	1
WPR-B	<US$ 1/day	9.2	8.7	*17.9*	<US$ 1/day	51.5	<US$ 1/day	0.9
	>US$ 1/day	48.3	33.8	*82.1*	>US$ 1/day	58.9	US$ 1/day	0.9
		57.6	*42.4*	***100.0***			>US$ 1/day	1
Total[a]	<US$ 1/day	8.1	10.3	*18.3*	<US$ 1/day	43.9	<US$ 1/day	0.7
	>US$ 1/day	51.0	30.7	*81.7*	>US$ 1/day	62.4	US$ 1/day	—
		59.0	*41.0*	***100.0***			>US$ 1/day	1

[a] Summary of the subregions in the table.

Table 24.44 Sample sizes for overweight mothers by subregion (DHS data)

Subregion	*n*
AFR-D	38 005
AFR-E	43 880
AMR-B	15 177
AMR-D	29 398
EMR-B	—
EMR-D	3 314
EUR-B	10 722
EUR-C	—
SEAR-B	—
SEAR-D	8 187
WPR-B	4 469

— No data.

therefore for mothers only. Overweight was defined as a body mass index (BMI) of between 25 and 29.9 kg/m^2. The overall proportion of overweight women within each subregion was estimated using data from chapter 8.

For WPR-B, for which we had no applicable DHS data set, prevalence information was obtained using the CHNS data set. To make the WPR-B values comparable with those obtained using the DHS data, analysis of body weight was restricted to adult women (aged >15 years) only. As shown in Table 24.44, there were sufficient sample sizes available for analysis in all subregions except EMR-B, EUR-C and SEAR-B. Because all analyses were limited to women with children aged <5 years and do not correct for pregnancy, the results must be treated with caution.

Full tabular results for US$ 2 and US$ 1 cut-offs are given in Tables 24.45 and 24.46, respectively. For both African subregions and SEAR-D (dominated by India) the prevalence of overweight among poor women was less than half that among non-poor women (Table 24.45). For WPR-B (dominated by China) there was no discernible socioeconomic gradient. For the remaining subregions (Central and South American countries, and the south-eastern European and central Asian countries) there was a tendency for overweight to be about 20% less common among poor women than among non-poor women.

3.9 Results summary—relative risks and risk differences

Relative risks

Table 24.47 shows the relative risks for the prevalence of each risk factor for people living on <US$ 1, US$ 1–2 and exactly US$ 2 per day

Table 24.45 Cell prevalence of overweight by poverty, prevalence of overweight by poverty and relative risk of overweight by poverty, by subregion, for a US$ 2 per day cut-off (women only)

		Overweight			*Prevalence of overweight by poverty (%)*		*Relative risk of overweight*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	5.4	50.1	*55.5*	<US$ 1/day	9.8	<US$ 1/day	0.4
	US$ 1–2/day	4.0	18.4	*22.4*	US$ 1–2/day	17.9	US$ 1–2/day	0.8
	>US$ 2/day	5.1	17.0	*22.1*	>US$ 2/day	22.9	US$ 2/day	0.9
		14.5	*85.5*	***100.0***			>US$ 2/day	1
AFR-E	<US$ 1/day	4.0	23.3	*27.3*	<US$ 1/day	14.8	<US$ 1/day	0.4
	US$ 1–2/day	9.8	26.4	*36.2*	US$ 1–2/day	27.2	US$ 1–2/day	0.7
	>US$ 2/day	14.2	22.3	*36.5*	>US$ 2/day	38.9	US$ 2/day	0.8
		28.1	*71.9*	***100.0***			>US$ 2/day	1
AMR-B	<US$ 1/day	2.5	8.5	*11.0*	<US$ 1/day	22.3	<US$ 1/day	0.8
	US$ 1–2/day	5.1	14.0	*19.1*	US$ 1–2/day	26.9	US$ 1–2/day	1.0
	>US$ 2/day	19.3	50.5	*69.8*	>US$ 2/day	27.6	US$ 2/day	1.0
		26.9	*73.1*	***100.0***			>US$ 2/day	1
AMR-D	<US$ 1/day	4.5	12.9	*17.4*	<US$ 1/day	26.0	<US$ 1/day	0.8
	US$ 1–2/day	7.9	18.4	*26.3*	US$ 1–2/day	29.9	US$ 1–2/day	0.9
	>US$ 2/day	17.8	38.5	*56.3*	>US$ 2/day	31.7	US$ 2/day	1.0
		30.2	*69.8*	***100.0***			>US$ 2/day	1
EMR-D	<US$ 1/day	2.3	17.0	*19.3*	<US$ 1/day	12.2	<US$ 1/day	0.7
	US$ 1–2/day	6.4	38.9	*45.3*	US$ 1–2/day	14.1	US$ 1–2/day	0.8
	>US$ 2/day	6.1	29.2	*35.3*	>US$ 2/day	17.3	US$ 2/day	0.9
		14.9	*85.1*	***100.0***			>US$ 2/day	1
EUR-B	<US$ 1/day	0.9	2.1	*3.0*	<US$ 1/day	29.8	<US$ 1/day	0.8
	US$ 1–2/day	4.7	9.9	*14.6*	US$ 1–2/day	32.4	US$ 1–2/day	0.8
	>US$ 2/day	32.0	50.3	*82.3*	>US$ 2/day	38.9	US$ 2/day	1.0
		37.6	*62.4*	***100.0***			>US$ 2/day	1
SEAR-D	<US$ 1/day	0.0	42.4	*42.4*	<US$ 1/day	0.1	<US$ 1/day	0.4
	US$ 1–2/day	0.1	42.7	*42.8*	US$ 1–2/day	0.2	US$ 1–2/day	0.7
	>US$ 2/day	0.0	14.8	*14.8*	>US$ 2/day	0.2	US$ 2/day	0.9
		0.1	*99.9*	***100.0***			>US$ 2/day	1
WPR-B	<US$ 1/day	3.9	14.0	*17.9*	<US$ 1/day	21.8	<US$ 1/day	1.1
	US$ 1–2/day	6.7	27.3	*34.0*	US$ 1–2/day	19.6	US$ 1–2/day	1.0
	>US$ 2/day	9.6	38.5	*48.1*	>US$ 2/day	20.1	US$ 2/day	1.1
		20.2	*79.8*	***100.0***			>US$ 2/day	1
Total[a]	<US$ 1/day	2.6	23.1	*25.7*	<US$ 1/day	9.9	<US$ 1/day	0.4
	US$ 1–2/day	4.7	29.2	*33.9*	US$ 1–2/day	13.8	US$ 1–2/day	0.6
	>US$ 2/day	9.2	31.2	*40.4*	>US$ 2/day	22.8	US$ 2/day	—
		16.0	*84.0*	***100.0***			>US$ 2/day	1

[a] Summary of the subregions in the table.

Table 24.46 Cell prevalence of overweight by poverty, prevalence of overweight by poverty and relative risk of overweight by poverty, by subregion, for a US$ 1 per day cut-off (women only)

		Overweight			*Prevalence of overweight by poverty (%)*		*Relative risk of overweight*	
Subregion		*Yes*	*No*					
AFR-D	<US$ 1/day	5.5	50.0	*55.5*	<US$ 1/day	10.0	<US$ 1/day	0.5
	>US$ 1/day	8.9	35.6	*44.5*	>US$ 1/day	20.1	US$ 1/day	0.8
		14.5	*85.5*	***100.0***			>US$ 1/day	1
AFR-E	<US$ 1/day	4.0	23.3	*27.3*	<US$ 1/day	14.7	<US$ 1/day	0.4
	>US$ 1/day	24.1	48.6	*72.7*	>US$ 1/day	33.1	US$ 1/day	0.6
		28.1	*71.9*	***100.0***			>US$ 1/day	1
AMR-B	<US$ 1/day	2.5	8.5	*11.0*	<US$ 1/day	22.3	<US$ 1/day	0.8
	>US$ 1/day	*24.4*	*64.6*	*89.0*	>US$ 1/day	27.4	US$ 1/day	0.9
		26.9	*73.1*	***100.0***			>US$ 1/day	1
AMR-D	<US$ 1/day	4.5	12.9	17.4	<US$ 1/day	26.1	<US$ 1/day	0.8
	>US$ 1/day	25.7	56.9	82.6	>US$ 1/day	31.1	US$ 1/day	0.9
		30.2	*69.8*	***100.0***			>US$ 1/day	1
EMR-D	<US$ 1/day	2.3	17.0	*19.3*	<US$ 1/day	12.0	<US$ 1/day	0.8
	>US$ 1/day	12.5	68.2	*80.7*	>US$ 1/day	15.5	US$ 1/day	0.8
		14.9	*85.1*	***100.0***			>US$ 1/day	1
EUR-B	<US$ 1/day	0.9	2.1	*3.0*	<US$ 1/day	29.4	<US$ 1/day	0.8
	>US$ 1/day	36.7	60.3	*97.0*	>US$ 1/day	37.9	US$ 1/day	0.8
		37.6	*62.4*	***100.0***			>US$ 1/day	1
SEAR-D	<US$ 1/day	0.0	42.4	*42.4*	<US$ 1/day	0.1	<US$ 1/day	0.5
	>US$ 1/day	0.1	57.5	*57.6*	>US$ 1/day	0.2	US$ 1/day	0.6
		0.1	*99.9*	***100.0***			>US$ 1/day	1
WPR-B	<US$ 1/day	3.6	14.3	*17.9*	<US$ 1/day	20.4	<US$ 1/day	1.0
	>US$ 1/day	16.6	65.5	*82.1*	>US$ 1/day	20.2	US$ 1/day	1.0
		20.2	*79.8*	***100.0***			>US$ 1/day	1
Total[a]	<US$ 1/day	2.5	23.2	*25.7*	<US$ 1/day	9.6	<US$ 1/day	0.5
	>US$ 1/day	14.0	60.3	*74.3*	>US$ 1/day	18.8	US$ 1/day	—
		16.0	*84.0*	***100.0***			>US$ 1/day	1

[a] Summary of the subregions in the table.

Table 24.47 Summary of relative risks by poverty: reference category >US$2 per day

Subregion		*Unimproved water and/or sanitation*	*Underweight (low weight-for-age)*	*Non-marital sex (men)*	*Non-marital sex (women)*	*Condom use (men)*	*Condom use (women)*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Body weight (women)*
AFR-D	<US$1/day	9.4	2.3	0.6	0.4	0.5	0.3	1.1	1.1	0.8	0.4
	US$1–2/day	2.9	1.4	0.9	0.9	0.9	0.7	1.1	0.9	0.8	0.8
	US$2/day	1.6	1.2	0.9	0.9	1.0	0.9	1.1	0.9	0.9	0.9
	>US$2/day	1	1	1	1	1	1	1	1	1	1
AFR-E	<US$1/day	4.6	2.6	0.9	0.6	0.3	0.1	2.0	0.7	0.5	0.4
	US$1–2/day	3.5	1.8	1.1	0.8	0.7	0.4	1.9	0.8	0.6	0.7
	US$2/day	2.7	1.4	1.0	1.0	0.8	0.6	1.7	0.9	0.7	0.8
	>US$2/day	1	1	1	1	1	1	1	1	1	1
AMR-B	<US$1/day	12.3	2.4	0.7	0.8	0.6	0.3	7.2	1.1	0.5	0.8
	US$1–2/day	5.7	1.8	0.9	0.6	0.8	0.4	4.7	1.2	0.7	1.0
	US$2/day	3.4	1.6	1.0	0.7	0.9	0.6	3.3	1.2	0.8	1.0
	>US$2/day		1	1	1	1	1	1	1	1	1
AMR-D	<US$1/day	8.9	3.7	0.7	0.7	0.4	0.3	14.6	0.6	0.8	0.8
	US$1–2/day	4.0	2.1	0.9	1.0	0.7	0.3	4.7	0.7	0.9	0.9
	US$2/day	2.3	1.6	1.0	1.0	0.8	0.5	2.3	0.8	1.0	1.0
	>US$2/day	1	1	1	1	1	1	1	1	1	1
EMR-B	<US$1/day	3.6	2.1	—	—	—	—	—	—	—	—
	US$1–2/day	3.2	1.9	—	—	—	—	—	—	—	—
	US$2/day	2.9	1.8	—	—	—	—	—	—	—	—
	>US$2/day	1	1	—	—	—	—	—	—	—	—

continued

Table 24.47 Summary of relative risks by poverty: reference category >US $2 per day (*continued*)

Subregion		Unimproved water and/or sanitation	Underweight (low weight-for-age)	Non-marital sex (men)	Non-marital sex (women)	Condom use (men)	Condom use (women)	Indoor air pollution	Tobacco use	Alcohol use	Body weight (women)
EMR-D	<US$ 1/day	15.1	1.7	—	—	—	—	4.0	1.7	—	0.7
	US$ 1–2/day	7.2	1.7	—	—	—	—	3.1	1.4	—	0.8
	US$ 2/day	3.2	1.5	—	—	—	—	1.7	1.1	—	0.9
	>US$ 2/day	1	1	—	—	—	—	1	1	—	1
EUR-B	<US$ 1/day	3.1	1.9	—	—	—	—	1.2	0.8	0.8	0.8
	US$ 1–2/day	2.7	1.6	—	—	—	—	1.2	0.8	0.7	0.8
	US$ 2/day	2.3	1.3	—	—	—	—	1.2	0.8	0.7	1.0
	>US$ 2/day	1	1	—	—	—	—	1	1	1	1
EUR-C	<US$ 1/day	11.8	2.4	—	0.4	—	1.0	1.3	1.1	0.8	—
	US$ 1–2/day	8.4	2.1	—	0.4	—	1.0	1.4	1.0	0.8	—
	US$ 2/day	5.4	1.8	—	0.5	—	1.0	1.3	1.0	0.8	—
	>US$ 2/day	1	1	—	1	—	1	1	1	1	—
SEAR-B	<US$ 1/day	2.0	3.3	—	—	—	—	—	—	—	—
	US$ 1–2/day	1.7	2.2	—	—	—	—	—	—	—	—
	US$ 2/day	1.4	1.7	—	—	—	—	—	—	—	—
	>US$ 2/day	1	1	—	—	—	—	—	—	—	—

SEAR-D	<US$ 1/day	5.0	2.1	—	—	—	—	3.5	—	—	0.4
	US$ 1–2/day	4.1	1.7	—	—	—	—	2.9	—	—	0.7
	US$ 2/day	2.1	1.3	—	—	—	—	1.8	—	—	0.9
	>US$ 2/day	1	1	—	—	—	—	1	—	—	1
WPR-B	<US$ 1/day	1.7	1.1	—	2.4	—	—	1.4	1.0	0.8	1.1
	US$ 1–2/day	1.3	1.4	—	1.9	—	—	1.4	1.0	0.9	1.0
	US$ 2/day	1.0	1.3	—	1.5	—	—	1.3	1.0	0.9	1.1
	>US$ 2/day	1	1	—	1	—	—	1	1	1	1
Total (crude)[a]	<US$ 1/day	3.3	3.1	0.7	0.6	0.4	0.3	2.2	0.9	0.7	0.4
	US$ 1–2/day	2.4	2.4	0.9	0.7	0.7	0.5	1.9	1.0	0.8	0.6
	US$ 2/day	—	—	—	—	—	—	—	—	—	—
	>US$ 2/day	1	1	1	1	1	1	1	1	1	1
Total (pooled)[b]	<US$ 1/day	7.9	2.5	0.7	0.6	0.5	0.2	3.8	1.1	0.7	0.5
	US$ 1–2/day	4.2	1.8	0.9	0.9	0.8	0.5	2.8	1.0	0.8	0.8
	US$ 2/day	—	—	—	—	—	—	—	—	—	—
	>US$ 2/day	1	1	1	1	1	1	1	1	1	1

— No data.

[a] Total refers to the 11 out of 14 subregions included in this report. The crude "total" estimate is derived by summing the estimated number of people in each subregion with each risk factor within each poverty stratum, then recalculating the relative risks at this total level. It is crude in so far as it does not allow for confounding at the subregional level of the relative risk association.

[b] Total refers to the 11 out of 14 subregions included in this report. Unlike the crude estimate, the pooled estimate uses Mantel-Haenszel weights to pool the relative risks across subregions. The actual DHS or LSMS sample sizes are used to calculate the Mantel-Haenszel weights.

compared to the baseline of >US$ 2 per day. The equivalent summary relative risks for the dichotomous poverty variable are shown in Table 24.48.

The only difference from the relative risk results presented previously is the inclusion of both a crude and a pooled "total" summary relative risk for the 11 subregions included in this chapter. The crude total relative risk is simply the estimated relative risk comparing people by poverty stratum, regardless of the subregion in which they live. This relative risk estimate is crude in the sense that it does not allow for confounding by subregion; nevertheless, it is easily understood.

The "pooled" relative risk is a Mantel-Haenszel pooled relative risk across subregions. The actual sample sizes for each subregion are used to create the weights. This total relative risk estimate seeks to summarize the within-subregion associations, and will vary from the crude total estimate if there is independent variation across subregions in both poverty levels and risk factor prevalence.

It should be noted that there is marked heterogeneity between subregions in the association of income poverty with most of the risk factors considered here. The association of income poverty with lack of water and sanitation is stronger in AMR-B, AFR-D, EMR-D and EUR-C, and that with indoor air pollution is stronger in AMR-B and AMR-D. Thus, the pooled relative risk estimates for the total developing world should be treated cautiously. Nevertheless, comparing the crude and pooled total relative risks in Table 24.47 reveals substantial differences for unimproved water and sanitation and indoor air pollution. In both these instances, the pooled relative risk is greater than the crude. Thus, there are (on average) stronger associations of poverty with these two risk factors within subregions than those suggested by a crude global analysis. Put another way, subregion confounds the association of individual income poverty with unsafe water and sanitation, since more poor subregions have higher prevalences of these two risk factors independent of personal income poverty.

Risk differences

All the summary statistics thus far in this chapter are relative risks. However, as the prevalence of risk factors varies by subregion, a constant relative risk across subregions will correspond to differences in the absolute difference in prevalence of risk factor between poverty levels. Tables 24.49 and 24.50 present risk difference estimates by subregion for the three-level and two-level poverty estimates, respectively.

For ease of interpretation, the risk differences are presented as the difference between the poor and the referent group. For example, it was estimated that the prevalence of underweight among 0–4 year olds living on <US$ 1 per day in AFR-D was 22.6% greater than that among 0–4 year olds living on >US$ 2 per day (Table 24.49). The actual risk factor prevalence in the referent group is also given. Thus, the prevalence of

Table 24.48 Summary of relative risks by poverty: reference category >US$ 1 per day

Subregion		*Unimproved water and/or sanitation*	*Underweight (low weight-for-age)*	*Non-marital sex (men)*	*Non-marital sex (women)*	*Condom use (men)*	*Condom use (women)*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Body weight (women)*
AFR-D	<US$ 1/day	4.2	1.8	0.6	0.5	0.6	0.4	1.1	1.2	0.9	0.5
	US$ 1/day	2.1	1.3	0.9	0.9	0.9	0.7	1.1	1.1	0.9	0.8
	>US$ 1/day	1	1	1	1	1	1	1	1	1	1
AFR-E	<US$ 1/day	1.9	1.7	0.8	0.7	0.4	0.1	1.6	0.8	0.5	0.4
	US$ 1/day	1.7	1.5	1.1	0.8	0.7	0.3	1.6	0.8	0.6	0.6
	>US$ 1/day	1	1	1	1	1	1	1	1	1	1
AMR-B	<US$ 1/day	6.2	2.1	0.7	0.9	0.7	0.3	4.5	1.0	0.5	0.8
	US$ 1/day	4.8	1.7	0.8	0.8	0.7	0.4	3.8	1.1	0.6	0.9
	>US$ 1/day	1	1	1	1	1	1	1	1	1	1
AMR-D	<US$ 1/day	4.4	2.7	0.7	0.7	0.4	0.3	5.9	0.7	0.9	0.8
	US$ 1/day	3.5	2.0	0.8	0.9	0.6	0.4	3.5	0.8	0.9	0.9
	>US$ 1/day	1	1	1	1	1	1	1	1	1	1
EMR-B	<US$ 1/day	3.0	1.9	—	—	—	—	—	—	—	—
	US$ 1/day	2.9	1.9	—	—	—	—	—	—	—	—
	>US$ 1/day	1	1	—	—	—	—	—	—	—	—
EMR-D	<US$ 1/day	3.3	1.2	—	—	—	—	1.5	1.3	—	—
	US$ 1/day	2.8	1.2	—	—	—	—	1.5	1.2	—	—
	>US$ 1/day	1	1	—	—	—	—	1	1	—	—
EUR-B	<US$ 1/day	2.2	1.7	—	—	—	—	1.2	0.9	0.8	0.8
	US$ 1/day	2.2	1.6	—	—	—	—	1.2	0.9	0.8	0.8
	>US$ 1/day	1	1	—	—	—	—	1	1	1	1
EUR-C	<US$ 1/day	5.7	2.1	—	0.4	—	1.0	1.2	1.1	0.8	—
	US$ 1/day	5.6	2.1	—	0.4	—	1.0	1.3	1.1	0.8	—
	>US$ 1/day	1	1	—	1	—	1	1	1	1	—

continued

Table 24.48 Summary of relative risks by poverty: reference category >US$ 1 per day (*continued*)

Subregion		*Unimproved water and/or sanitation*	*Underweight (low weight-for-age)*	*Non-marital sex (men)*	*Non-marital sex (women)*	*Condom use (men)*	*Condom use (women)*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Body weight (women)*
SEAR-B	<US$ 1/day	1.5	2.1	—	—	—	—	2.6	—	—	—
	US$ 1/day	1.5	2.0	—	—	—	—	2.1	—	—	—
	>US$ 1/day	1	1	—	—	—	—	1	—	—	—
SEAR-D	<US$ 1/day	1.5	1.4	—	—	—	—	1.5	—	—	0.5
	US$ 1/day	1.5	1.3	—	—	—	—	1.4	—	—	0.6
	>US$ 1/day	1	1	—	—	—	—	1	—	—	1
WPR-B	<US$ 1/day	2.3	0.9	—	—	—	—	1.2	1.0	0.9	1.0
	US$ 1/day	2.3	1.0	—	—	—	—	1.2	1.0	0.9	1.0
	>US$ 1/day	1	1	—	—	—	—	1	1	1	
Total (crude)[a]	<US$ 1/day	2.4	1.9	0.7	0.7	0.5	0.3	1.5	0.9	0.7	0.5
	US$ 1/day	—	—	—	—	—	—	—	—	—	—
	>US$ 1/day	1	1	1	1	1	1	1	1	1	1
Total (pooled)[b]	<US$ 1/day	3.0	1.7	0.7	0.6	0.5	0.3	1.6	1.0	0.8	0.6
	US$ 1/day	—	—	—	—	—	—	—	—	—	—
	>US$ 1/day	1	1	1	1	1	1	1	1	1	1

— No data.

[a] Total refers to the 11 out of 14 subregions included in this report. The crude "total" estimate is derived by summing the estimated number of people in each subregion with each risk factor within each poverty stratum, then recalculating the relative risks at this total level. It is crude in so far as it does not allow for confounding at the subregional level of the relative risk association.

[b] Total refers to the 11 out of 14 subregions included in this report. Unlike the crude estimate, the pooled estimate uses Mantel-Haenszel weights to pool the relative risks across subregions. The actual DHS or LSMS sample sizes are used to calculate the Mantel-Haenszel weights.

Table 24.49 Absolute percentage differences in risk factor prevalence by subregion by three-level poverty variable: reference group >US$ 2 per day prevalence

Subregion		*Underweight (low weight-for-age)*	*Unimproved water and/or sanitation*	*Non-marital sex (men)*	*Non-marital sex (women)*	*Condom use (men)*	*Condom use (women)*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Body weight (women)*
AFR-D	<US$ 1/day	22.6	63.2	–23.7	–12.3	–21.4	–22.2	9.5	1.8	–10.0	–13.1
	US$ 1–2/day	7.8	14.2	–7.8	–3.2	–4.0	–9.0	7.6	–1.0	–7.2	–5.0
	US$ 2/day	3.3	4.7	–3.7	–1.8	–2.0	–4.4	4.8	–1.7	–4.0	–1.9
	>US$ 2/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 2/day**	**17.9**	**7.6**	**53.6**	**22.2**	**46.9**	**31.5**	**66.4**	**14.7**	**45.5**	**22.9**
AFR-E	<US$ 1/day	28.2	72.1	–4.6	–7.5	–35.9	–31.7	46.7	–7.3	–31.6	–24.2
	US$ 1–2/day	14.9	49.6	2.9	–3.7	–16.8	–20.1	46.7	–4.1	–21.9	–11.7
	US$ 2/day	7.4	33.4	–1.7	–0.9	–9.7	–13.5	37.3	–1.6	–15.5	–6.9
	>US$ 2/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 2/day**	**17.9**	**20.0**	**36.0**	**19.0**	**52.9**	**34.8**	**53.3**	**26.5**	**60.6**	**38.9**
AMR-B	<US$ 1/day	5.5	87.1	–13.3	–3.6	–21.0	–29.9	64.2	1.7	–39.3	–5.3
	US$ 1–2/day	3.2	36.5	–6.1	–7.1	–10.5	–23.6	37.8	5.4	–23.4	–0.7
	US$ 2/day	2.5	18.5	–2.3	–6.3	–3.3	–18.0	24.2	4.8	–17.3	0.2
	>US$ 2/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 2/day**	**3.8**	**7.7**	**47.1**	**19.7**	**59.9**	**42.5**	**10.3**	**29.5**	**75.0**	**27.6**
AMR-D	<US$ 1/day	18.8	79.8	–15.7	–1.2	–31.4	–45.1	87.8	–12.7	–11.2	–5.6
	US$ 1–2/day	7.9	30.7	–5.1	0.1	–14.4	–40.3	44.6	–8.3	–5.7	–1.8
	US$ 2/day	4.1	12.8	–2.1	–0.2	–8.2	–30.4	16.1	–6.7	–1.4	0.2
	>US$ 2/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

continued

Table 24.49 Absolute percentage differences in risk factor prevalence by subregion by three-level poverty variable: reference group >US$ 2 per day prevalence (*continued*)

Subregion		*Underweight (low weight-for-age)*	*Unimproved water and/or sanitation*	*Non-marital sex (men)*	*Non-marital sex (women)*	*Condom use (men)*	*Condom use (women)*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Body weight (women)*
Reference prevalence[a]	**>US$ 2/day**	**7.0**	**10.1**	**47.3**	**4.1**	**50.7**	**60.1**	**12.2**	**31.9**	**65.9**	**31.7**
EMR-B	<US$ 1/day	7.8	35.1	—	—	—	—	—	—	—	
	US$ 1–2/day	6.8	29.8	—	—	—	—	—	—	—	—
	US$ 2/day	6.0	26.0	—	—	—	—	—	—	—	—
	>US$ 2/day	0.0	0.0	—	—	—	—	—	—	—	—
Reference prevalence[a]	**>US$ 2/day**	**7.5**	**13.7**	**—**	**—**	**—**	**—**	**—**	**—**	**—**	**—**
EMR-D	<US$ 1/day	12.6	72.7	—	—	—	—	66.2	11.5	—	–5.2
	US$ 1–2/day	12.5	32.0	—	—	—	—	45.3	7.2	—	–3.2
	US$ 2/day	7.7	11.2	—	—	—	—	15.0	2.2	—	–1.1
	>US$ 2/day	0.0	0.0	—	—	—	—	0.0	0.0	—	0.0
Reference prevalence[a]	**>US$ 2/day**	**17.0**	**5.2**	**—**	**—**	**—**	**—**	**21.9**	**17.6**	**—**	**17.3**
EUR-B	<US$ 1/day	6.3	32.4	—	—	—	—	9.3	–5.6	–13.6	–9.1
	US$ 1–2/day	4.0	25.9	—	—	—	—	9.6	–8.2	–17.3	–6.5
	US$ 2/day	2.0	19.4	—	—	—	—	8.9	–8.2	–16.6	–1.8
	>US$ 2/day	0.0	0.0	—	—	—	—	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 2/day**	**6.8**	**15.4**	**—**	**—**	**—**	**—**	**39.9**	**36.7**	**64.9**	**38.9**

EUR-C	<US$ 1/day	2.8	20.2	—	–9.7	—	–0.9	6.3	2.7	–21.1	—
	US$ 1–2/day	2.2	14.0	—	–8.6	—	–0.4	8.3	1.0	–15.8	—
	US$ 2/day	1.7	8.2	—	–7.6	—	–0.5	6.6	–0.9	–13.7	—
	>US$ 2/day	0.0	0.0	—	0.0	—	0.0	0.0	0.0	0.0	—
Reference prevalence[a]	**>US$ 2/day**	**2.0**	**1.9**	**—**	**15.4**	**—**	**22.6**	**20.7**	**34.4**	**89.1**	**—**
SEAR-B	<US$ 1/day	34.9	21.2	—	—	—	—	—	—	—	—
	US$ 1–2/day	19.0	15.4	—	—	—	—	—	—	—	—
	US$ 2/day	10.8	9.2	—	—	—	—	—	—	—	—
	>US$ 2/day	0.0	0.0	—	—	—	—	—	—	—	—
Reference prevalence[a]	**>US$ 2/day**	**15.5**	**22.1**	**—**	**—**	**—**	**—**	**—**	**—**	**—**	**—**
SEAR-D	<US$ 1/day	28.0	64.7	—	—	—	—	70.8	—	—	–0.1
	US$ 1–2/day	18.6	50.0	—	—	—	—	54.1	—	—	–0.1
	US$ 2/day	7.0	17.0	—	—	—	—	22.5	—	—	0.0
	>US$ 2/day	0.0	0.0	—	—	—	—	0.0	—	—	0.0
Reference prevalence[a]	**>US$ 2/day**	**26.1**	**16.0**	**—**	**—**	**—**	**—**	**29.2**	**—**	**—**	**0.2**
WPR-B	<US$ 1/day	1.1	31.0	—	—	—	—	28.5	1.4	–10.7	1.7
	US$ 1–2/day	5.5	16.1	—	—	—	—	25.6	–0.3	–7.6	–0.5
	US$ 2/day	4.0	2.3	—	—	—	—	18.4	–0.7	–4.4	3.0
	>US$ 2/day	0.0	0.0	—	—	—	—	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 2/day**	**13.9**	**47.1**	**—**	**—**	**—**	**—**	**64.3**	**34.6**	**62.0**	**20.1**

— No data.

[a] The actual percentage prevalence of the risk factor in the >US$ 2 per day reference category.

Table 24.50 Absolute difference in risk factor prevalence by subregion by two-level poverty variable: reference group >US$ 1 per day prevalence

Subregion		*Underweight (low weight-for-age)*	*Unimproved water and/or sanitation*	*Non-marital sex (men)*	*Non-marital sex (women)*	*Condom use (men)*	*Condom use (women)*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Body weight (women)*
AFR-D	<US$ 1/day	17.3	53.1	−18.9	−10.4	−19.0	−17.0	5.1	2.4	−5.9	−10.1
	US$ 1/day	7.5	17.4	−6.8	−2.6	−4.7	−7.6	4.7	1.5	−4.9	−4.7
	>US$ 1/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 1/day**	**22.6**	**16.4**	**49.2**	**20.4**	**44.7**	**26.5**	**70.6**	**14.2**	**41.6**	**20.1**
AFR-E	<US$ 1/day	19.2	41.7	−6.2	−5.4	−27.0	−21.4	19.5	−6.0	−77.3	−18.4
	US$ 1/day	13.6	32.6	3.7	−3.8	−15.4	−16.1	49.6	−5.3	−80.2	−13.8
	>US$ 1/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 1/day**	**25.8**	**46.2**	**37.5**	**17.1**	**44.4**	**24.7**	**80.5**	**24.6**	**50.5**	**33.1**
AMR-B	<US$ 1/day	4.8	79.7	−11.9	−2.1	−18.8	−24.8	62.1	0.9	−34.8	−5.1
	US$ 1/day	3.3	59.0	−9.4	−4.1	−15.8	−20.7	49.7	3.9	−26.6	−2.9
	>US$ 1/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 1/day**	**4.5**	**15.5**	**45.7**	**18.2**	**57.6**	**37.4**	**17.8**	**30.6**	**70.0**	**27.4**
AMR-D	<US$ 1/day	16.0	68.5	−14.1	−1.2	−26.7	−32.0	57.0	−9.6	−9.0	−5.0
	US$ 1/day	9.5	50.1	−8.0	−0.3	−18.4	−26.3	100.0	−6.2	−8.4	−3.8
	>US$ 1/day	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 1/day**	**9.6**	**20.2**	**45.7**	**4.1**	**46.1**	**47.3**	**43.0**	**29.2**	**64.1**	**31.1**

EMR-B	<US$ 1/day	7.2	32.0	—	—	—	—	—	—	—	—
	US$ 1/day	7.0	30.9	—	—	—	—	—	—	—	—
	>US$ 1/day	0.0	0.0	—	—	—	—	—	—	—	—
Reference prevalence[a]	**>US$ 1/day**	**8.0**	**15.9**	**—**	**—**	**—**	**—**	**—**	**—**	**—**	**—**
EMR-D	<US$ 1/day	5.5	54.1	—	—	—	—	27.1	5.6	—	–3.5
	US$ 1/day	4.6	41.7	—	—	—	—	23.7	3.6	—	–3.4
	>US$ 1/day	0.0	0.0	—	—	—	—	0.0	0.0	—	0.0
Reference prevalence[a]	**>US$ 1/day**	**24.0**	**23.3**	**—**	**—**	**—**	**—**	**50.0**	**22.0**	**—**	**15.5**
EUR-B	<US$ 1/day	5.0	23.1	—	—	—	—	8.3	–4.3	–10.6	–8.4
	US$ 1/day	4.8	22.5	—	—	—	—	8.6	–4.8	–11.5	–8.0
	>US$ 1/day	0.0	0.0	—	—	—	—	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 1/day**	**7.4**	**19.5**	**—**	**—**	**—**	**—**	**41.3**	**35.5**	**62.2**	**37.9**
EUR-C	<US$ 1/day	2.7	22.2	—	–8.1	—	–0.8	4.7	2.5	–18.0	—
	US$ 1/day	2.6	21.7	—	–8.0	—	–0.7	7.5	2.2	–16.0	—
	>US$ 1/day	0.0	0.0	—	0.0	—	0.0	0.0	0.0	0.0	—
Reference prevalence[a]	**>US$ 1/day**	**2.5**	**4.7**	**—**	**13.4**	**—**	**22.5**	**22.5**	**34.7**	**85.5**	**—**
SEAR-B	<US$ 1/day	26.1	14.1	—	—	—	—	35.8	—	—	—
	US$ 1/day	23.1	13.7	—	—	—	—	70.7	—	—	—
	>US$ 1/day	0.0	0.0	—	—	—	—	0.0	—	—	—
Reference prevalence[a]	**>US$ 1/day**	**24.1**	**29.1**	**—**	**—**	**—**	**—**	**64.2**	**—**	**—**	**—**

continued

Table 24.50 Absolute difference in risk factor prevalence by subregion by two-level poverty variable: reference group >US$ 1 per day prevalence (*continued*)

Subregion		*Underweight (low weight-for-age)*	*Unimproved water and/or sanitation*	*Non-marital sex (men)*	*Non-marital sex (women)*	*Condom use (men)*	*Condom use (women)*	*Indoor air pollution*	*Tobacco use*	*Alcohol use*	*Body weight (women)*
SEAR-D	<US$ 1/day	14.2	27.7	—	—	—	—	28.6	—	—	–0.1
	US$ 1/day	10.8	27.6	—	—	—	—	30.4	—	—	–0.1
	>US$ 1/day	0.0	0.0	—	—	—	—	0.0	—	—	0.0
Reference prevalence[a]	**>US$ 1/day**	**39.9**	**53.1**	**—**	**—**	**—**	**—**	**71.4**	**—**	**—**	**0.2**
WPR-B	<US$ 1/day	–1.2	62.6	—	—	—	—	17.4	1.5	–7.4	0.2
	US$ 1/day	–0.2	61.1	—	—	—	—	18.1	0.1	–6.9	0.6
	>US$ 1/day	0.0	0.0	—	—	—	—	0.0	0.0	0.0	0.0
Reference prevalence[a]	**>US$ 1/day**	**16.2**	**46.9**	**—**	**—**	**—**	**75.0**	**34.5**	**58.9**	**20.2**	

— No data.

[a] The actual percentage prevalence of the risk factor in the >US$ 2 per day reference category.

underweight among children living on >US$2 per day in AFR-D was 17.9%. Accordingly the prevalence of underweight among children living on <US$1 per day can be deduced from the data in Table 24.49, i.e. 22.6+17.9 = 40.5%.

While the relative risks across subregions for underweight were somewhat similar, the risk differences by poverty varied markedly given the different overall prevalences of underweight between subregions. There were also notable differences in the risk differences across subregions for unimproved water and sanitation and for indoor air pollution.

Finally, the negative risk differences correspond to those risk factors where the prevalence among the poor was less than that among the non-poor.

4. Systematic literature reviews

4.1 Tobacco use

Literature on developing countries

Smoking associated with low socioeconomic status (SES)

The International Clinical Epidemiology Network (INCLEN) study collected data on risk factors for cardiovascular disease in men from 12 centres in seven developing countries (INCLEN 1994). It found significant SES trends (using education, occupation or income) for increased smoking among men of low SES in two centres in China and one in rural Thailand. No significant associations were obtained from the centres in Brazil, Chile, Colombia, Indonesia and the Philippines.

China. Marked gradients in smoking prevalence for men and women (increasing prevalence with lower levels of educational attainment) have been described in an urban Chinese population (Yu et al. 2000). Over the period 1989–1996 the gradient became less steep in men (owing largely to an increase in smoking by better educated men), whereas in women it changed little over the same period. This study also showed that less educated men smoked more cigarettes per day, but that this gap also decreased over the period 1989–1996. Another urban study revealed that those with lower and middle levels of education (particularly men but also women) had the highest prevalence of smoking (Koong et al. 1991). In contrast, a cohort study of urban male workers reported no association of SES with smoking prevalence, but daily cigarette consumption was significantly greater among men with low educational attainment (Siegrist et al. 1990). Surveys in three provinces of rural China also showed an increased prevalence of smoking among adults with low educational attainment (Hu and Tsai 2000). Also, among urban adolescents, prevalence of smoking was found to be associated with lower parental SES (Zhu et al. 1996).

South Asia. Low educational attainment was found to be associated with smoking in men and women in urban populations in India (Narayan et al. 1996; Singh et al. 1998). One of these studies also reported that the smoking of beedi or chutta was associated with a low level of education and with manual occupations (Narayan et al. 1996). Another study in India also reported this finding for beedi smoking, and that smoking this type of cigarette (compared to other cigarettes) was independently associated with a higher frequency of respiratory symptoms and poorer lung function on clinical testing (Chhabra et al. 2001). In rural populations in India, the same pattern of increased smoking prevalence with lower SES was described for educational attainment in both men and women (Gupta et al. 1994) and for social class in men (Singh et al. 1997a) (with the smoking prevalence for women in this latter study being very low). Similarly, a national survey in Pakistan found that, among both males and females, illiteracy was strongly associated with higher smoking rates (Alam 1998). Also, among Bangladeshi male adolescents, a much higher rate of smoking was reported among slum dwellers with no formal education than among students (Ahsan et al. 1998).

Other Asian countries. In the Republic of Korea, smoking prevalence was found to be associated with low educational attainment in men but not in women (women had a very low smoking prevalence) (Chung et al. 1992). This was also the case for men in Cambodia (Smith et al. 1998). Studies of rural populations in Malaysia found the prevalence of smokeless tobacco use to be significantly higher among less educated women (Gan 1995, 1998). In the latter study, however, education was not associated with prevalence of tobacco smoking.

Middle East. In Saudi Arabian adults in three regions of the country, smoking prevalence was higher among uneducated people and among certain occupations of low SES such as manual workers (Jarallah et al. 1999). Similarly, in Bahrain (Hamadeh and Musaiger 2000) and Kuwait (Memon et al. 2000) low educational attainment was associated with increased smoking prevalence in adults.

Other countries. In urban Brazil, increased smoking prevalence was associated with low educational attainment in men and low social class in women (Duncan et al. 1993). Similarly, it was associated with more poor and uneducated adults in Nigeria (Obot 1990), men and women with low educational attainment in Tonga (Woodward et al. 1994) and men with low educational attainment on a Caribbean island (the results for women were not statistically significant) (Grol et al. 1997b). In the latter study, however, men and women of high SES who did smoke consumed significantly more cigarettes per day.

Smoking associated with high SES

A gradient of increased smoking prevalence with higher SES was described in an urban study in Brazil for women of high social class (with no significant gradient among men and in contrast to the other Brazilian study described above) (Martins et al. 1995). This pattern was also described for adolescent smoking in the United Arab Emirates (Bener and al-Ketbi 1999) (based on paternal educational attainment and family income) and Ghana (for high SES homes, though no precise data were provided) (Amonoo-Lartson and Pappoe 1992).

Smoking not associated with SES

As mentioned above, the INCLEN study (INCLEN 1994) found no significant associations between smoking prevalence and SES from the centres in Brazil, Chile, Colombia, Indonesia and the Philippines. Similar findings were obtained from surveys of villagers in India (Singh et al. 1997b) and of a town and village population in India (using level of educational attainment) (Chaturvedi et al. 1998). In this latter study, however, there was a higher rate of smoking among the unemployed than among the employed. Surveys of women in five Indian cities also found no significant gradient for smoking prevalence (Singh et al. 1999).

Literature on developed countries

Smoking associated with low SES

A review by Stellman and Resnicow (1997) found that in most developed countries, the prevalence of cigarette smoking is currently higher in low-SES groups. However, they note that in some of these countries smoking had been more prevalent among the higher social classes during the first half of the twentieth century. Studies examined in this review that reported an occupational class gradient by smoking prevalence (increased smoking among low-SES groups) come from Australia (with gradients for men and women); the United Kingdom of Great Britain and Northern Ireland (based on multiple surveys); and the United States of America (three studies cited with gradients for both men and women). The same pattern was seen for SES gradients determined by educational attainment in studies from France; Italy and Spain (for men under 65 years only); the United Kingdom; and the United States (seven studies cited). In the French study it was reported that less educated men much preferred nonfilter cigarettes manufactured with black tobacco. Similarly, there are survey data from the United States showing that better educated men tended to smoke cigarettes with lower tar yields.

The same pattern was seen when considering SES as determined by income in Australia (for men and women), the United Kingdom (based on housing data to reflect SES) (Stellman and Resnicow 1997) and the United States (owing principally to smoking cessation among the more affluent).

Similarly, the INTERSALT study involving 52 centres in 18 developed and 12 developing countries (ICRG 1988) found that smoking prevalence was related to low educational attainment (Stamler et al. 1992). Also, a review of smoking among adolescents identified four studies in which low SES was associated with increased smoking prevalence in developed countries (Canada, Finland, New Zealand and Norway) (Tyas and Pederson 1998).

Other data not detailed in the above-mentioned reviews also conform to this pattern of increased smoking prevalence by low-SES groups for:

- adults in the United Kingdom, based on occupational class for men and women (Acheson 1998);
- young people leaving school in the United Kingdom, for educational disadvantage independent of parental SES and parental smoking status (Glendinning et al. 1994; Green et al. 1991);
- students in the United States, for school performance and smoking prevalence and attempts to quit smoking (Hu et al. 1998);
- pregnant women in the United States, by educational attainment (Pamuk et al. 1998);
- adult New Zealanders, based on national data from the census (Borman et al. 1999; Crampton et al. 2000; Davis et al. 1999; Hay and Foster 1984), national health surveys (Howden-Chapman and Tobias 2000; Ministry of Health 1999) and other community surveys (Jackson et al. 1990; Klemp et al. 1998; Statistics New Zealand/ Ministry of Health 1993; Whitlock et al. 1997); and
- pregnant women in New Zealand (occupational class) (Fergusson et al. 1998).

Furthermore, low-SES adults who do not smoke were found to suffer from increased exposure to second-hand smoke (Whitlock et al. 1998).

Tobacco-related disease and low SES

Stellman and Resnicow (1997) reported that "a large body of evidence confirms the inverse association of lung cancer and social class in many developed countries". These authors noted that in some situations it was "satisfactorily demonstrated" that such gradients were attributable to social class gradients in tobacco use. In general, however, the SES patterning of lung cancer cannot be fully explained by SES patterns of tobacco use (given some likely additional role for various dietary and occupational risk factors and access to health care). This review provided data for the following countries, all of which demonstrated increased tobacco-related disease rates among low-SES groups:

- Denmark, where there was a very strong occupational class gradient (lung and bladder cancer) in men and women;

- England and Wales, for cancer incidence and mortality;
- Italy, with a very strong relationship between educational attainment and various tobacco-related cancers (but not lung cancer); and
- New Zealand, where smoking patterns explained much of the increased mortality risk for social classes III and IV.

Other data from New Zealand also suggest that SES gradients for tobacco-related disease reflect SES gradients for smoking rates (higher smoking rates with lower SES) (Kawachi et al. 1991; Pearce 1997). Indeed, the latest analysis on smoking and inequality in New Zealand suggests that "at least one-third of the deprivation gradient in life expectancy at birth or older ages, one-quarter of the corresponding ethnic disparity, and one-fifth of the corresponding gender gap, is accounted for by tobacco consumption" (Ministry of Health 2001).

Smoking associated with high SES

The review by Stellman and Resnicow (1997) reported that in the United Kingdom in the late 1940s and early to mid-1950s, smoking was more common among the higher social classes (based on data from multiple surveys). This review also reported on Spanish data indicating that smoking prevalence was relatively higher among college-educated men aged >65 years. Also, there was a very strong association with higher levels of education and smoking among younger Spanish women (in marked contrast to the data for men) (1987 data). More recent data from Spain suggest that the gap in smoking cessation rates by educational level has become wider over time (i.e. cessation rates are higher for those with higher levels of education) (Fernandez et al. 2001). This pattern was also apparent for women with higher educational levels. Similarly, other data from Spain suggest that the quitting rate is higher among women aged 25–44 years with non-manual occupations compared to those with manual occupations (Regidor et al. 2001). A review of smoking among adolescents identified studies in five developed countries (Iceland, Japan, New Zealand, Sweden and the United Kingdom) in which higher personal income was associated with higher levels of smoking (Tyas and Pederson 1998).

Smoking not associated with SES

One review identified two studies that reported non-significant effects of parental education on adolescent smoking in Canada and the United States (Tyas and Pederson 1998). However, these studies examined maternal education only. In the United Kingdom there was little occupational class gradient in the proportion of children who had ever smoked (although some gradient existed for higher average consumption of cigarettes with lower occupational class) (Acheson 1998). It has been suggested that smoking in adolescence is an indicator of occupational

class of destination rather than occupational class of origin (Glendinning et al. 1994).

Discussion

The relationship

In general in developing countries, lower SES was associated with higher prevalence of smoking. The association was generally stronger for men. There was also a much lower smoking prevalence among women in most developing countries. Nevertheless, some studies in the developing world actually show the opposite pattern, with high-SES adults and adolescents having higher smoking rates, while other studies show no association. This overall pattern contrasts with that for other cardiovascular disease risk factors in developing countries (cholesterol level, obesity, blood pressure and to some extent physical inactivity). This might reflect the low cost of tobacco products in many developing countries and the influence of advertising and the mass media in promoting their use. Nevertheless, there is a need for more historical data on smoking in developing countries to better interpret this pattern.

In developed countries the gradient of increased smoking with lower SES is even more predominant. There are still some exceptions, however, such as for adolescent smoking in some settings and for Spanish women. Furthermore, the SES gradient for tobacco-related disease and overall mortality is likely to be partly explained by the SES gradient for tobacco use.

In both developing and developed countries there is some evidence that the type of tobacco product smoked by low-SES adults may also be more hazardous (e.g. beedis in India) than that consumed by high-SES adults. Similarly, the number of cigarettes consumed daily may be higher among low-SES adults.

Possible mechanism

Several factors are likely to be involved in the relatively increased prevalence of smoking by lower-SES groups.

- Cultural effects relating to health values and risk perception may be relevant. Similarly, neighbourhood effects may be important, such as the influence of disadvantaged neighbourhoods on smoking in men (though not women) described in the United States by Ross (2000).
- Parental influences on adolescent smoking may possibly be more important in low-SES populations (e.g. in countering the benefits of education about the risks of smoking).
- Access to cheap tobacco in tobacco-producing rural areas of the developing world may be relevant.

- Lower educational attainment may lead to poorer knowledge of the hazards of smoking and of how to obtain support for quitting.
- The relative lack of restrictions on smoking at the workplace for many blue-collar workers could reduce incentives to quit (compared to those in higher occupational classes).
- Lower rates of smoking cessation in low-SES groups might be related to a greater burden of psychosocial stress and the effect of others smoking.

Advertising by the tobacco industry is probably an important determinant of smoking trends (Stellman and Resnicow 1997), particularly the aggressive marketing in developing countries (Connolly 1992; Mackay 1992).

4.2 Hazardous alcohol use

Literature on developing countries

Hazardous alcohol use associated with low SES

In China, one cohort study of urban male workers reported that high levels of daily alcohol consumption were significantly more prevalent among men with low educational attainment (Siegrist et al. 1990). Similarly, a Brazilian study of urban residents (using a logistic regression model) found that both heavy drinking and alcohol dependence were associated with low educational attainment and low income (Moreira et al. 1996). Another study in a Brazilian city found that low social class was significantly associated with excessive alcohol consumption in both men and women (Duncan et al. 1993).

A study in a Caribbean island reported that men and women of low SES (based on educational attainment) drank more alcohol per week (Grol et al. 1997b). However, high-SES women were more likely to be regular drinkers than low-SES women. Of the two African studies identified, one in Ethiopia found that problem drinking (based on the CAGE screening instrument) was associated with a lower level of educational attainment among residents of Addis Ababa (Kebede and Alem 1999). Similarly, in a Nigerian study, alcohol consumption was higher in those with low SES (Bunker et al. 1992). In Brazil, the results of one study suggested that higher rates of oesophageal cancer among lower-SES groups are likely to be attributable to the higher use of sugar-cane spirit, black tobacco and mate in these groups (Bouchardy et al. 1993). In contrast, data from Colombia indicate that the incidence of alcohol-related cancers tended to show positive social class gradients (Cuello et al. 1982).

Hazardous alcohol use associated with high SES

In an Ethiopian study, problem drinking (based on the CAGE screening instrument) was associated with higher income and education among men or women in a rural district (Alem et al. 1999) (in contrast to another Ethiopian study described above). In Thailand, a study reported that hazardous alcohol use (based on AUDIT scores) was associated with higher income among those seeking emergency treatment (Lapham et al. 1998). By contrast, the study found that those with a university degree were significantly *less* likely to have positive AUDIT scores. A study in the Republic of Korea found a trend towards greater alcohol consumption with increasing years of education among women, but no such trend among men (Chung et al. 1992). Alcohol intake among high-SES males explained a small but statistically significant part of the relationship between hypertension and high SES in villagers in North India (Singh et al. 1997b). This suggests that at least some members of this population were drinking fairly large quantities of alcohol.

No association with SES

A study in a metropolitan area in Brazil found no statistically significant differences in alcohol use associated with SES, although there was a trend towards higher rates of drinking among low-SES men (Martins et al. 1995).

Literature on developed countries

Hazardous alcohol use associated with low SES

A major review of alcohol drinking, social class and cancer suggested a very likely role for alcohol drinking in the observed negative social class gradients for alcohol-related cancers in men in France, Italy and New Zealand (Moller and Tonnesen 1997). Evidence that was less strong, but still suggestive of such a role, was also reported for men in Denmark, Switzerland and the United Kingdom. This review identified studies showing associations between hazardous alcohol use and low SES in Denmark (for average number of drinks per week in men and women); Finland (for rates of alcohol intoxication and rates of hospital admission for acute alcohol-related conditions for men and women); France (for prevalence of "heavy drinkers" in men); Sweden (alcoholism and high alcohol consumption in young people); Switzerland (for alcoholism); the United Kingdom (for the proportion of heavy drinkers in manual vs non-manual occupations among men); and the United States (for heavy consumption—five or more drinks on one occasion in men and women).

The major review by Colhoun et al. (1998) on hypertension and SES reported that higher alcohol consumption by low-SES men explains part, though not all, of the association between SES and blood pressure in men in developed countries. This collectively considered the findings of

studies in Australia, Israel, the Netherlands, Norway, the United Kingdom (two studies) and the United States (two studies). Similarly, the INTERSALT study, involving 52 centres in 18 developed and 12 developing countries, found that adjusting for alcohol (along with BMI, smoking, and sodium and potassium excretion) halved the association between low SES and higher blood pressure in men so that it was no longer statistically significant (Stamler et al. 1992).

A combined analysis of population samples from France, Italy, Spain and Switzerland (Pequignot et al. 1988) found that daily alcohol consumption was higher in male manual workers than in male professionals, whereas in women there was no significant variation by occupational group. More recent data from the United Kingdom indicated that 10% of low-SES men (social classes IV and V) were dependent on alcohol compared to 5% of high-SES men (social classes I and II) (Acheson 1998). Similarly, in the United States, heavy alcohol use was reported to be higher among those with poorer education among men and women in virtually all ethnic groups (Pamuk et al. 1998).

Data from surveys in 48 American states showed that hazardous consumption of alcohol was significantly more frequent in households with below-median income and in those with a lower educational level (for five or more drinks on one occasion at least once a week during the previous year) (Midanik and Clark 1994). For weekly drinking, however, the pattern was the opposite, being higher in those with above-median income and a higher level of education.

In a national survey in New Zealand based on AUDIT scores, lower social class was associated with an increased risk of a hazardous pattern of alcohol use (Howden-Chapman and Tobias 2000). There was some suggestion that lower educational attainment was associated with a more hazardous pattern of alcohol use, but this was not statistically significant. Earlier New Zealand work indicated that the pattern of drinking "high quantities but less often" was more common in the lower social classes, along with an increased prevalence of high daily consumption (Casswell and Gordon 1984).

Hazardous alcohol use associated with high SES

The review by Moller and Tonnesen (1997) reported a French study indicating that high-SES women were more likely to be regular consumers of alcohol, and that in Sweden high alcohol consumption was associated with high SES among older people. The review by Colhoun et al. (1998) on hypertension and SES reported that higher alcohol consumption among high-SES women had been identified in studies in Australia and England (two studies) and in the INTERSALT study, where some of 52 centres collected data on alcohol.

A study in Wales revealed a tendency towards a higher prevalence of consumption of 22 drinks or more per week in the higher social classes (Farrow et al. 1988). A study of three regions in the United Kingdom

reported that weekly alcohol consumption increased with household income in both men and women (Crawford 1988), but that manual workers drank more alcohol than non-manual workers. In the United States, a combined analysis of 10 surveys showed some positive gradient for the frequency of alcohol drinking with SES (Knupfer 1989). However, there was a weakly inverse gradient for "frequent drunkenness". Among employees in Minneapolis-St Paul, the frequency of alcohol drinking was associated with high SES in women but not in men (Jeffery et al. 1991). A national survey in New Zealand (Statistics New Zealand/Ministry of Health 1993) reported a higher alcohol intake with greater educational attainment, in contrast to the other results from New Zealand detailed above.

No association with SES

The review by Colhoun et al. (1998) reported on two British studies that showed no SES gradient with alcohol consumption. Similarly, there was no association with social class among a sample of attendees at a London health centre (King 1986). In New Zealand, a child cohort study found that family SES did not have a marked impact on drinking behaviour, but that those from low-SES families in this cohort drank more at age 15 years (Connolly 1992).

The international review of alcohol drinking, social class and cancer by Moller and Tonnesen (1997) found no evidence for such an association for Japan, except that women in the highest social classes had slightly elevated mortality from liver cirrhosis. There was a similar finding for Finland and for Sweden, except that pharyngeal cancer in Swedish women was more prevalent in low-SES women. Nevertheless, these two countries do show some evidence for SES gradients in hazardous alcohol use (as detailed above).

Discussion

The relationship

The general pattern in developed countries is for lower-SES men to have a more hazardous pattern of alcohol consumption in terms of high or excessive intake of alcohol. This pattern is also evident in the distribution of alcohol-related disease, particularly alcohol-related cancers. The pattern for women can resemble that for men, although there are also many studies indicating excessive alcohol use among high-SES women. Different studies in the same countries sometimes show contrasting patterns that may reflect temporal trends, regional variation and different methodological approaches.

The pattern is also mixed in developing countries, although in this review more studies were identified that indicated that lower SES was associated with a more hazardous pattern of alcohol consumption. This pattern is similar to that found for tobacco use, but it does contrast with

the SES gradient seen for other disease risk factors where high SES is associated with increased risk (e.g. for cholesterol level, obesity, blood pressure and to some extent physical inactivity).

Possible mechanisms

Some of the factors likely to be involved in the excessive use of alcohol by lower-SES groups include:

- cultural influences, including the presence of greater numbers of alcohol users in certain areas and the social acceptability of heavy drinking;
- the existence of psychological stressors associated with poverty and inequality from which alcohol users may seek to escape;
- access to cheap alcohol, especially in rural areas of the developing world and where alcohol is not taxed; and
- the effect of lower educational attainment on poorer knowledge of the hazards of excessive alcohol use.

High levels of alcohol use among higher-SES groups (especially women) may reflect in particular cultural patterns and the role of additional income in settings where alcohol is relatively expensive. The alcohol industry may also target advertising at those with the highest incomes. It is also plausible that alcohol is used as a coping mechanism by high-SES women who have heavy demands on their time.

Some of the studies in this review also reported higher levels of abstinence among lower-SES groups. This behavioural pattern may have health benefits in those aged <50 years (e.g. in terms of injury prevention), but abstinence could be considered a potential risk factor when considering the evidence for cardioprotective effects of moderate alcohol consumption in older populations (Rimm et al. 1996).

4.3 Illicit drug use

Literature on developing countries

Substance use associated with poverty

In an Indian city, substance abuse by adolescents (covering tobacco, alcohol and cannabis) was more common among those from Hindu families with low educational status and low family income (Kushwaha et al. 1992). Similarly, in Nepal, a small qualitative study suggested that urban poverty was a factor in substance use (Jutkowitz et al. 1997). In Brazil, youths living on the streets were significantly (eight times) more likely to use drugs (generally inhalants and marijuana) than those living at home (Pinto et al. 1994). Other work has documented relatively heavy drug use by street children in Brazilian cities, such as half of 7–8-year-old street children in São Paulo (Carlini 1990). A study of military con-

scripts in Kuwait found that amphetamine use was associated with lower educational attainment and unemployment (Bilal et al. 1992).

Substance use associated with high SES

A survey of students in an Indian city found that regular cannabis users "were mainly from professional colleges, hailing from metropolitan cities and with relatively higher amount of pocket money at their disposal" (Reddy et al. 1993). Also, a study of marijuana-related hospital admissions in Malawi found that the marijuana-abusing group had significantly *more* schooling than the control group (Carr et al. 1994).

No relationship/inadequate information

In Zimbabwe, there were no particularly pronounced differences in the use of inhalants and cannabis by students of different SES subgroups, in contrast to that of tobacco and alcohol (Eide and Acuda 1996). A review of the use of non-prescription psychoactive substances in Nigeria reported that: "as a result of insufficient information, it could not be established who uses the substances more among different age, religious, social class, educational, and occupational categories" (Omoluabi 1995).

Literature on developed countries

Substance use associated with low SES

A United States study reported that people living in areas with high rates of poverty had relatively high rates of illicit drug use (marijuana or cocaine/crack) in the past year compared to a national sample (Ensminger et al. 1997). Also, frequent cocaine use in the United States was reported to be more prevalent in low-income urban areas than elsewhere (Brownsberger 1997). Another United States study found that social disadvantage was moderately associated with drug-related behaviour (Boardman et al. 2001). This work suggested that this was related to increased social stressors and higher levels of psychological distress among residents of disadvantaged neighbourhoods. Supporting evidence for such an association also comes from United States data on the early wave of injecting drug users with AIDS, who were characteristically poor or from ethnic minorities (Schrager et al. 1991).

Poverty and lower levels of education were associated with marijuana use in pregnancy in a United States review (Day et al. 1993). This review also reported work that found that cocaine use in pregnancy was associated with low educational attainment. In another United States review (Hans 1999), the studies reviewed suggested similar rates of substance use during pregnancy by women of different racial and social class categories (for all substances including alcohol). However, it did report that black women and poorer women were more likely to use illicit substances, particularly cocaine. Another United States review of

substance use in pregnancy also noted its association with poverty (Howell et al. 1999). A more recent study found that neighbourhood poverty was associated with higher rates of the use by pregnant women of cocaine, opiates, amphetamines and all illicit drugs (Finch et al. 2001).

A Swedish study suggested that low SES at the time of birth, relative to the general population, might be associated with dependence on amphetamines as an adult (Nyberg et al. 1992). However, this relationship did not exist for opiate dependence. In New Zealand, there is evidence that the Maori population has a significantly greater rate of cannabis use (Fergusson and Horwood 2000) and a relatively high rate of treatment for cannabis use (Adamson et al. 2000; Pomare et al. 1995). Similarly, Australian school survey data show that a relatively disadvantaged ethnic group (Aboriginal and Torres Strait Islander students) were more likely to have tried cannabis and other illicit substances (as well as cigarettes) than other students (Forero et al. 1999).

In terms of educational attainment and substance use, student survey data in the United States indicate that lower parental educational level was significantly associated with higher rates of substance use (Robinson et al. 1987), although "substance use" covered alcohol and tobacco as well as illicit drugs. A New Zealand cohort study found that adolescents who had less well educated mothers had significantly higher "deviant peer affiliations", i.e. associations with delinquent or substance-using peers (Fergusson and Horwood 1999). These substances included alcohol, tobacco and other drugs.

Substance use associated with high SES

Some United States research suggested higher rates of substance use among affluent, suburban or white youths than among poor, non-white, urban youths (Luthar and D'Avanzo 1999). Other work suggested minimal SES patterning for substance use: one United States study reported that illicit drug use was only 1.3 times more likely in disadvantaged neighbourhoods compared to the least disadvantaged ones (Saxe et al. 2001). However, visible drug sales were 6.3 times more likely to be reported in these disadvantaged neighbourhoods.

Discussion

The relationship

The available literature in general suggests a relationship between SES disadvantage and increased prevalence of use of illicit substances. This relationship appears to exist in both developing and developed countries. Nevertheless, some studies suggest that, for some substances in some settings, the opposite relationship exists. A further complication is that not all the studies on this relationship appropriately adjust for relevant confounding variables such as ethnic status and religious affiliation. In India,

for example, caste was described as being a factor in determining substance use (Sharma 1996).

Possible mechanisms

Many factors are likely to be significant in the relationship between poverty, low SES and drug use. For example, Finch et al. (2001) describe how neighbourhood poverty may increase substance use in a number of ways: ". . . the presence of greater numbers of substance users, access to substances, greater overall deviance, the social acceptability of substance use . . .". These authors also consider that contextual effects might be involved: ". . . the increased stressors of poor housing conditions, a lack of health care and substance use services, and a general lack of social services within the community" (Finch et al. 2001). Specific United States data highlight the importance of exposure, with youths living in the most disadvantaged neighbourhoods being six times more likely to have been offered cocaine (Crum et al. 1996).

In drug production areas, it is likely that poor farmers and their neighbours will have relatively easy access to these substances at low cost. Evidence from Malawi suggests that the low cost of marijuana is a factor influencing its usage (Carr et al. 1994). Also, poverty may sustain illicit drug production in settings where other cash crops are less profitable.

On the demand side, it is plausible that lower educational attainment is associated with poorer knowledge of the hazards of illicit drug use. Also, the hardship of poverty may be associated with a greater demand for the temporary psychological escape achievable through drug use. Self-perceptions of low status may also be relevant, given that low social status may affect dopamine receptors in the brain that increase the addictiveness of some drugs (Morgan et al. 2002).

4.4 Hypertension/high blood pressure

Literature on developing countries

A narrative systematic review covering literature published between 1966 and 1996 identified 13 studies in developing countries (Colhoun et al. 1998). Of these, a direct relationship between higher SES and higher blood pressure (BP) or hypertension was reported in five studies in India (three studies), Kenya and Nigeria. Another Indian study reported a non-significant direct relationship. The opposite pattern (of low SES associated with higher BP or hypertension) was reported in four studies in Brazil, India and Senegal and among South African Zulus. A further Indian study reported no relationship, while two studies in the Caribbean found conflicting results for males and females. Another review of cardiovascular disease risk factor data from 12 centres in seven developing countries also examined SES (three centres in Thailand, two each in Brazil, Chile and China and one each in

Colombia, Indonesia and the Philippines) (INCLEN 1994). For BP the associations with SES tended to be negative (i.e. similar to those from the developed world).

High BP associated with high SES

A number of studies conducted in Asia and Africa showed this pattern, e.g. villagers in India (Singh et al. 1997b), urban residents in India (Singh et al. 1998), urban workers in Nigeria (Ekpo et al. 1992) and urban civil servants in two Nigerian cities (Markovic et al. 1998; Omokhodion et al. 2000). However, while another Nigerian study found this pattern for men it found the reverse for women, i.e. higher BP in those with less educational attainment (Kadiri et al. 1999). There was a positive association between BP and income among a rural population in Papua New Guinea, but this held for women only and did not reach statistical significance. Another Chinese study found that men paid according to a new, highly demanding salary system exhibited significant increases in systolic BP between the first and second screenings (Siegrist et al. 1990).

High BP associated with low SES

Data from an urban population in China indicated that high BP and hypertension in men and women were associated with low SES (Yu et al. 2000). Over a seven-year period the SES gradient declined for men (systolic and diastolic BP). Over this period there was also an increase in systolic BP among better-educated women, while the reverse occurred in women with less education. A study in Saudi Arabia also found an association between high BP or hypertension and low SES (Wahid Saeed et al. 1996).

Literature on developed countries

The narrative systematic review by Colhoun et al. (1998) identified 57 relevant studies in developed countries. The reviewers concluded that lower SES was associated with higher mean BP in almost all studies in developed countries. However, the magnitude of the association varied and was generally quite small (with age-adjusted mean systolic BP differences of about 2–3 mmHg between the highest and lowest SES groups). Even so, this difference is still important in terms of its consequences for public health at the population level. This inverse gradient was both stronger and more consistently found in women than in men. An earlier review (the INTERSALT study) examined data on BP and educational attainment from 47 centres around the world (in 18 developed and 12 developing countries) (Stamler et al. 1992). It found an inverse association for men in 28 centres and for women in 38 centres. Age-adjusted systolic BP in men was significantly higher (1.3 mmHg higher for 10 fewer years of education), while for women the association was even more significant at 4.5 mmHg higher. Yet when the adjustment

included five lifestyle factors (24-hour sodium and potassium excretion, BMI, alcohol intake and smoking) these estimates were reduced by about 50%. Furthermore, after this adjustment the inverse association was no longer significant for men. The results obtained for diastolic BP were similar to those for systolic BP. Of ten studies on children in developed countries, only four reported significant differences in BP associated with SES (Colhoun et al. 1998). All four showed that systolic BP was higher in children of low SES, although in two the association disappeared after adjustment, e.g. for weight or fitness.

Discussion

The relationship

In developing countries there is a mixed picture, although slightly more studies have suggested a direct association between high SES and high BP. This pattern is consistent with that generally seen for the SES distribution of other cardiovascular disease risk factors in developing countries (obesity, hypertension, cholesterol level and to some extent physical inactivity). Nevertheless, in a number of developing countries the pattern is similar to the fairly consistent pattern in the developed world, with high BP associated with low SES. The pattern seen in developed countries is consistent with the evidence that mortality from hypertension-related diseases such as ischaemic heart disease, hypertensive heart disease, stroke and end-stage renal disease also shows an association with low SES (Colhoun et al. 1998).

Possible mechanisms

Colhoun et al. (1998) suggested that the SES gradient in BP seen in the developed world was unlikely to be explained by differential treatment rates. Indeed, these authors considered that a substantial part of it was accounted for by the SES gradient in BMI (see also section 4.7). Also, alcohol consumption across SES groups accounted for part of the association in men (although this issue was examined in only a few studies). Furthermore, these reviewers considered there to be little evidence that adverse psychosocial factors associated with low SES caused chronic elevations in BP. Nevertheless, they noted that stronger relationships might exist between BP and more specific stressful aspects of low SES such as unemployment or job insecurity.

The INTERSALT study reported that those with less education had on average higher sodium excretion, lower potassium excretion, greater body mass and higher alcohol intake (Stamler et al. 1992). All of these are factors that tend to be associated with increased BP.

For some developing countries, the review by Colhoun et al. (1998) suggested that the direct association between higher SES and higher BP may reflect a higher prevalence of obesity and higher salt and alcohol intakes among those of higher SES. Other dietary factors (e.g. more

animal products) and less physical activity in these groups might also play a role. Furthermore, a Brazilian review concluded that the higher people's "cultural consonance" the lower their BP (with cultural consonance referring to how individuals are able to approximate, in their own behaviour, shared cultural models of life) (Dressler and Santos 2000).

In those developing countries where low SES is associated with high BP, poor maternal and perinatal nutrition may be important factors. Indeed, there is growing evidence from animal studies (Edwards et al. 1999; Ozanne 2001) and human studies (Roseboom et al. 2001) for the Barker hypothesis on the fetal origins of adult disease.

Regardless of the mechanisms involved, it is important to note that there is evidence that the intensive medical management of hypertension can abolish SES differences in hypertension-associated mortality (Anonymous 1977).

4.5 Serum cholesterol and lipid levels

Literature on developing countries

Adverse lipid profile associated with high SES

Data from 12 centres in seven countries (the INCLEN study) found such a relationship for total cholesterol (TC) in six centres. These included centres in Brazil (for occupational class and income); Chile; China (for occupational class); the Philippines (for education, occupational class and income); and Thailand (one centre for income and one for education) (INCLEN 1994).

China. A study of an urban population reported that high-SES people had a more unfavourable serum lipid profile (for TC, HDL-c, LDL-c and triglycerides [TG]) than those in lower-SES groups (Yu et al. 2002). This significant association was especially apparent in men, and education seemed to be the most important predictor of serum lipids compared to the other SES indicators of occupation and income. A cohort study of urban male workers found that blood lipid profiles were less favourable in the better-educated groups (Siegrist et al. 1990). LDL-c and LDL-c/HDL-c ratio were significantly higher and HDL-c was significantly lower. Similar results were obtained using occupational class.

India. A study of an urban population found a significant SES gradient for both men and women for hypercholesterolaemia, TC, LDL-c, HDL-c and TG (Singh et al. 1998). Similarly, surveys of women in five Indian cities found positive gradients by social class for TC and HDL-c (Singh et al. 1998). A study of an elderly urban population also found an association between hypercholesterolaemia and income (Singh et al. 1995). In rural India, one study found a significant SES gradient for hypercholesterolaemia, for TC and for TG in women (Singh et al. 1997a), but the

trends for higher LDL-c and HDL-c in both men and women were not significant.

Other countries. In a white rural South African population, the prevalence of hypercholesterolaemia increased with income for men but there was no association with education or between HDL-c and SES (Rossouw et al. 1990). Low HDL-c was more prevalent in women (by income and educational attainment) but, in contrast to the men, there was less hypercholesterolaemia with increasing level of income.

A study in Mauritius found that professional and skilled men had significantly higher mean TG, higher LDL-c and lower mean HDL-c than unskilled workers (Pereira et al. 1998). Unskilled and partly skilled women had lower TG compared with "homemakers". Also, in contrast to the men, professional and skilled women had higher HDL-c than unskilled women (with no significant association for LDL-c).

A study in Turkey found an association between income and TC in urban men and women and rural men, but not rural women (Onat et al. 1992). Another Turkish study found higher TC and HDL-c levels in high-SES than in low-SES children (Mahley et al. 2001). The high-SES group also consumed more saturated fat of animal origin and less carbohydrate.

Lipid profile and development

Cholesterol levels in the population appear to increase with the modernization and economic development of a country. For example, a study of adult men in 13 countries in 1980 reported that mean TC levels were low in those from Africa, intermediate in those from Hungary, Pakistan, the Philippines, Poland, Suriname and the Mediterranean countries, and high in those from Finland and the Netherlands (Knuiman et al. 1982). HDL-c tended to be lower in men from Africa, Asia and Suriname than in those from Europe (with the highest values for both TC and HDL-c being found in Finnish men). The HDL-c:TC ratio was on average slightly higher in those from Africa than it was in those from Europe and from Asia and Suriname.

Another multi-country study found low TC levels in Amazonian Indians and Africans (traditional Bantu population) compared to Italian and Polish populations (Pavan et al. 1999). Other studies found relatively low TC levels in other traditional societies with diets based on complex carbohydrates and vegetables (Mancilha-Carvalho and Crews 1990). Even in developing countries, such a Zimbabwe, there appears to be an increase in TC levels in urban relative to rural populations (Allain et al. 1997).

Adverse lipid profile associated with low SES

Some of the results for specific aspects of the lipid profile detailed in studies described above fit this pattern (e.g. low-SES white rural South

African women and hypercholesterolaemia, and unskilled women in Mauritius and HDL-c). Furthermore, data from the INCLEN study revealed such a relationship in one centre (Shanghai in China for low occupational class and raised TC) (INCLEN 1994). In India it was reported that TC and LDL-c were high in urban slum dwellers (Misra et al. 2001), but this study did not involve a comparison group.

No association (lipid profile and SES)

Data from the INCLEN study found no significant relationship for TC in five centres in Brazil, Chile, Colombia, Indonesia and Thailand. In a rural population in western India there was no association of SES with the prevalence of hypercholesterolaemia (for TC, LDL-c, HDL-c and TG) in men and women by educational status. Similarly, SES was not related to HDL-c levels in adults, in contrast to the results for Turkish children detailed above (INCLEN 1994).

Literature on developed countries

A review of cardiovascular disease risk factors and SES by Pickering (1999) reported that the SES gradient for TC was generally minimal or non-existent. He reviewed studies in England (two studies), Norway (one study) and the United States (four studies). However, further work has subsequently been published, some of which is described below.

Adverse lipid profile associated with high SES

In a multi-country study (mainly in developed countries) the lipid profile was worse with increasing years of schooling for men in China, Poland and the Russian Federation (for raised TC, LDL-c and TG and lower HDL-c) (Perova et al. 2001). However, the findings were less consistent for women and for Israeli and American blacks of both sexes. Also, the opposite pattern was seen among white men from the United States.

A study in Scotland found that TC levels were higher in the non-manual social class groups, for both men and women (Davey Smith et al. 1998). A study of working Japanese men showed that those in higher employment grades had lower levels of HDL-c (Martikainen et al. 2001).

A study of an economically depressed agricultural area in the United States revealed that those with low educational attainment and income below the poverty level had lower mean TC levels than the rest of the population in the area (Gold and Franks 1990), despite the fact that the low-SES groups in this population had a greater intake of dietary cholesterol.

Adverse lipid profile associated with low SES

Perova et al. (2001) showed that white men in the United States with low educational attainment had an adverse lipid profile. Similarly, a national survey in the United States found that for whites, HDL-c levels

were highest for those in the highest category of earnings (although blacks generally had lower levels of HDL-c with increased earnings) (Linn et al. 1989). In a multivariate model, predictors of higher HDL-c included a higher frequency of alcohol intake and reported high physical activity, with smoking and high BMI being strongly negatively related to HDL-c levels. The authors of this national study suggested that the findings "support previous findings in selected populations in the United States".

In Finland, cholesterol levels were found to be higher in blue-collar than white-collar workers, for both men and women (Vartiainen et al. 1998). This study found that the decline in cholesterol levels between 1972 and 1987 was similar for men in both groups, but greater among white-collar women than blue-collar women. Similarly, in Norway, poverty during childhood was positively associated with age-adjusted levels of TC in adults (Arnesen and Forsdahl 1985). In Sweden, women of high SES were found to have a better lipid profile than low-SES women (lower levels of TC, LDL-c and TG and raised HDL-c), but there was no relation between occupational status and lipid levels for men.

In the Whitehall II longitudinal study of men and women in the United Kingdom, it was found that high occupational status was associated with lower cardiovascular disease risk factors, including HDL-c levels (Brunner et al. 1999). This study also reported that social position in childhood was associated with HDL-c level in women. Another study of working men in England found that higher occupational class was associated with a favourable lipid profile (for HDL-c and TC) (Martikainen et al. 2001).

In Greece, TC levels were found to be inversely associated with educational level in both sexes, and especially in women (Benetou et al. 2000). The authors report that this SES pattern contrasted with the previous pattern that existed two decades before in Greece. HDL-c levels were inversely associated with educational level in men, but the association was not clear for women.

No association (lipid profile and SES)

In a German study, no poverty-related differences were found for prevalence of hypercholesterolaemia, despite a much higher prevalence of obesity in persons with an income below the poverty line (Helmert et al. 1997). In a national survey in New Zealand there was no overall SES gradient in TC levels (Howden-Chapman and Tobias 2000). In European men and Maori women, however, relatively higher TC levels were seen with high SES while, in contrast, higher TC levels were seen in low-SES European women (based on educational attainment).

TC levels for children in a disadvantaged inner-city population in the United States were not significantly different from those of other children from several large studies of North American populations (Wadowski et al. 1994).

Discussion

The relationship

In the developing world, the predominant pattern is for the population as a whole to develop an adverse lipid profile along with industrialization and urbanization. Furthermore, within developing countries, high-SES groups appear to have the most hazardous lipid profiles. This pattern is consistent with that generally seen for the SES distribution of other cardiovascular disease risk factors in developing countries (obesity, hypertension and to some extent physical inactivity).

The picture is more mixed in the developed world, but the predominant pattern is for low-SES groups to have the most adverse lipid profiles (particularly in the most developed countries in this group). Such a pattern is also consistent with that generally seen for the SES distribution of other cardiovascular disease risk factors in developed countries (smoking, obesity, hypertension and physical inactivity). It is also consistent with the relatively higher rates of cardiovascular disease among low-SES groups in many developed countries.

Possible mechanisms

In the developing world, it is likely that dietary change is a key factor in the development of adverse lipid profiles associated with industrialization and urbanization. Higher total calories and increased intakes of animal products are probably important. Nevertheless, other factors may be relevant: a Brazilian study found significant interactions between elevated cholesterol levels and smoking, obesity and physical inactivity (Martins et al. 1995).

In developed countries, the generally poorer diet of low-SES groups is also likely to be a factor in generating adverse lipid profiles. For example, increased dietary intake of cholesterol was described among various United States populations, including poor elderly people (Prothro and Rosenbloom 1999), poor children (Casey et al. 2001) and children whose parents smoked (a crude marker for low SES) (Johnson et al. 1996). Similarly, the higher BMI of low-SES groups is likely to play a role: a 13-country survey by Knuiman et al. (1982) found a direct relationship of BMI with TC and an inverse relationship with HDL-c and HDL-c:TC ratio.

Physical inactivity may also be relevant (given that it contributes to obesity) and there is some evidence that it is associated with a poorer lipid profile, although studies have reported conflicting results (Sowers et al. 1995). Regular consumption of modest quantities of alcohol may also play a role in improving the lipid profile of high-SES groups. For example, a recent French study found that wine drinkers had higher HDL-c levels than non-wine drinkers (Ruidavets et al. 2002). However, the authors reported that such differences became non-significant after adjustment for SES parameters.

Furthermore, there may be psychosocial pathways for determining lipid profiles in different populations. For example, a case–control study in England suggested that "lifestyle incongruity" is associated with higher cholesterol levels (Dressler et al. 1992). This variable is a type of status incongruence, involving the degree to which lifestyle (material consumption and status-enhancing behaviour) exceeds occupational status.

Low birth weight may play a role in influencing lipid metabolism in the adult. This is one component of the still controversial "Barker hypothesis" discussed in section 4.4.

In some developed countries there may also be SES patterning of cholesterol screening, awareness of the risks of elevated cholesterol levels, willingness to change one's diet to reduce cholesterol levels and access to cholesterol-lowering drugs.

4.6 Overweight/obesity

Literature on developing countries

Obesity associated with high SES

A review of 144 published studies on the relationship between SES and obesity was published by Sobal and Stunkard (1989). It revealed a strong direct relationship between SES and obesity in women in 10 studies in developing and non-Western societies (with another study showing no relationship and one showing the opposite relationship). For men, 12 studies found a relationship and two found no relationship. In children and adolescents, this pattern was seen in 14 studies in girls (with two other studies showing no relationship) and in 13 studies in boys (with two others showing no relationship). A subsequent review by Colhoun et al. (1998) reported that BMI and SES were measured in seven out of 13 studies on BP and SES in developing countries. In four of these studies—in rural India (two studies), rural Kenya and urban Nigeria—BMI increased with higher SES. In Senegal, BMI did not vary with SES and in two studies the relationship of SES to BMI was not reported. These reviewers concluded that, overall, the SES differences in BMI "probably account for a substantial part, though not all, of the SES–BP association".

The INCLEN study collected cardiovascular disease risk factor data on men from 12 centres in seven developing countries (INCLEN 1994). It found significant trends for increasing BMI with higher SES in five out of the 12 centres. These were in urban China (by occupation and income), urban Indonesia (by education), urban Philippines (by education, occupation and income), rural Thailand (by education and income) and urban Thailand (by occupation and income). A recent analysis of obesity in women based on nationally representative surveys (39 surveys from 38 developing countries) (Martorell et al. 2000) reported that in developing countries, such as those in sub-Saharan Africa, obesity levels were higher among better educated women and urban women. Yet in

more developed countries, such as those in Latin America and central and eastern Europe, obesity levels were more equally distributed in the general population. At a national level, levels of obesity in countries increased sharply up to a GNP of US$ 1500 per capita (1992 values) and then changed little thereafter.

Another multi-country study of developing and developed countries used data for children aged 6–18 years from nationwide surveys in China, the Russian Federation and the United States (NHANES III, 1988–1994) (Wang 2001). It found that higher-SES subjects were more likely to be obese in China and the Russian Federation, while in the United States low-SES groups were at higher risk. A review of preschool children by Martorell et al. (2000) examined 71 national nutrition surveys since 1986 from 50 developing countries. It found that in 22 countries (mainly those from Latin America, the Caribbean, the Middle East and North Africa) overweight was significantly more common in children of mothers with higher education. Overweight was also significantly more common in urban areas in 24 countries (and only more common in rural areas in two countries). At national level, the prevalence of overweight among preschool children tended to increase with increasing GNP ($r = 0.28$, $P = 0.05$).

In China, a nationwide longitudinal survey conducted in 1989 and 1991 (Popkin et al. 1995a) found evidence that increased household income was significantly associated with increased BMI. Urban residence and higher income were associated with lower energy intake, higher fat intake and lower levels of physical activity compared to rural residence and other income categories. Diet was considered to be a particularly important determinant of body weight in this population (Paeratakul et al. 1998).

Obesity associated with low SES

The INCLEN study (INCLEN 1994) found significant trends for increasing BMI with lower SES in two out of 12 centres. These were in urban Colombia (by education and occupation) and rural Thailand (by occupation). The review by Sobal and Stunkard (1989) found no studies reporting an association between obesity and low SES. Similarly, the review of overweight among preschool children from 50 developing countries found none in which overweight was significantly more common in the children of poorly educated women (Martorell et al. 2000).

A review of obesity in developing countries (Popkin et al. 1995b) cited evidence for increased obesity among low-SES groups from urban Brazil. Another recent review by Pena and Bacallao (2000) reported such a finding for urban Chile, for women in Uruguay and for urban Peru. A book on obesity in Latin America described the higher prevalence of obesity in middle-income compared to high-income women (Monteiro 2000).

A significant gradient was described for BMI and obesity prevalence by education level for urban women in China (i.e. greatest among those with the least education) (Yu et al. 2000). Over the period 1989–1996 this gradient declined for women, along with a decrease in BMI across all educational strata. The results for men suggest no significant gradient in BMI and obesity and no significant time trends.

No association (obesity and SES)

The review by Colhoun et al. (1998) reported one country (Senegal) where there was no association found between BMI and SES. Also, the INCLEN study (INCLEN 1994) found no significant associations for BMI and SES in five out of 12 of the centres. These were in Brazil (two sites), Chile (two sites) and China. The review of overweight among preschool children from developing countries by Martorell et al. (2000) found no association with maternal education in 28 out of 50 countries. The review by Sobal and Stunkard (1989) found no association in one out of 11 studies of women. The equivalent figures for men, boys and girls, respectively, were 2/14, 2/15 and 2/16. One multi-country study examined the prevalence of the coexisting underweight and overweight individuals in households using survey data from Brazil, China and the Russian Federation (Doak et al. 2000). It found no clear pattern in the prevalence of underweight/overweight in households by income, but this coexistence pattern was highest in the urban environment in all three countries.

Literature on developed countries

Obesity associated with low SES

The review by Colhoun et al. (1998) reported that adjustment for BMI in 26 studies (along with other variables) usually attenuated observed SES gradients in BP (increasing BP with lower SES). Furthermore, in five studies in Germany, the Netherlands, Sweden and the United States (two studies), in which adjustment was made for age and BMI only, this SES gradient was abolished completely. Some of the studies described showed stronger gradients for women compared to men for increased BMI and low SES (Shewry et al. 1992) and in some settings the association applied only to women. Similarly, the INTERSALT study found that increased BMI was significantly associated with lower SES, particularly for women (Stamler et al. 1992).

Surveys associated with WHO's MONICA (MONItoring of CArdiovascular diseases) Project collected data on BMI and years of schooling for 26 populations in the period 1979–1996 (25 developed countries plus China) (Molarius et al. 2000). An analysis of the combined data found that, for women, almost all populations (22 out of 26) showed a statistically significant association between lower educational level and increased BMI (the difference between the highest and the lowest edu-

cational tertiles ranged from –3.3 to 0.4 kg/m^2). For men, there was the same association in six out of 26 of the populations (with the equivalent range being from –1.5 to 2.2 kg/m^2). In about two thirds of the populations, the differences in BMI between the educational levels increased over the 10-year period.

The earlier review of 144 published studies on SES and obesity by Sobal and Stunkard (1989) also revealed a strong inverse relationship among women in developed societies. There were 28 United States studies showing this relationship for women and 18 from other developed countries (with only seven studies showing no relationship). For men there were 12 United States studies and 22 studies from other developed countries that found this relationship (while 20 found the opposite relationship and 11 found no relationship). In children and adolescents this pattern was seen in 13 studies for girls (with eight showing the opposite relationship and 11 showing no relationship) and similarly in 11 studies in boys (with nine showing the opposite relationship and 14 showing no relationship).

More recently, a systematic review by Parsons et al. (1999) found that in developed countries there was a strong consistent relationship observed between low SES in early life and overweight in adulthood. These reviewers found that women who changed social class (social mobility) adopted the prevalence of obesity of the class they joined, while this association was not present in men. More recent survey data from various developed countries is consistent with the relationship described above of increased BMI being associated with low SES, for example New Zealand (e.g. for education) (Dryson et al. 1992; Howden-Chapman and Tobias 2000; Statistics New Zealand/Ministry of Health 1993); the United Kingdom (for occupational class among men and women) (Acheson 1998); and the United States (for education among men and women) (Pamuk et al. 1998).

Obesity associated with high SES

The review by Sobal and Stunkard (1989) found this relationship for men in three United States studies and in eight studies from other developed countries. It was found in only one study of women (among migrants in Belgium). In the 26 MONICA Project populations this association was reported in two populations for men (in Poland and the Russian Federation) and none in women (Molarius et al. 2000). Among children, this relationship was reported for eight studies in girls and nine in boys.

No association (obesity and SES)

The review by Sobal and Stunkard (1989) found no association between SES and obesity in men in 11 studies, in women in seven studies, in boys in 14 studies and in girls in 11 studies. Of the 26 MONICA Project surveys, 18 found no significant association between BMI and SES in

men and four found no such association for women (Molarius et al. 2000). Similarly, the lack of an association was described for many studies for men and children in the earlier review by Sobal and Stunkard (1989).

The systematic review by Parsons et al. (1999) found no clear relationship between SES in early life and overweight as a child, even though a relationship was reported for low SES in early life and overweight as an adult. These reviewers reported that very few of the relevant studies investigating SES considered confounding by parental obesity. Such findings are consistent with an examination of United States survey data on 12 states between 1980 and 1989, which found in general that children from low-income families did not have a greater prevalence of overweight than children from higher-income families (Yip et al. 1993). (In two states, however, low-income school-aged children and adolescents did show significant increases in body weight in relation to height.)

Discussion

The relationship

A very large body of evidence from studies in developed countries indicates that the predominant gradient is that of lower SES being associated with increased body weight. This relationship appears very strong and consistent for women, but only six of the 26 MONICA Project surveys found such a relationship for men. Also, a systematic review found no clear relationship between SES in early life and childhood obesity, although a relationship was reported for low SES in early life and adult obesity.

The predominant pattern in developing countries is for higher-SES men and women to have higher body weight. Nevertheless, some studies have found no relationship and in some regions (e.g. Latin America) the opposite pattern appears to be emerging. Among preschool children there is also evidence that being overweight was significantly more common in those with mothers with higher education (in 22 out of 50 countries).

Possible mechanisms

In developing countries the association between BMI and higher SES is probably related to a "nutrition transition" (Popkin 2001) to diets higher in fat. A trend towards poorer diets (e.g. increased dietary fat) with increasing income was described for China and the Philippines in a review by Popkin et al. (1995b). Societal attitudes to obesity and thinness, as described by Sobal and Stunkard (1989) may also be relevant in determining dietary restraint and participation in physical activity. For example, mild obesity appears to be equated with physical attractiveness in some countries. Fetal and infant nutrition may also be relevant in determining subsequent obesity, although a recent review of this evidence

in a developing country context suggests the relationship is not as clear for obesity as it is for hypertension (Dryson and Martorell 2000). Reduced physical activity was considered to be a relevant factor for obesity in high-SES groups (see also section 4.7), but reduced physical activity might also be an important factor for obesity in low-SES groups (Torun 2000).

In developed countries it is likely that nutrition also plays a key role in the association between low SES and increased BMI. A review by Potter (1997) reported evidence of poorer nutrition with low SES in a number of developed countries: Australia (three studies), Canada, Denmark, Finland and the United States (four studies). In all these studies the dietary pattern was suggestive of that associated with the greatest risk of excessive calorie intake (i.e. low in vegetables, fruit or whole-grain cereals and/or high in fat, meat, butter or whole milk). In particular, the relationship between dietary fat intake and obesity was fairly well established (Bray and Popkin 1998). Similarly a meta-analysis of 11 studies from seven European countries found lower vegetable and fruit intake in low-SES groups (using education and occupation) for both men and women (Irala-Estevez et al. 2000). Data from other studies provide further evidence for this pattern, for example for New Zealand (Howden-Chapman and Tobias 2000) and the United Kingdom (Acheson 1998). The SES gradient for breastfeeding, e.g. in the United Kingdom (Acheson 1998), might also be relevant in determining subsequent BMI in childhood.

In some developed countries, physical inactivity might well play a role in the SES–BMI gradient (see section 4.7). Other explanations include "a complex mix of poverty of information and skills, a distorted view of "status foods" and an approach to dietary priorities typical of the 1940s, which was then much more focused on protein and calories" (Potter 1997). Furthermore, cultural attitudes to thinness and dietary restraint may be important (Sobal and Stunkard 1989).

When considering potential explanations for the relationship between SES of origin and adult obesity, Power and Parsons (2000) suggested that SES of origin may actually be confounded by parental body size. Alternative explanations include "(1) nutrition in infancy and childhood, either over- or undernutrition, followed subsequently by over nutrition; (2) psychological factors, possibly involving emotional deprivation in childhood; (3) cultural or social norms regarding dietary restraint and attitudes to fatness that may be acquired during childhood".

4.7 Physical inactivity

Literature on developing countries

Physical inactivity associated with low SES

In an urban Chinese population, low level of education was found to be associated with leisure-time physical inactivity in both men and women

(Yu et al. 2000). However, this gradient declined over the seven-year study period, particularly for men. Other Chinese survey data indicate that urban residents are far more sedentary than rural residents (Paeratakul et al. 1998). A study of a metropolitan area in Brazil found that high-SES women were less likely to be physically inactive (Martins et al. 1995). This was also the pattern for the skilled working class relative to the unskilled working class. A study in a Caribbean island reported that low-SES men and women (based on educational attainment) were reported to exercise less often (Grol et al. 1997b).

Physical inactivity associated with high SES

In India, physical inactivity was reported to be more prevalent among those of higher SES in women in five cities (Singh et al. 1999), in men and women in villages (for SES based on occupation) (Singh et al. 1997b) and in men and women in a city (Singh et al. 1998). For SES based on educational attainment, the same pattern was seen in another study of men and women in Indian villages (Gupta et al. 1994). This pattern was also reported for men in a Brazilian city for overall physical activity, although there was no association for leisure-time activity (Duncan et al. 1993). In contrast, the higher-SES women in this Brazilian city had higher levels of physical activity. A study in Puerto Rico reported that urban men with more education had lower levels of physical activity (Sorlie and Garcia-Palmieri 1990). Also, in Mauritius, unskilled workers were reported to be significantly more physically active than members of other occupational groups (attributable largely to occupational rather than to leisure-time physical activity) (Pereira et al. 1998).

Literature on developed countries

Physical inactivity associated with low SES

In the United States, national survey data indicated increased leisure-time physical inactivity with lower SES (including people who were less educated, who lived below the poverty line and who lived in low-income households) (Crespo et al. 1999). Previous national survey data also showed this pattern (USDHHS 2001). However, these indicators still do not seem to explain the higher prevalence of leisure-time physical inactivity among specific populations (African Americans and Mexican Americans) (Crespo et al. 2000).

Consistent with these results are studies of specific United States population groups, for example for income and education in a disadvantaged urban community (Diez-Roux 1998); educational level and women among members of a health maintenance organization in California (Sternfeld et al. 1999); and educational level and income among Minneapolis-St Paul residents). The latter study reported time trends for leisure-time physical activity, indicating greater gains among men with

low SES and among less affluent women. There is also United States evidence that living in an poor area is independently associated with physical inactivity and in a decline in physical activity levels over time (Yen and Kaplan 1998).

Similarly, the Whitehall II longitudinal study in the United Kingdom found that low-status occupation was associated with leisure-time physical inactivity for both men and women (Brunner et al. 1999). Adjustment for earlier SES position (using father's social class and own education level simultaneously) did not weaken the effects of SES position as an adult. Other United Kingdom work showed non-participation in sport to be associated with low SES, and no significant changes in this gradient occurred between 1984 and 1993 (Bartley et al. 2000).

Survey data from Germany have also suggested that lack of regular exercise (for both sexes) was one of the most striking poverty-related differences in terms of cardiovascular disease risk factors (Helmert et al. 1997). Other German work has reported that high-SES men and women were, respectively, four and three times more likely to have an active leisure time than those with low SES (Mensink et al. 1997). Similarly, survey data on elderly people of low SES in Switzerland suggest that they get less physical exercise (Abelin and Schlettwein-Gsell 1986). In the Netherlands, a cohort study reported that lower levels of educational attainment were associated with physical inactivity (Schrijvers et al. 1999), as did a cross-sectional study (Droomers et al. 1998).

A Swedish cohort study (Lindstrom et al. 2001) reported increased levels of physical inactivity among skilled and unskilled male manual workers and unskilled female manual workers compared to high-level non-manual employees. However, it was found that adjusting for social participation almost completely erased the SES differences (social participation included involvement in political parties and organizations and was a strong predictor of increased physical activity). In Australia, it was reported that multicentre cross-sectional community survey data indicate a "clear socioeconomic gradient between leisure-time physical activity and education attainment" (Bennett 1995). Nevertheless, time-trend data indicate that walking for recreation or exercise has become more popular among older men of low education. Another Australian study found that a higher proportion of students participating in organized sport and health and fitness activities came from the higher-SES group (based on parental occupation) (Blanksby et al. 1996).

Physical inactivity associated with high SES

Survey data from a Spanish city between 1983 and 1994 found that physical inactivity in men was always more prevalent in social classes I and II (for total daily activity including activity at work) (Borrell et al. 2000). This difference increased over time as more people of advantaged classes became physically inactive. Among men in the United Kingdom, those

in the manual classes had a higher level of physical activity than those in the non-manual classes, owing partly to work-related physical activity (Acheson 1998). However, the general pattern for the United Kingdom is for physical inactivity to be associated with low SES (as discussed above).

No relationship

In New Zealand, the most recent national survey found no significant relationship between physical inactivity and social class or educational attainment (Howden-Chapman and Tobias 2000). This was in contrast to earlier New Zealand national survey data, which showed a gradient for lower levels of physical inactivity with higher educational attainment (for more than four hours of exercise in the previous seven days) (Statistics New Zealand/Ministry of Health 1993).

Discussion

The relationship

In general, physical inactivity is associated with low SES in developed countries, with a fairly marked gradient. Nevertheless, there are some exceptions, such as the reverse pattern among men in Spain. The relationship is more mixed for developing countries, with opposing gradients seen for China and India (with the latter showing the pattern of the developed world).

For the developed countries, this relationship appears to have impacts on disease outcomes. A United States review suggested that SES gradients in cardiovascular disease may be partly attributable to SES gradients in physical activity (Pickering 1999). A Dutch cohort study also reported that physical inactivity partly explained the relationship between poor educational attainment and mortality rates (Schrijvers et al. 1999). In general terms, the pattern of low SES being associated with physical inactivity is consistent with the evidence from developed countries that low SES in men is associated with increased risk of developing cardiovascular disease (Gonzalez et al. 1998).

Possible mechanisms

Physical inactivity and low SES. In the United Kingdom, barriers to exercise among low-income people include illness or disability, lack of money and lack of transport (Chinn et al. 1999). Fear of crime may also have an effect on walking and the use of parks in some neighbourhoods, along with access to such parks, recreational facilities and gymnasiums. Swedish data suggest that high-SES groups have increased social participation, and it may be that leisure-time physical activity is partly mediated by a higher extent of encouragement/peer pressure to participate in physical activity (Lindstrom et al. 2001). United States data also indicate that social support and confidence in one's ability to continue to

exercise when faced with other pressures and demands is important in determining levels of physical activity in women (Sternfeld et al. 1999).

Physical inactivity and high SES. In at least some parts of the developing world, the move to sedentary occupations among those with high SES appears to be an important factor in the SES patterning of physical inactivity (Singh et al. 1998). A decline in walking to work among members of this group may also be important (Duncan et al. 1993). Also, in some low-income neighbourhoods there may be "contagion effects" where a culture of being out on the streets and walking around prevails (Ross 2000), possibly combined with higher-density housing. In the United Kingdom, barriers to participation in physical activity among high-income people include lack of leisure-time and lack of motivation (Chinn et al. 1999).

4.8 Blood-lead levels

Literature on developing countries

A review of lead poisoning in children in China found 17 relevant publications. Those children residing in industrial areas and areas with heavy traffic had average blood-lead levels of 21.8–67.9 μg/dl (Shen et al. 1996). The percentages of blood-lead levels above 10 μg/dl, which defines lead poisoning in children, ranged from 64.9 to 99.5. A more recent study in a Chinese city (Wuxi City) in children aged 1–5 years also found that blood-lead levels were significantly higher for those living in industrial areas (Gao et al. 2001).

In a survey in India of children aged from 6 months to 6 years, risk factors for elevated blood-lead levels were living in houses painted with lead-based paint, recent exposure to lead-based paint and the use of the eye cosmetic "ma". However, they also found that high caste was a risk factor. No association was found with traffic, parental occupational exposure or nutritional status.

In a study in a Mexican city, low family income was one of the most important factors related to raised blood-lead levels in children, along with family use of lead-glazed pottery and the use of animal fat in cooking (Azcona-Cruz et al. 2000). Another study examined blood-lead levels in pregnant women in Mexico City attending public hospital prenatal clinics (representing primarily women of low SES) and private hospitals (primarily women of high SES) (Farias et al. 1996). Overall, blood-lead levels were significantly higher in the public hospital group. Consumption of milk products significantly reduced lead levels in the higher-SES group, and taking calcium supplements lowered blood-lead levels in those women whose diets were deficient in calcium. The authors concluded that avoiding the use of lead-glazed ceramics, consuming diets rich in calcium and, if needed, taking calcium supplements would be

expected to result in a substantial lowering of blood-lead levels, especially in pregnant women of low SES.

A study of 4–12-year-old schoolchildren in Bangladesh found that elevated blood-lead levels were significantly associated with low parental education (odds ratio of 2.7, 95% CI 2.0–3.8) and living close to major roads (odds ratio of 2.3, 95% CI 1.2–4.3) along with soil-eating and increasing age (Kaiser et al. 2001). The authors considered that living near major roads may be related to low SES in this population.

A study in South Africa (KwaZulu-Natal) examined children aged 3–5 and 8–10 years (Nriagu et al. 1997). Household factors that were significantly associated with blood-lead levels included distance from tarred roads, overcrowding, hygienic habits in the household and the burning of solid waste for heating or cooking, all of which are related to poverty and low SES. The lack of significant associations with risk behaviour in this study was attributed by the authors to "the over-riding influence of high levels of contaminated dusts both indoor and outdoor". Another study in South Africa found that low parental education and income were associated with raised blood-lead levels, along with dusty homes, homes in a poor state of repair and overcrowding (von Schirnding et al. 1991).

Literature on developed countries

A recent review (Tong et al. 2000b) reported that in developed countries chronic exposure to low levels of lead is still a public health issue, "especially among some minorities and socioeconomically disadvantaged groups". Similarly, a review in the United States reported that poor and minority children have higher rates of elevated blood-lead levels (Powell and Stewart 2001).

National survey data on United States children aged 1–5 years found that independent risk factors for elevated blood-lead levels included: "being of minority race/ethnicity, living in an older home, residing in the northeast or midwest regions of the United States, being on Medicaid, having a head of household with <12 years of education, and having a history of anemia" (Kaufmann et al. 2000). Some of these factors are associated with SES disadvantage (living in older housing and low level of education). An earlier national survey also reported associations between blood-lead levels in children and low income and low parental educational attainment (Pirkle et al. 1994).

Another study in the United States found that lead concentrations in the tibia were significantly higher in men who did not graduate from high school than in men with ≥4 years of college education (Elreedy et al. 1999). Living in an area with a poor level of education was associated with a significantly higher tibia lead level among those not graduating from high school, but not among college graduates. The authors concluded that the influence of individual SES on cumulative lead

exposure is modified by geographical area. This study also identified a significant relationship between low individual annual income and increased bone lead levels.

In Australia, a cohort study found that residential area and father's place of employment were the two variables most strongly predictive of a child's blood-lead concentration at the end of primary school (Baghurst et al. 1999). A poorer-quality home environment was also found to be an independent contributor to blood-lead concentrations. Other work on this cohort, however, did not find statistically significant interactions between lifetime average blood-lead concentration and parents' occupational status (used as a surrogate of SES) (Tong et al. 2000a).

In New Zealand, a study of a cohort at age 21 years found significant associations between higher blood-lead levels and high-risk occupational activities, living close to a main road, smoking and recreational exposure (Fawcett et al. 1996). Some of these risk factors have associations with SES (occupation and smoking). However, a study of this same cohort at age 11 years found no significant correlation between blood-lead levels and SES (Silva et al. 1986).

Discussion

The relationship

The overall pattern in both developed and developing countries is that poverty and low SES are associated with elevated blood-lead levels in children and adults. Nevertheless, there is still a minority of studies in which such a relationship is not apparent (for example, high caste Indian children, who may have greater exposure to cosmetics and crayons containing lead).

Possible mechanisms

The literature suggests a number of possible mechanisms for a relationship between poverty and elevated blood-lead levels.

- *Air pollution from vehicle exhausts.* People living near major roads are likely to have increased exposure to airborne lead in countries where lead is still added to petrol. In the United States, air pollution exposures and associated health risks appear to affect disproportionately populations that are poor and non-Caucasian (see section 3.5 for further details).
- *Older and dilapidated housing.* Older houses have an increased probability of having been painted with lead-based paint, and these tend to be concentrated in older urban centres where many disadvantaged groups live. In these settings, and where houses are not properly maintained, children are more likely to ingest flakes of lead-based paint.

- *Housing in the developing world*. It has been suggested that the type of construction of some low-income housing and its location close to roads may make it difficult to keep these houses free from lead in dust and soil (Kaiser et al. 2001).
- *Emissions from industry*. Poor people are more likely to be exposed to such emissions, since they more often live in or close to heavily industrialized areas. One United States review noted that: "minority and poor families disproportionately live in communities with landfills, hazardous waste facilities, incinerators, industrial plants, and old housing with poor indoor air quality and lead-based paint" (Powell and Stewart 2001).
- *Occupational hazards*. Certain occupations are associated with increased risks of high blood-lead levels (e.g. lead smelting and battery manufacture). Many of these blue-collar factory workers are in relatively disadvantaged groups. Also, these workers sometimes bring lead dust into the home environment and thereby expose family members. In developing countries poor people can be exposed to lead in cottage industries.
- *Smoking*. Smoking is a risk factor for high blood-lead levels and is generally more common in low-income groups.
- *Home remedies*. Some home remedies used in developing countries are known to contain high levels of lead (Azcona-Cruz et al. 2000). Poor people might be more likely to use such remedies than those who can afford to seek treatment from a health professional.
- *Poor nutrition*. The dangers of lead exposure are increased by several dietary conditions (Mahaffey 1990) that are more frequently present in economically disadvantaged children.
- *Knowledge of risks and preventive measures*. Poorly educated parents may be less likely to know of the risks of lead exposure and to make use of preventive measures (such as frequent hand-washing with soap and water; thoroughly cleaning fruit and vegetables; regularly dusting and sweeping the house; and preventing children sucking fingers, toys or pencils or eating soil).
- *Overcrowding*. In a South African study (von Schirnding et al. 1991) it was suggested that "the over-crowded nature of the homes could have a direct bearing on the quality of the care-giving environment, providing opportunity for children's activities to go unsupervised. This could lead young children to be more exposed to accessible sources of lead associated with poor housing conditions".

Poverty and low SES might also actually accentuate the effect of lead poisoning. An Australian cohort study found that, after adjustment for

a wide range of covariates, children from socially disadvantaged backgrounds were more sensitive to the effects of lead on their IQ relative to those of a higher SES (Tong et al. 2000a). This finding is generally consistent with the results of some (but not all) other studies of children. It is also consistent with the results of a recent animal study (Schneider et al. 2001).

5. Discussion

We estimated associations between risk factors and absolute poverty across the developing subregions of the world, and calculated attributable risks based on these, but without adjustment for confounding. Many of the causes of ill-health that we studied tend to weigh most heavily on the poorest groups in each of the 11 subregions. However, the socioeconomic gradient is not uniform across all the risk factors, and in some instances the associations vary between subregions. For some risk factors, variation between subregions in overall risk factor prevalence appears to be more important than variation by individual-level poverty within subregions. Our findings are based on a combination of quantitative analyses of existing survey data, and literature reviews for selected risk factors with some limitations.

5.1 Limitations

Conceptual limitations

We examined the individual-level association of poverty with risk factors, whereas previous work had examined the global differences in health status and poverty using country-level analyses (e.g. Gwatkin et al. 1999). Our presentation of individual-level analyses is arguably preferable to country- or subregional-level analyses of poverty and health. However, it is important not to let our analyses distract from two issues: societies tend to adopt health behaviours collectively; and the greatest income inequalities in the world are between countries rather than within countries. Regarding the former point, it is interesting to note that the between-region variation in the overall prevalence of risk factors such as tobacco use, indoor air pollution and unsafe water and sanitation was more strongly associated with individual risk than income poverty.

There are several other limitations to our analyses of the association of poverty with risk factors, including the inability to make causal inferences, lack of consideration of the role of time-lags, exclusion of residual income heterogeneity within level of absolute poverty, and exclusion of the heterogeneity of the association of poverty with risk factors within and between countries in each subregion. Consequently, the results must be interpreted as a mapping of risk factors by income poverty. They should *not* be interpreted as a determination of the causal association

of poverty and risk factors or as a determination of the causal impact fractions of poverty for various risk factors.

Quantitative analyses

The major limitation was data availability. First, both poverty and risk factor data were available for only some countries within each subregion, requiring extrapolations from those countries with data. Second, some subregions had no data at all for some risk factors (e.g. unsafe sex, alcohol), limiting the number of subregions for which we could conduct analyses. Third, all results were based on survey data with their own random and systematic errors.

It was necessary to make methodological assumptions in order to map risk factors by absolute poverty. The most important of these assumptions was that we could use the association of risk factors with a composite measure of socioeconomic position (e.g. asset scores in DHS data) to estimate the (unobserved) association with income poverty.

The definition and measurement of risk factors also need to be considered when interpreting the findings. In some instances, the variables used are indirect measures. Tobacco use, for example, was estimated in some subregions from household expenditure on tobacco (which would probably skew the distribution of use by socioeconomic group towards the better off). Other variables may not fully capture all aspects of risk. For example, alcohol use was also based on expenditure data.

5.2 Key findings

Having accepted the limitations described above, many striking results appeared that are highly unlikely to be spurious. For several important risk factors the effect of poverty appears to be very strong. Lack of improved water and sanitation, for instance, was 10–15 times more common in households living on <US$1 per day than in those living on more than >US$2 per day. Indeed, lack of adequate water and sanitation is widely accepted as being so closely aligned to poverty that it is used as a variable in asset scores to measure poverty. (Note that we did not use water or sanitation in the estimation of asset scores.) Worldwide, we estimated that 36–51% of instances of inadequate water and/or sanitation could be prevented if the prevalence of inadequate water and/or sanitation among people living on <US$2 per day was the same as that of people living on >US$2 per day (assuming a causal relationship).

We found a consistently strong association of absolute poverty with underweight children in all subregions except WPR-B, which is dominated by the China analyses (see Table 24.19). Previous research also found the same strong association, including for China (Chen 1996; Li et al. 1999; Popkin et al. 1993). Thus, our estimate of there being little association of poverty with underweight children in China is probably due to the small sample size in that survey. Putting China aside, the con-

sistency of the *relative* association of absolute poverty with underweight children across other subregions was notable. Put another way, regardless of where children live in the world the relative increase in underweight children with increasing socioeconomic deprivation is remarkably consistent. Across all developing countries, we estimated that 23–37% of cases of underweight could be prevented if the prevalence of underweight among children living on <US$ 2 per day was changed to that of children living on >US$ 2 per day. Assuming a causal relationship, poverty appears to have a strong and important association with child malnutrition.

The third risk factor in our quantitative analyses that demonstrated a strong association with poverty was indoor air pollution—although with some interesting heterogeneity between subregions (see Table 24.35). Regardless of the level of poverty in AFR-E, exposure to indoor air pollution appeared to be very high. In other subregions (often based on analyses for just one country per subregion), the associations appeared somewhat variable with regard to overall level of exposure to indoor air pollution and the relative risk association. Nevertheless, there was no exception to the general rule that those living in absolute poverty are exposed more frequently to indoor air pollution. Using the population impact fraction estimates (US$ 2 per day cut-off), we estimate that a third to a half of the exposure to indoor air pollution in developing countries could be averted if the poor had the exposure prevalence of the non-poor (assuming a causal relationship).

Beyond these three risk factors (underweight, water and sanitation and indoor air pollution), the associations we found in our quantitative analyses tended to be weaker and more variable between subregions. Data were limited in our quantitative analyses on unsafe sex, and the most reliable overall conclusion was of no consistent pattern of unsafe sex by poverty within subregions. However, behind the summary variable of "unsafe sex", it was interesting that "sex with a non-marital partner" and "condom use among those having sex with a non-marital partner" were less common among poor males and females. Also, there was some indication of variation in the association by sex. In AFR-E for instance, unsafe sex was associated with poverty among men but there was no apparent association for women (see Tables 24.26 and 24.27).

We carried out quantitative analyses for outdoor air, although these were not conducted in the same manner as for other risk factors. The results for outdoor air pollution show lower exposures to fine particles in those living on <US$ 1 per day in all subregions studied (see Table 24.37). Our method assigned people in urban or rural areas the same outdoor air pollution exposure regardless of income. Because rural areas have lower air pollution, and poor people tend to live in rural areas, this means that the poorest (on average) are exposed to lower levels of outdoor air pollution. However, this result is likely to obscure some

reverse socioeconomic patterns within urban centres, where it is the poorest people who live in the more polluted areas. This within-urban socioeconomic difference is likely to partially offset the urban–rural effect that gives rise to our results, but by how much is unclear. Without doubt, however, our method overstated the degree to which the poorest people are protected from outdoor air pollution relative to those who are less poor.

The quantitative analyses for overweight were based on DHS data (probably superior to LSMS data) supplemented by CHNS data, but only analyses for *maternal* overweight were possible. Nevertheless, a clear pattern emerged. Among the poorest subregions of the world (Africa and South East Asia) poor women were only half as likely to be overweight as the non-poor, whereas among subregions without quite the same depth of poverty (south-eastern Europe, central Asia, Central and South America) poor women were only about 20% less likely to be overweight. The relationship in WPR-B (dominated by China) was flat. We are uncertain about the accuracy of the data for China, but Yu et al. (2000) observed a higher prevalence of obesity among the least educated women in urban China. Our literature review tended to support the pattern of obesity becoming a greater problem among lower socioeconomic groups in South American countries. Popkin et al. (1995b) commented that the socioeconomic gradient in developing countries is shifting towards a developed world pattern of increased obesity among groups of low SES. Indeed, this is already occurring in some urbanized parts of Latin America. It is interesting to speculate whether the patterns we found were simply due to individuals in more poor subregions living in deeper poverty or whether this pattern is due to the socioeconomic gradient starting to "tip the other way" in countries with increasing affluence.

The final two risk factors for which we were able to conduct quantitative analyses—tobacco and alcohol consumption—showed weak and variable associations between subregions. The data used in the analyses were not of high quality, being often based on expenditure data and income. Summarizing the analyses as a (rather crude) estimate, there was little "net" association: the "total" population impact fractions for subregions with data combined were less than 10%. Nevertheless, there were some indications of between-subregion variability. For example, it is reasonable to conclude that if poor people in Africa (AFR-D and AFR-E) adopted the lifestyle of the non-poor, the prevalence of alcohol consumption might increase by approximately 20–40%.

In terms of tobacco and alcohol consumption, we may expect a variable association between subregions. It has often been the experience of the developed world that higher socioeconomic groups initially adopt adverse behavioural patterns (such as smoking) then discard them on learning of the health consequences. For various reasons, possibly asso-

ciated with economics or marketing, lower socioeconomic groups take up such behavioural patterns later. It is not unreasonable to expect the current pattern among developing countries to reflect historical patterns in the developed world. Moreover, it seems likely that such transitions might be at different stages in different subregions. Our quantitative analyses are not inconsistent with this conjecture, but neither can they offer strong support for it. For example, one might expect to see a stronger association of poverty with tobacco in subregions with relatively high prevalence of tobacco use, assuming that prevalence is related to the stage of the tobacco "epidemic". This was not apparent in our data.

Regarding the international literature on tobacco, the general pattern in developing countries is for a higher prevalence of smoking in those of lower SES, with a generally stronger association for men. Nevertheless, some studies in the poorer countries actually show the opposite pattern, with adults and adolescents of high SES having higher smoking rates, while some studies show no association. Our quantitative analyses mostly demonstrated a null association of poverty with prevalence of tobacco use except in EMR-D (i.e. Pakistan), where there seemed to be higher prevalence of tobacco use among more poor people (see Table 24.39). It should be noted that our quantitative analyses may have underestimated tobacco use among poor people, owing to our having to rely on expenditure data. Conversely, it is possible that the published literature has tended to arise from less poor areas of the developing world.

Our literature review demonstrated a tendency towards more hazardous alcohol consumption among lower socioeconomic groups in the developing world. However, the pattern varied between countries and studies. Our quantitative analyses were based on the prevalence of alcohol use, as indicated by expenditure on alcohol. This approach might have underestimated an adverse socioeconomic gradient for *hazardous* alcohol consumption.

The local effects of poverty are important, but it must be noted that variations in prevalence *between* subregions are frequently greater than the variations in risk factors by income *within* subregions. Tobacco use and exposure to indoor air pollution are two examples. The effect of poverty on smoking rates might be considered as a (not insignificant) ripple on a much larger wave, with "peak" subregions having smoking prevalence figures four times greater than those of the subregions in the "trough". Global tobacco control must deal with both the forces that generate disadvantage for lower income groups within populations and those forces (mostly on the supply side) operating at the level of countries and subregions.

We have included detailed literature reviews for risk factors for which we were unable to conduct quantitative analyses. For cholesterol and

lipids, hypertension and physical inactivity, the general patterns in the developing world were of either a mixed pattern or of more adverse profiles among higher socioeconomic groups. However, just as with obesity and overweight, it seems likely that in the future the average profile for these risk factors will worsen in the developing world, and become more concentrated among lower socioeconomic groups.

Acknowledgements

We thank Cara Marshall for assisting with the retrieval of journal articles; Emmanuela Gakidou for advice on factor analysis; and Majid Ezzati, Alan Lopez and Anthony Rodgers for ongoing guidance and advice. We are also indebted to the other risk factor teams, who forwarded data and reports to assist in the analysis.

Funding for this work was provided by WHO.

Notes

1 See preface for an explanation of this term.

2 For the 11 subregions with DHS data.

3 Referred to as scenarios in chapter 16.

References

Abelin T, Schlettwein-Gsell D (1986) [Handicaps and needs of the elderly. A multifactor epidemiologic study under urban conditions]. *Schweizerische Medizinische Wochenschrift. Journal Suisse de Médecine*, **116**:1524–1542.

Acheson D (1998) *Independent inquiry into inequalities in health report*. The Stationery Office, London.

Adamson SJ, Sellman JD, Futterman-Collier A et al. (2000) A profile of alcohol and drug clients in New Zealand: results from the 1998 national telephone survey. *New Zealand Medical Journal*, **113**:414–416.

Agarwal AK, Yunus M, Khan A, Ahmad J (1994) A clinical-epidemiological study of hypertension in rural population of Jawan Block, Distt, Aligarh (UP) India. *Journal of the Royal Society of Health*, **114**:17–19.

Ahsan H, Underwood P, Atkinson D (1998) Smoking among male teenagers in Dhaka, Bangladesh. *Preventive Medicine*, **27**:70–76.

Alam SE (1998) Prevalence and pattern of smoking in Pakistan. *JPMA—Journal of the Pakistan Medical Association*, **48**:64–66.

Alem A, Kebede D, Kullgren G (1999) The epidemiology of problem drinking in Butajira, Ethiopia. *Acta Psychiatrica Scandinavica, Supplementum*, **397**:77–83.

Allain TJ, Wilson AO, Gomo ZA, Adamchak DJ, Matenga JA (1997) Diet and nutritional status in elderly Zimbabweans. *Age and Ageing*, **26**:463–470.

al-Mannai A, Dickerson JW, Morgan JB, Khalfan H (1996) Obesity in Bahraini adults. *Journal of the Royal Society of Health*, **116**:30–2, 37–40.

al-Nuaim AA, Bamgboye EA, al-Rubeaan KA, al-Mazrou Y (1997) Overweight and obesity in Saudi Arabian adult population, role of socio-demographic variables. *Journal of Community Health*, **22**:211–223.

al-Nuaim AR, al-Rubeaan K, al-Mazrou Y, al-Attas O, al-Daghari N, Khoja T (1996) High prevalence of overweight and obesity in Saudi Arabia. *International Journal of Obesity and Related Metabolic Disorders*, **20**:547–552.

Amonoo-Lartson R, Pappoe ME (1992) Prevalence of smoking in secondary schools in the greater Accra region of Ghana. *Social Science and Medicine*, **34**:1291–1293.

Anonymous (1977) Race, education and prevalence of hypertension. *American Journal of Epidemiology*, **106**:351–361.

Arnesen E, Forsdahl A (1985) The Tromso Heart Study: coronary risk factors and their association with living conditions during childhood. *Journal of Epidemiology and Community Health*, **39**:210–214.

Azcona-Cruz MI, Rothenberg SJ, Schnaas L, Zamora-Munoz JS, Romero-Placeres M (2000) Lead-glazed ceramic ware and blood lead levels of children in the city of Oaxaca, Mexico. *Archives of Environmental Health*, **55**:217–222.

Baghurst PA, Tong S, Sawyer MG, Burns J, McMichael AJ (1999) Sociodemographic and behavioural determinants of blood lead concentrations in children aged 11–13 years. The Port Pirie Cohort Study. *Medical Journal of Australia*, **170**:63–67.

Bartley M, Fitzpatrick R, Firth D, Marmot M (2000) Social distribution of cardiovascular disease risk factors: change among men in England 1984–1993. *Journal of Epidemiology and Community Health*, **54**:806–814.

Bener A, al-Ketbi LM (1999) Cigarette smoking habits among high school boys in a developing country. *Journal of the Royal Society of Health*, **119**: 166–169.

Benetou V, Chloptsios Y, Zavitsanos X, Karalis D, Naska A, Trichopoulou A (2000) Total cholesterol and HDL-cholesterol in relation to socioeconomic status in a sample of 11 645 Greek adults: the EPIC study in Greece. European prospective investigation into nutrition and cancer. *Scandinavian Journal of Public Health*, **28**:260–265.

Bennett S (1995) Cardiovascular risk factors in Australia: trends in socioeconomic inequalities. *Journal of Epidemiology and Community Health*, **49**: 363–372.

Bilal AM, Khattar MA, Hassan KI, Berry D (1992) Psychosocial and toxicological profile of drug misuse in male army conscripts in Kuwait. *Acta Psychiatrica Scandinavica*, **86**:104–107.

Blakely T (2002) *The New Zealand census-mortality study: socioeconomic inequalities and adult mortality 1991–94*. Ministry of Health, Wellington.

Blakely T, Pearce N (2002) Socio-economic position is more than just NZDep. *New Zealand Medical Journal*, **115**:109–111.

Blakely T, Woodward A, Pearce N, Salmond C, Kiro C, Davis P (2002) Socioeconomic factors and mortality among 25–64 year olds followed from 1991 to 1994: The New Zealand Census-Mortality Study. *New Zealand Medical Journal*, **115**:93–97.

Blanksby BA, Anderson MJ, Douglas GA (1996) Recreational patterns, body composition and socioeconomic status of western Australian secondary school students. *Annals of Human Biology*, **23**:101–112.

Blum A (1997) Cancer prevention: Preventing tobacco-related cancers. In: *Cancer: principles and practice of oncology.* DeVita V, Hellman S, Rosenburg S, eds. Lippincott-Raven, Philadelphia, PA.

Boardman JD, Finch BK, Ellison CG, Williams DR, Jackson JS (2001) Neighborhood disadvantage, stress, and drug use among adults. *Journal of Health and Social Behavior*, **42**:151–165.

Borman B, Wilson N, Mailing C (1999) Socio-demographic characteristics of New Zealand smokers: results from the 1996 census. *New Zealand Medical Journal*, **112**:460–463.

Borrell C, Rue M, Pasarin MI, Rohlfs I, Ferrando J, Fernandez E (2000) Trends in social class inequalities in health status, health-related behaviors, and health services utilization in a southern European urban area (1983–1994). *Preventive Medicine*, **31**:691–701.

Bouchardy C, Parkin DM, Khlat M et al. (1993) Education and mortality from cancer in Sao Paulo, Brazil. *Annals of Epidemiology*, **3**:64–70.

Bray GA, Popkin BM (1998) Dietary fat intake does affect obesity! *American Journal of Clinical Nutrition*, **68**:1157–1173.

Brownsberger W (1997) Prevalence of frequent cocaine use in urban poverty areas. *Contemporary Drug Problems*, **24**:349–371.

Brunner E, Shipley MJ, Blane D, Smith GD, Marmot MG (1999) When does cardiovascular risk start? Past and present socioeconomic circumstances and risk factors in adulthood. *Journal of Epidemiology and Community Health*, **53**:757–764.

Bunker CH, Ukoli FA, Nwankwo MU et al. (1992) Factors associated with hypertension in Nigerian civil servants. *Preventive Medicine*, **21**:710–722.

Carlini EA (1990) Research is badly needed to improve programmes for the prevention and treatment of drug abuse and drug dependence in Brazil. *Drug and Alcohol Dependence*, **25**:169–173.

Carr S, Ager A, Nyando C, Moyo K, Titeca A, Wilkinson M (1994) A comparison of chamba (marijuana) abusers and general psychiatric admissions in Malawi. *Social Science and Medicine*, **39**:401–406.

Casey PH, Szeto K, Lensing S, Bogle M, Weber J (2001) Children in food-insufficient, low-income families: prevalence, health, and nutrition status. *Archives of Pediatrics and Adolescent Medicine*, **155**:508–514.

Casswell S, Gordon A (1984) Drinking and occupational status in New Zealand men. *Journal of Studies on Alcohol*, **45**:144–148.

Chaturvedi HK, Phukan RK, Zoramtharga K, Hazarika NC, Mahanta J (1998) Tobacco use in Mizoram, India: Sociodemographic differences in pattern. *Southeast Asian Journal of Tropical Medicine and Public Health*, **29**:66–70.

Chen CM (1996) Nutrition status of the Chinese people. *Biomedical and Environmental Sciences*, **9**:81–92.

Chen S, Ravallion M (2000) *How did the world's poorest fare in the 1990s?* World Bank, Washington, DC. Accessed 29/11/01 at http://www.worldbank.org/research/povmon itor/pdfs/methodology.pdf.

Chhabra SK, Rajpal S, Gupta R (2001) Patterns of smoking in Delhi and comparison of chronic respiratory morbidity among beedi and cigarette smokers. *Indian Journal of Chest Diseases and Allied Sciences*, **43**:19–26.

Chinn DJ, White M, Harland J, Drinkwater C, Raybould S (1999) Barriers to physical activity and socioeconomic position: implications for health promotion. *Journal of Epidemiology and Community Health*, **53**:191–192.

Chung MH, Chung KK, Chung CS, Raymond JS (1992) Health-related behaviors in Korea: Smoking, drinking, and perinatal care. *Asia-Pacific Journal of Public Health*, **6**:10–15.

Colhoun H, Hemingway H, Poulter NR (1998) Socioeconomic status and blood pressure: an overview analysis. *Journal of Human Hypertension*, **12**:91–110.

Commission on Macroeconomics and Health (2001) *Investing in health for economic development*. World Health Organization, Geneva.

Connolly GN (1992) Worldwide expansion of transnational tobacco industry. *Journal of the National Cancer Institute. Monographs*, **12**:29–35.

Crampton P, Salmond C, Woodward A, Reid P (2000) Socioeconomic deprivation and ethnicity are both important for anti-tobacco health promotion. *Health Education and Behavior*, **27**:317–327.

Crawford A (1988) Self-reported alcohol consumption among population subgroups in three areas of Britain. *Drug and Alcohol Dependence*, **21**:161–167.

Crespo CJ, Ainsworth BE, Keteyian SJ, Heath GW, Smit E (1999) Prevalence of physical inactivity and its relation to social class in U.S. adults: results from the third national health and nutrition examination survey, 1988–1994. *Medicine and Science in Sports and Exercise*, **31**:1821–1827.

Crespo CJ, Smit E, Andersen RE, Carter-Pokras O, Ainsworth BE (2000) Race/ethnicity, social class and their relation to physical inactivity during leisure time: results from the third national health and nutrition examination survey, 1988–1994. *American Journal of Preventive Medicine*, **18**:46–53.

Crum RM, Lillie-Blanton M, Anthony JC (1996) Neighborhood environment and opportunity to use cocaine and other drugs in late childhood and early adolescence. *Drug and Alcohol Dependence*, **43**:155–161.

Cuello C, Correa P, Haenszel W (1982) Socioeconomic class differences in cancer incidence in Cali, Colombia. *International Journal of Cancer*, **29**:637–643.

Davey Smith G, Hart C, Watt G, Hole D, Hawthornem V (1998) Individual social class, area-based deprivation, cardiovascular disease risk factors, and

mortality: the Renfrew and Paisley Study. *Journal of Epidemiology and Community Health*, **52**:399–405.

Davis P, McLeod K, Ransom M, Ongley P, Pearce N, Howden-Chapman P (1999) The New Zealand socioeconomic index: developing and validating an occupationally-derived indicator of socioeconomic status. *Australian and New Zealand Journal of Public Health*, **23**:27–33.

Day NL, Cottreau CM, Richardson GA (1993) The epidemiology of alcohol, marijuana, and cocaine use among women of childbearing age and pregnant women. *Clinical Obstetrics and Gynecology*, **36**:232–245.

Delpeuch F, Cornu A, Massamba JP, Traissac P, Maire B (1994) Is body mass index sensitively related to socioeconomic status and to economic adjustment? A case study from the Congo. *European Journal of Clinical Nutrition*, **48**:S141–147.

Delpeuch F, Traissac P, Martin-Prevel Y, Massamba JP, Maire B (2000) Economic crisis and malnutrition: socioeconomic determinants of anthropometric status of preschool children and their mothers in an African urban area. *Public Health Nutrition*, **3**:39–47.

Dhurandhar NV, Kulkarni PR (1992) Prevalence of obesity in Bombay. *International Journal of Obesity and Related Metabolic Disorders*, **16**:367–375.

Diez-Roux A (1998) Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American Journal of Public Health*, **88**:216–222.

Doak CM, Adair LS, Monteiro C, Popkin BM (2000) Overweight and underweight coexist within households in Brazil, China and Russia. *Journal of Nutrition*, **130**:2965–2971.

Dressler WW, Evans P, Gray DJ (1992) Status incongruence and serum cholesterol in an English general practice. *Social Science and Medicine*, **34**:757–762.

Dressler WW, Santos JE (2000) Social and cultural dimensions of hypertension in Brazil: a review. *Cadernos de Saude Publica*, **16**:303–315.

Drever F, Whitehead M, eds. (1997) *Health inequalities*. Office for National Statistics, London.

Droomers M, Schrijvers CT, van de Mheen H, Mackenbach JP (1998) Educational differences in leisure-time physical inactivity: a descriptive and explanatory study. *Social Science and Medicine*, **47**:1665–1676.

Dryson E, Martorell R (2000) Poor fetal and child growth and later obesity and chronic disease: relevance for Latin America. In: *Obesity and poverty: a new public health challenge*. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.

Dryson E, Metcalf P, Baker J, Scragg R (1992) The relationship between body mass index and socioeconomic status in New Zealand: ethnic and occupational factors. *New Zealand Medical Journal*, **105**:233–235.

Duncan BB, Schmidt MI, Achutti AC, Polanczyk CA, Benia LR, Maia AA (1993) Socioeconomic distribution of noncommunicable disease risk factors in urban Brazil: the case of Porto Alegre. *Bulletin of the Pan American Health Organization*, **27**:337–349.

Edwards R, Brown JS, Hodgson P, Kyle D, Reed D, Wallace B (1999) An action plan for tobacco control at regional level. *Public Health*, **113**:165–170.

Eide AH, Acuda SW (1996) Adolescents' drug use in Zimbabwe—comparing two recent studies. *Central African Journal of Medicine*, **42**:128–135.

Ekpo EB, Udofia O, Eshiet NF, Andy JJ (1992) Demographic, life style and anthropometric correlates of blood pressure of Nigerian urban civil servants, factory and plantation workers. *Journal of Human Hypertension*, **6**:275–280.

Elreedy S, Krieger N, Ryan PB, Sparrow D, Weiss ST, Hu H (1999) Relations between individual and neighborhood-based measures of socioeconomic position and bone lead concentrations among community-exposed men: the normative aging study. *American Journal of Epidemiology*, **150**:129–141.

Ensminger ME, Anthony JC, McCord J (1997) The inner city and drug use: initial findings from an epidemiological study. *Drug and Alcohol Dependence*, **48**:175–184.

Evans T, Whitehead M, Diderichsen F, Bhuiya A, Wirth M (2001) *Challenging inequalities in health: from ethics to action*, Oxford University Press, Oxford.

Farias P, Borja-Aburto VH, Rios C, Hertz-Picciotto I, Rojas-Lopez M, Chavez-Ayala R (1996) Blood lead levels in pregnant women of high and low socioeconomic status in Mexico City. *Environmental Health Perspectives*, **104**: 1070–1074.

Farrow SC, Charny MC, Lewis PC (1988) A community survey of alcohol consumption. *Alcohol and Alcoholism*, **23**:315–322.

Fawcett JP, Williams SM, Heydon JL, Walmsley TA, Menkes DB (1996) Distribution of blood lead levels in a birth cohort of New Zealanders at age 21. *Environmental Health Perspectives*, **104**:1332–1335.

Fergusson DM, Horwood LJ (1999) Prospective childhood predictors of deviant peer affiliations in adolescence. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **40**:581–592.

Fergusson DM, Horwood LJ (2000) Cannabis use and dependence in a New Zealand birth cohort. *New Zealand Medical Journal*, **113**:156–158.

Fergusson DM, Woodward LJ, Horwood LJ (1998) Maternal smoking during pregnancy and psychiatric adjustment in late adolescence. *Archives of General Psychiatry*, **55**:721–727.

Fernandez E, Schiaffino A, Garcia M, Borras JM (2001) Widening social inequalities in smoking cessation in Spain, 1987–1997. *Journal of Epidemiology and Community Health*, **55**:729–730.

Filmer D, Pritchett L (1988) *Estimating wealth effects without expenditure, data or tears*. Development Economics Research Group, Washington, DC.

Finch BK, Vega WA, Kolody B (2001) Substance use during pregnancy in the state of California, USA. *Social Science and Medicine*, **52**:571–583.

Forero R, Bauman A, Chen JX, Flaherty B (1999) Substance use and sociodemographic factors among aboriginal and Torres Strait Islander school students in New South Wales. *Australian and New Zealand Journal of Public Health*, **23**:295–300.

Gan CY (1995) Smokeless tobacco use among rural Kadazan women in Sabah, Malaysia. *Southeast Asian Journal of Tropical Medicine and Public Health*, **26**:291–296.

Gan CY (1998) Tobacco usage among rural Bajaus in Sabah, Malaysia. *Southeast Asian Journal of Tropical Medicine and Public Health*, **29**:643–648.

Gao W, Li Z, Kaufmann RB et al. (2001) Blood lead levels among children aged 1 to 5 years in Wuxi City, China. *Environmental Research*, **87**:11–19.

Ge K, Weisell R, Guo X et al. (1994) The body mass index of Chinese adults in the 1980s. *European Journal of Clinical Nutrition*, **48**:S148–154.

Glanville J, Lefebvre C (2000) Identifying systematic reviews: key resources. *Evidence-Based Mental Health*, **3**:68–69.

Glendinning A, Shucksmith J, Hendry L (1994) Social class and adolescent smoking behaviour. *Social Science and Medicine*, **38**:1449–1460.

Gold MR, Franks P (1990) The social origin of cardiovascular risk: an investigation in a rural community. *International Journal of Health Services*, **20**: 405–416.

Gonzalez MA, Rodriguez Artalejo F, Calero JR (1998) Relationship between socioeconomic status and ischaemic heart disease in cohort and case-control studies: 1960–1993. *International Journal of Epidemiology*, **27**: 350–358.

Green G, Macintyre S, West P, Ecob R (1991) Like parent like child? Associations between drinking and smoking behaviour of parents and their children. *British Journal of Addiction*, **86**:745–758.

Grol ME, Eimers JM, Alberts JF et al. (1997a) Alarmingly high prevalence of obesity in Curacao: data from an interview survey stratified for socioeconomic status. *International Journal of Obesity and Related Metabolic Disorders*, **21**:1002–1009.

Grol ME, Halabi YT, Gerstenbluth I, Alberts JF, O'Niel J (1997b) Lifestyle in Curacao. Smoking, alcohol consumption, eating habits and exercise. *West Indian Medical Journal*, **46**:8–14.

Gupta R, Gupta VP, Ahluwalia NS (1994) Educational status, coronary heart disease, and coronary risk factor prevalence in a rural population of India. *British Medical Journal*, **309**:1332–1336.

Gwatkin D, Guillot M, Heuveline P (1999) The burden of disease among the global poor. *The Lancet,* **354**:586–589.

Hadi A (2000) A participatory approach to sanitation: experience of Bangladeshi NGOs. *Health Policy and Planning*, **15**:332–337.

Hamadeh RR, Musaiger AO (2000) Lifestyle patterns in smokers and non-smokers in the state of Bahrain. *Nicotine and Tobacco Research*, **2**:65–69.

Hans SL (1999) Demographic and psychosocial characteristics of substance-abusing pregnant women. *Clinics in Perinatology*, **26**:55–74.

Hay DR, Foster FH (1984) Intercensal trends in cigarette smoking in New Zealand 2: social and occupational factors. *New Zealand Medical Journal*, **97**:395–398.

Helmert U, Mielck A, Shea S (1997) Poverty, health, and nutrition in Germany. *Reviews on Environmental Health*, **12**:159–170.

Hoa DP, Thanh HT, Hojer B, Persson LA (1995) Young child feeding in a rural area in the Red River Delta, Vietnam. *Acta Paediatrica*, **84**:1045–1049.

Hodge AM, Dowse GK, Toelupe P, Collins VR, Imo T, Zimmet PZ (1994) Dramatic increase in the prevalence of obesity in Western Samoa over the 13 year period 1978–1991. *International Journal of Obesity and Related Metabolic Disorders*, **18**:419–428.

Howden-Chapman P, Tobias M, eds. (2000) *Social inequalities in health: New Zealand 1999,* Ministry of Health, Wellington.

Howell EM, Heiser N, Harrington M (1999) A review of recent findings on substance abuse treatment for pregnant women. *Journal of Substance Abuse Treatment*, **16**:195–219.

Hu TW, Lin Z, Keeler TE (1998) Teenage smoking, attempts to quit, and school performance. *American Journal of Public Health*, **88**:940–943.

Hu TW, Tsai YW (2000) Cigarette consumption in rural China: survey results from 3 provinces. *American Journal of Public Health*, **90**:1785–1787.

ICRG (1988) Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. INTERSALT Cooperative Research Group. *British Medical Journal*, **297**: 319–328.

INCLEN (1994) Socioeconomic status and risk factors for cardiovascular disease: a multicentre collaborative study in the international clinical epidemiology network (INCLEN). The INCLEN multicentre collaborative group. *Journal of Clinical Epidemiology*, **47**:1401–1409.

Irala-Estevez JD, Groth M, Johansson L, Oltersdorf U, Prattala R, Martinez-Gonzalez MA (2000) A systematic review of socioeconomic differences in food habits in Europe: consumption of fruit and vegetables. *European Journal of Clinical Nutrition*, **54**:706–714.

Jackson R, Beaglehole R, Yee RL, Small C, Scragg R (1990) Trends in cardiovascular risk factors in Auckland, 1982 to 1987. *New Zealand Medical Journal*, **103**:363–365.

Jarallah JS, al-Rubeaan KA, al-Nuaim AR, al-Ruhaily AA, Kalantan KA (1999) Prevalence and determinants of smoking in three regions of Saudi Arabia. *Tobacco Control*, **8**:53–56.

Jeffery RW, French SA, Forster JL, Spry VM (1991) Socioeconomic status differences in health behaviors related to obesity: the healthy worker project. *International Journal of Obesity*, **15**:689–696.

Johnson RK, Wang MQ, Smith MJ, Connolly G (1996) The association between parental smoking and the diet quality of low-income children. *Pediatrics*, **97**:312–317.

Jutkowitz JM, Spielmann H, Koehler U, Lohani J, Pande A (1997) Drug use in Nepal: the view from the street. *Substance Use and Misuse*, **32**:987–1004.

Kadiri S, Walker O, Salako BL, Akinkugbe O (1999) Blood pressure, hypertension and correlates in urbanised workers in Ibadan, Nigeria: a revisit. *Journal of Human Hypertension*, **13**:23–27.

Kaiser R, Henderson AK, Daley WR et al. (2001) Blood lead levels of primary school children in Dhaka, Bangladesh. *Environmental Health Perspectives*, **109**:563–566.

Kaufman J, Cooper R (1999) Seeking causal explanations in social epidemiology. *American Journal of Epidemiology*, **150**:113–120.

Kaufmann RB, Clouse TL, Olson DR, Matte TD (2000) Elevated blood lead levels and blood lead screening among us children aged one to five years: 1988–1994. *Pediatrics*, **106**:E79.

Kawachi I, Marshall S, Pearce N (1991) Social class inequalities in the decline of coronary heart disease among New Zealand men, 1975–1977 to 1985–1987. *International Journal of Epidemiology*, **20**:393–398.

Kebede D, Alem A (1999) The epidemiology of alcohol dependence and problem drinking in Addis Ababa, Ethiopia. *Acta Psychiatrica Scandinavica, Supplementum*, **397**:30–34.

Kikafunda JK, Walker AF, Collett D, Tumwine JK (1998) Risk factors for early childhood malnutrition in Uganda. *Pediatrics*, **102**:E45.

King M (1986) At risk drinking among general practice attenders: Prevalence, characteristics and alcohol-related problems. *British Journal of Psychiatry*, **148**:533–540.

Klemp P, Robertson MC, Stansfield S, Klemp JA, Harding E (1998) Factors associated with smoking and the reasons for stopping in Maori and Europeans: a comparative study. *New Zealand Medical Journal*, **111**:148–501.

Knuiman JT, West CE, Burema J (1982) Serum total and high density lipoprotein cholesterol concentrations and body mass index in adult men from 13 countries. *American Journal of Epidemiology*, **116**:631–642.

Knupfer G. (1989) The prevalence in various social groups of eight different drinking patterns, from abstaining to frequent drunkenness: analysis of 10 US surveys combined. *British Journal of Addiction*, **11**:1305–1318.

Koong SL, Serdula MK, Williamson DF, Malison MD, Davis RM (1991) Smoking prevalence in the United States and Taipei City, Taiwan. *American Journal of Preventive Medicine*, **7**:161–165.

Krieger N (2001) A glossary for social epidemiology. *Journal of Epidemiology and Community Health*, **56**:693–700.

Kunst A, Groenhof F, Mackenbach J, The EU Working Group on Socioeconomic Inequalities in Health (1998a) Mortality by occupational class among men 30–64 years in 11 European countries. *Social Science and Medicine*, **46**:1459–1476.

Kunst A, Groenhof F, Mackenbach J, The EU Working Group on Socioeconomic Inequalities in Health (1998b) Occupational class and cause specific mortality in middle-aged men in 11 European countries: comparison of population based studies. *British Medical Journal*, **316**:1636–1642.

Kushwaha KP, Singh YD, Rathi AK, Singh KP, Rastogi CK (1992) Prevalence and abuse of psychoactive substances in children and adolescents. *Indian Journal of Pediatrics*, **59**:261–268.

Lapham SC, Skipper BJ, Brown P, Chadbunchachai W, Suriyawongpaisal P, Paisarnsilp S (1998) Prevalence of alcohol problems among emergency room patients in Thailand. *Addiction*, **93**:1231–1239.

Leon D, Walt G, eds. (2001) *Poverty, inequality and health*. Oxford University Press, Oxford.

Li Y, Guo G, Shi A, Anme T, Ushijima H (1999) Prevalence and correlates of malnutrition among children in rural minority areas of China. *Pediatrics International*, **41**:549–556.

Lindstrom M, Hanson BS, Ostergren PO (2001) Socioeconomic differences in leisure-time physical activity: the role of social participation and social capital in shaping health related behaviour. *Social Science and Medicine*, **52**:441–451.

Linn S, Fulwood R, Rifkind B et al. (1989) High density lipoprotein cholesterol levels among us adults by selected demographic and socioeconomic variables. The Second National Health and Nutrition Examination Survey 1976–1980. *American Journal of Epidemiology*, **129**:281–294.

Luthar SS, D'Avanzo K (1999) Contextual factors in substance use: a study of suburban and inner-city adolescents. *Development and Psychopathology*, **11**:845–867.

Mackay J (1992) US tobacco export to Third World: Third World War. *Journal of the National Cancer Institute. Monographs*, **12**:25–28.

Mackenbach J, Kunst A (1997) Measuring the magnitude of socioeconomic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Social Science and Medicine*, **44**:757–771.

Mackenbach J, Kunst A, Cavelaars A, Groenhof F, Geurts J (1997) Socioeconomic inequalities in morbidity and mortality in western Europe. *The Lancet*, **349**:1655–1659.

Mahaffey KR (1990) Environmental lead toxicity: nutrition as a component of intervention. *Environmental Health Perspectives*, **89**:75–78.

Mahley RW, Arslan P, Pekcan G et al. (2001) Plasma lipids in Turkish children: impact of puberty, socioeconomic status, and nutrition on plasma cholesterol and HDL. *Journal of Lipid Research*, **42**:1996–2006.

Mancilha-Carvalho JJ, Crews DE (1990) Lipid profiles of Yanomamo Indians of Brazil. *Preventive Medicine*, **19**:66–75.

Markovic N, Bunker CH, Ukoli FA, Kuller LH (1998) John Henryism and blood pressure among Nigerian civil servants. *Journal of Epidemiology and Community Health*, **52**:186–190.

Martikainen P, Ishizaki M, Marmot MG, Nakagawa H, Kagamimori S (2001) Socioeconomic differences in behavioural and biological risk factors: a comparison of a Japanese and an English cohort of employed men. *International Journal of Epidemiology*, **30**:833–838.

Martins IS, Coelho LT, Casajus MI, Okani ET (1995) Smoking, consumption of alcohol and sedentary life style in population grouping and their relationships with lipemic disorders. *Revista de Saude Publica*, **29**:38–45.

Martorell R, Kettel Khan L, Hughes ML, Grummer-Strawn LM (2000) Overweight and obesity in preschool children from developing countries. *International Journal of Obesity and Related Metabolic Disorders*, **24**:959–967.

Martorell R, Khan LK, Hughes ML, Grummer-Strawn LM (1998) Obesity in Latin American women and children. *Journal of Nutrition*, **128**:1464–1473.

Memon A, Moody PM, Sugathan TN et al. (2000) Epidemiology of smoking among Kuwaiti adults: prevalence, characteristics, and attitudes. *Bulletin of the World Health Organization*, **78**:1306–1315.

Mensink GB, Loose N, Oomen CM (1997) Physical activity and its association with other lifestyle factors. *European Journal of Epidemiology*, **13**:771–778.

Midanik LT, Clark WB (1994) The demographic distribution of US drinking patterns in 1990: Description and trends from 1984. *American Journal of Public Health*, **84**:1218–1222.

Misra A, Pandey RM, Devi JR, Sharma R, Vikram NK, Khanna N (2001) High prevalence of diabetes, obesity and dyslipidaemia in urban slum population in northern India. *International Journal of Obesity*, **25**:1722–1729.

Ministry of Health (1999) *Taking the pulse: The 1996/97 New Zealand Health Survey*. Ministry of Health, Wellington.

Ministry of Health (2001) *Inhaling inequality: tobacco's contribution to health inequality in New Zealand*. Ministry of Health, Wellington.

Molarius A, Seidell JC, Sans S, Tuomilehto J, Kuulasmaa K (2000) Educational level, relative body weight, and changes in their association over 10 years: an international perspective from the WHO MONICA Project. *American Journal of Public Health*, **90**:1260–1268.

Moller H, Tonnesen H (1997) Alcohol drinking, social class and cancer. *IARC Scientific Publications*, **138**:251–263.

Monteiro C (2000) The epidemiologic transition in Brazil. In: *Obesity and poverty: a new public health challenge*. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.

Monteiro CA, Mondini L, de Souza AL, Popkin BM (1995) The nutrition transition in Brazil. *European Journal of Clinical Nutrition*, **49**:105–113.

Moreira LB, Fuchs FD, Moraes RS et al. (1996) Alcoholic beverage consumption and associated factors in Porto Alegre, a southern Brazilian city: a population-based survey. *Journal of Studies on Alcohol*, **57**:253–259.

Morgan D, Grant KA, Gage HD et al. (2002) Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration. *Nature Neuroscience*, **5**:169–174.

Naidu AN, Rao NP (1994) Body mass index: a measure of the nutritional status in Indian populations. *European Journal of Clinical Nutrition*, **48**:S131–140.

Narayan KM, Chadha SL, Hanson RL et al. (1996) Prevalence and patterns of smoking in Delhi: cross sectional study. *British Medical Journal*, **312**: 1576–1579.

Nriagu J, Jinabhai CC, Naidoo R, Coutsoudis A (1997) Lead poisoning of children in Africa, II. Kwazulu/Natal, South Africa. *Science of the Total Environment*, **197**:1–11.

Nyberg K, Allebeck P, Eklund G, Jacobson B (1992) Socioeconomic versus obstetric risk factors for drug addiction in offspring. *British Journal of Addiction*, **87**:1669–1676.

Obot IS (1990) The use of tobacco products among Nigerian adults: a general population survey. *Drug and Alcohol Dependence*, **26**:203–208.

Omokhodion FO, Umar US, Ogunnowo BE (2000) Prevalence of low back pain among staff in a rural hospital in Nigeria. *Occupational Medicine (Oxford)*, **50**:107–110.

Omoluabi PF (1995) A review of the incidence of nonprescription psychoactive substance use/misuse in Nigeria. *International Journal of the Addictions*, **30**:445–458.

Onat A, Surdum-Avci G, Senocak M, Ornek E, Gozukara Y (1992) Plasma lipids and their interrelationship in Turkish adults. *Journal of Epidemiology and Community Health*, **46**:470–476.

Ozanne SE (2001) Metabolic programming in animals. *British Medical Bulletin*, **60**:143–152.

Paeratakul S, Popkin BM, Keyou G, Adair LS, Stevens J (1998) Changes in diet and physical activity affect the body mass index of Chinese adults. *International Journal of Obesity and Related Metabolic Disorders*, **22**:424–431.

Pamuk E, Makuc D, Heck k, Reuben C, Lochner K (1998) *Socioeconomic status and health chartbook. Health, United States, 1998*. National Center for Health Statistics, Hyattsville, MD.

Parsons TJ, Power C, Logan S, Summerbell CD (1999) Childhood predictors of adult obesity: a systematic review. *International Journal of Obesity and Related Metabolic Disorders*, **23**:S1–107.

Pavan L, Casiglia E, Braga LM et al. (1999) Effects of a traditional lifestyle on the cardiovascular risk profile: the Amondava population of the Brazilian Amazon. Comparison with matched African, Italian and Polish populations. *Journal of Hypertension*, **17**:749–756.

Pearce N (1997) Why study socioeconomic factors and cancer? *IARC Scientific Publications*, **138**:17–23.

Pena M, Bacallao J (2000) Obesity among the poor: an emerging problem in Latin America and the Caribbean. In: *Obesity and poverty: a new public health challenge*. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.

Pequignot G, Crosignani P, Terracini B et al. (1988) A comparative study of smoking, drinking and dietary habits in population samples in France, Italy, Spain and Switzerland. III. Consumption of alcohol. *Revue d'Épidémiologie et de Santé Publique*, **36**:177–185.

Pereira MA, Kriska AM, Collins VR et al. (1998) Occupational status and cardiovascular disease risk factors in the rapidly developing, high-risk population of Mauritius. *American Journal of Epidemiology*, **148**:148–159.

Perova NV, Davis CE, Tao S et al. (2001) Multi-country comparison of plasma lipid relationship to years of schooling in men and women. *International Journal of Epidemiology*, **30**:371–379.

Pickering T (1999) Cardiovascular pathways: socioeconomic status and stress effects on hypertension and cardiovascular function. *Annals of the New York Academy of Sciences*, **896**:262–277.

Pinto JA, Ruff AJ, Paiva JV et al. (1994) HIV risk behavior and medical status of underprivileged youths in Belo Horizonte, Brazil. *Journal of Adolescent Health*, **15**:179–185.

Pirkle JL, Brody DJ, Gunter EW et al. (1994) The decline in blood lead levels in the United States. The National Health and Nutrition Examination Surveys (NHANES). *Journal of the American Medical Association*, **272**: 284–291.

Pomare E, Keefe-Ormsby V, Ormsby C et al. (1995) *Hauora: Maori standards of health III*. Te Ropu Rangahau Hauora a Eru Pomare; Wellington School of Medicine, Wellington.

Popkin BM (2001) The nutrition transition and obesity in the developing world. *Journal of Nutrition*, **131**:S871–873.

Popkin BM, Keyou G, Zhai F, Guo X, Ma H, Zohoori N (1993) The nutrition transition in China: a cross-sectional analysis. *European Journal of Clinical Nutrition*, **47**:333–346.

Popkin BM, Paeratakul S, Zhai F, Ge K (1995a) Dietary and environmental correlates of obesity in a population study in China. *Obesity Research*, **3**:S135–143.

Popkin BM, Paeratakul S, Zhai F, Ge K (1995b) A review of dietary and environmental correlates of obesity with emphasis on developing countries. *Obesity Research*, **3**:S145–153.

Potter JD (1997) Diet and cancer: Possible explanations for the higher risk of cancer in the poor. *IARC Scientific Publications*, **138**:S265–283.

Powell DL, Stewart V (2001) Children. The unwitting target of environmental injustices. *Pediatric Clinics of North America*, **48**:1291–1305.

Power C, Parsons T (2000) Nutritional and other influences in childhood as predictors of adult obesity. *Proceedings of the Nutrition Society*, **59**:267–272.

Prothro JW, Rosenbloom CA (1999) Description of a mixed ethnic, elderly population. I. Demography, nutrient/energy intakes, and income status. *Journals of Gerontology, Series A—Biological Sciences and Medical Sciences*, **54**:M315–324.

Prüss A, Kay D, Fewtrell L, Bartam J (2001) *Estimating the burden of disease due to water, sanitation, and hygiene at the global level*. World Health Organization, Geneva.

Radebe BZ, Brady P, Siziya S, Todd H (1996) Maternal risk factors for childhood malnutrition in the Mazowe district of Zimbabwe. *Central African Journal of Medicine*, **42**:240–244.

Reddy DC, Singh SP, Tiwari IC, Shukla KP, Srivastava MK (1993) An epidemiological study of cannabis abuse among college students of Varanasi. *Indian Journal of Public Health*, **37**:10–15.

Regidor E, Gutierrez-Fisac JL, Calle ME, Navarro P, Dominguez V (2001) Trends in cigarette smoking in Spain by social class. *Preventive Medicine*, **33**: 241–248.

Ricci JA, Becker S (1996) Risk factors for wasting and stunting among children in Metro Cebu, Philippines. *American Journal of Clinical Nutrition*, **63**: 966–975.

Rimm EB, Klatsky A, Grobbee D, Stampfer MJ (1996) Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine, or spirits. *British Medical Journal*, **312**:731–736.

Robinson TN, Killen JD, Taylor CB et al. (1987) Perspectives on adolescent substance use. A defined population study. *Journal of the American Medical Association*, **258**:2072–2076.

Roseboom TJ, van der Meulen JH, Osmond C, Barker DJ, Ravelli AC, Bleker OP (2001) Adult survival after prenatal exposure to the Dutch famine 1944–45. *Paediatric and Perinatal Epidemiology*, **15**:220–225.

Ross CE (2000) Walking, exercising, and smoking: does neighborhood matter? *Social Science and Medicine*, **51**:265–274.

Rossouw JE, Jooste PL, Steenkamp HJ, Thompson ML, Jordaan PC, Swanepoel AS (1990) Socioeconomic status, risk factors and coronary heart disease. The CORIS baseline study. *South African Medical Journal*, **78**:82–85.

Ruidavets JB, Ducimetiere P, Arveiler D et al. (2002) Types of alcoholic beverages and blood lipids in a French population. *Journal of Epidemiology and Community Health*, **56**:24–28.

Sakamoto N, Wansorn S, Tontisirin K, Marui E (2001) A social epidemiologic study of obesity among preschool children in Thailand. *International Journal of Obesity and Related Metabolic Disorders*, **25**:389–394.

Saxe L, Kadushin C, Beveridge A et al. (2001) The visibility of illicit drugs: implications for community-based drug control strategies. *American Journal of Public Health*, **91**:1987–1994.

Schneider JS, Lee MH, Anderson DW, Zuck L, Lidsky TI (2001) Enriched environment during development is protective against lead-induced neurotoxicity. *Brain Research*, **896**:48–55.

Schrager L, Friedland G, Feiner C, Kahl P (1991) Demographic characteristics, drug use, and sexual behavior of I.V. drug users with AIDS in Bronx, New York. *Public Health Reports*, **106**:78–84.

Schrijvers CT, Stronks K, van de Mheen HD, Mackenbach JP (1999) Explaining educational differences in mortality: the role of behavioral and material factors. *American Journal of Public Health*, **89**:535–540.

Sen A (1999) *Development as freedom*. Alfred A. Knopf, New York.

Sharma HK (1996) Sociocultural perspective of substance use in India. *Substance Use and Misuse*, **31**:1689–1714.

Shen X, Rosen JF, Guo D, Wu S (1996) Childhood lead poisoning in China. *Science of the Total Environment*, **181**:101–109.

Shewry MC, Smith WC, Woodward M, Tunstall-Pedoe H (1992) Variation in coronary risk factors by social status: results from the Scottish Heart Health Study. *British Journal of General Practice*, **42**:406–410.

Siegrist J, Bernhardt R, Feng ZC, Schettler G (1990) Socioeconomic differences in cardiovascular risk factors in China. *International Journal of Epidemiology*, **19**:905–910.

Silva PA, Hughes P, Faed JM (1986) Blood lead levels in Dunedin 11 year old children. *New Zealand Medical Journal*, **99**:179–183.

Singh RB, Beegom R, Mehta AS et al. (1999) Social class, coronary risk factors and undernutrition, a double burden of diseases, in women during transition, in five Indian cities. *International Journal of Cardiology*, **69**:139–147.

Singh RB, Niaz MA, Ghosh S et al. (1995) Epidemiological study of coronary artery disease and its risk factors in an elderly urban population of north India. *Journal of the American College of Nutrition*, **14**:628–634.

Singh RB, Niaz MA, Thakur AS, Janus ED, Moshiri M (1998) Social class and coronary artery disease in a urban population of north India in the Indian Lifestyle and Heart Study. *International Journal of Cardiology*, **64**:195–203.

Singh RB, Sharma JP, Rastogi V et al. (1997a) Social class and coronary disease in rural population of north India. The Indian Social Class and Heart Survey. *European Heart Journal*, **18**:588–595.

Singh RB, Sharma JP, Rastogi V, Niaz MA, Singh NK (1997b) Prevalence and determinants of hypertension in the Indian Social Class and Heart Survey. *Journal of Human Hypertension*, **11**:51–56.

Smith M, Umenai T, Radford C (1998) Prevalence of smoking in Cambodia. *Journal of Epidemiology*, **8**:85–89.

Statistics New Zealand/Ministry of Health (1993) *A picture of health*. Statistics New Zealand and Ministry of Health, Wellington.

Sobal J, Stunkard AJ (1989) Socioeconomic status and obesity: a review of the literature. *Psychological Bulletin*, **105**:260–275.

Sorlie P, Backlund E, Keller J (1995) US mortality by economic, demographic, and social characteristics: the National Longitudinal Mortality Study. *American Journal of Public Health*, **85**:949–956.

Sorlie PD, Garcia-Palmieri MR (1990) Educational status and coronary heart disease in Puerto Rico: the Puerto Rico Heart Health Program. *International Journal of Epidemiology*, **19**:59–65.

Sowers M, Gonzalez Villalpando C, Stern MP, Fox C, Mitchell BD (1995) Relationships between physical activity, insulin levels and lipids in non-diabetic low income residents of Mexico City: The Mexico City Diabetes Study. *Archives of Medical Research*, **26**:133–140.

Stamler R, Shipley M, Elliott P, Dyer A, Sans S, Stamler J (1992) Higher blood pressure in adults with less education. Some explanations from INTERSALT. *Hypertension*, **19**:237–241.

Stellman SD, Resnicow K (1997) Tobacco smoking, cancer and social class. *IARC Scientific Publications*, **138**:229–250.

Sternfeld B, Ainsworth BE, Quesenberry CP (1999) Physical activity patterns in a diverse population of women. *Preventive Medicine*, **28**:313–323.

Taylor R, Cumming R, Woodward A, Black M (2001) Passive smoking and lung cancer: a cumulative meta-analysis. *Australian and New Zealand Journal of Public Health*, **25**:203–211.

Tong S, McMichael AJ, Baghurst PA (2000a) Interactions between environmental lead exposure and sociodemographic factors on cognitive development. *Archives of Environmental Health*, **55**:330–335.

Tong S, von Schirnding YE, Prapamontol T (2000b) Environmental lead exposure: a public health problem of global dimensions. *Bulletin of the World Health Organization*, **78**:1068–1077.

Torun B (2000) Physical activity patterns in Central America. In: *Obesity and poverty: a new public health challenge*. Pena M, Bacallao J, eds. Pan American Health Organization, Washington, DC.

Tyas SL, Pederson LL (1998) Psychosocial factors related to adolescent smoking: a critical review of the literature. *Tobacco Control*, **7**:409–420.

Ueshima H, Zhang XH, Choudhury SR (2000) Epidemiology of hypertension in China and Japan. *Journal of Human Hypertension*, **14**:765–769.

UNDP (2000) *Human development report 2000: human rights and human development*. United Nations Development Programme, Oxford University Press, New York.

USDHHS (2001) *The Surgeon General's call to action to prevent and decrease overweight and obesity*. U.S. Department of Health and Human Services. Public Health Service, Office of the Surgeon General, Rockville, MD.

Vartiainen E, Pekkanen J, Koskinen S, Jousilahti P, Salomaa V, Puska P (1998) Do changes in cardiovascular risk factors explain the increasing socioeconomic difference in mortality from ischaemic heart disease in Finland? *Journal of Epidemiology and Community Health*, **52**:416–419.

von Schirnding YE, Fuggle RF, Bradshaw D (1991) Factors associated with elevated blood lead levels in inner city Cape Town children. *South African Medical Journal*, **79**:454–456.

Wadowski SJ, Karp RJ, Murray-Bachmann R, Senft C (1994) Family history of coronary artery disease and cholesterol: screening children in a disadvantaged inner-city population. *Pediatrics*, **93**:109–113.

Wagstaff A (2000) Socioeconomic inequalities in child mortality: comparisons across nine developing countries. *Bulletin of the World Health Organization*, **78**:19–29.

Wagstaff A, Watanabe N (2001) *Socioeconomic inequalities in child malnutrition in the developing world*. World Bank, Washington, DC.

Wahid Saeed AA, al Shammary FJ, Khoja TA, Hashim TJ, Anokute CC, Khan SB (1996) Prevalence of hypertension and sociodemographic characteristics of adult hypertensives in Riyadh City, Saudi Arabia. *Journal of Human Hypertension*, **10**:583–587.

Wang Y (2001) Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. *International Journal of Epidemiology*, **30**:1129–1136.

Whitlock G, MacMahon S, Vander Hoorn S, Davis P, Jackson R, Norton R (1997) Socioeconomic distribution of smoking in a population of 10 529 New Zealanders. *New Zealand Medical Journal*, **110**:327–330.

Whitlock G, MacMahon S, Vander Hoorn S, Davis P, Jackson R, Norton R (1998) Association of environmental tobacco smoke exposure with socioeconomic status in a population of 7725 New Zealanders. *Tobacco Control*, **7**:276–280.

Woodward A, Newland H, Kinahoi M (1994) Smoking in the Kingdom of Tonga: report from a national survey. *Tobacco Control*, **3**:41–45.

World Bank (2001a) *World development report 2000/2001: attacking poverty,* Oxford University Press, New York.

World Bank (2001b) *World development indicators 2001*. CD-ROM. World Bank. Washington, DC.

Yen IH, Kaplan GA (1998) Poverty area residence and changes in physical activity level: Evidence from the Alameda County Study. *American Journal of Public Health*, **88**:1709–1712.

Yip R, Scanlon K, Trowbridge F (1993) Trends and patterns in height and weight status of low-income US children. *Critical Reviews in Food Science and Nutrition*, **33**:409–421.

Yu Z, Nissinen A, Vartiainen E, Hu G, Tian H, Guo Z (2002) Socioeconomic status and serum lipids: a cross-sectional study in a Chinese urban population. *Journal of Clinical Epidemiology*, **55**:143–149.

Yu Z, Nissinen A, Vartiainen E, Song G, Guo Z, Tian H (2000) Changes in cardiovascular risk factors in different socioeconomic groups: seven year trends in a Chinese urban population. *Journal of Epidemiology and Community Health*, **54**:692–696.

Zhu BP, Liu M, Shelton D, Liu S, Giovino GA (1996) Cigarette smoking and its risk factors among elementary school students in Beijing. *American Journal of Public Health*, **86**:368–75.

Appendix A: DHS asset scores

Table A.1 Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score

Subregion	*Country*	*Variable value*	*Electricity availability*	*Urban–rural status*	*Educational level*	*Floor material*
AFR-D	Benin					
		0	4 648	547	4 026	
		1	806	317	1 018	2 591
		2		955	425	17
		3		3 672	22	2 883
AFR-D	Burkina Faso					
		0	5 503	963	5 240	
		1	738	337	643	3 965
		2		351	535	
		3		4 794	27	2 480
AFR-D	Cameroon					
		0	2 309	1 379	1 329	
		1	2 854	340	1 992	2 223
		2		990	2 057	12
		3		2 792	123	3 266
AFR-D	Chad					
		0	7 081	1 355	5 365	
		1	364	467	1 620	6 879
		2		1 191	454	
		3		4 441	15	564
AFR-D	Comoros					
		0	2 011	214	1 635	
		1	1 031		788	1 382
		2		689	600	42
		3		2 147	27	1 626
AFR-D	Ghana					
		0	2 921	635	1 737	
		1	1 919	232	813	869
		2		718	2 188	6
		3		3 258	105	3 968
AFR-D	Guinea					
		0	5 248	1 351	5 361	
		1	1 442	284	721	3 361
		2		790	530	74
		3		4 328	141	3 318
AFR-D	Liberia					
		0	2 266		3 347	
		1	2 969	933	1 027	806
		2		1 011	808	4 199
		3		3 295	57	234
AFR-D	Madagascar					
		0	5 655	1 286	1 465	
		1	1 398	351	3 439	485
		2		739	1 972	4 861
		3		4 684	182	1 714

continued

Table A.1 Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (*continued*)

Subregion	*Country*	*Variable value*	*Electricity availability*	*Urban–rural status*	*Educational level*	*Floor material*
AFR-D	Mali					
		0	8679	1265	7773	
		1	976	1708	1218	7310
		2		536	685	5
		3		6195	28	2389
AFR-D	Niger					
		0	6573	1048	6066	
		1	963	1315	934	5721
		2			549	
		3		5214	28	1856
AFR-D	Nigeria					
		0	5033	661	3706	
		1	4682	1750	2692	3508
		2		807	2891	24
		3		6592	521	6278
AFR-D	Senegal					
		0	5724	1017	6020	
		1	2474	1083	1705	3430
		2		1122	795	
		3		5371	73	5163
AFR-D	Togo					
		0	6877	1417	4423	
		1	1671	763	2800	2074
		2		869	1302	9
		3		5520	44	6486
AFR-E	Burundi					
		0	2955		2924	
		1	1015	619	814	3450
		2		31	200	23
		3		3320	31	497
AFR-E	Central African Republic	0	5570	1207	3083	
		1	298		2039	4813
		2		1267	730	20
		3		3410	32	1051
AFR-E	Côte d'Ivoire					
		0	4448	1264	4909	
		1	3625	1296	2032	1619
		2		1292	1112	2
		3		4247	46	6478
AFR-E	Ethiopia					
		0	10811	2015	10586	
		1	3902	1507	2530	11759
		2		1021	2092	158
		3		10824	159	3450

Table A.1 Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (*continued*)

Subregion	*Country*	*Variable value*	*Electricity availability*	*Urban–rural status*	*Educational level*	*Floor material*
AFR-E	Kenya					
		0	6 830	884	1 010	
		1	981	341	4 719	5 134
		2		241	2 004	63
		3		6 415	148	2 684
AFR-E	Malawi					
		0	4 485	336	1 834	
		1	356	568	2 633	3 720
		2		412	369	
		3		3 533	13	1 129
AFR-E	Mozambique					
		0	7 522	2 507	3 434	
		1	1 166	510	4 844	5 795
		2		1 204	486	368
		3		4 558	15	2 616
AFR-E	Namibia					
		0	4 227	666	799	
		1	1 169		2 674	3 391
		2		1 225	1 859	92
		3		3 530	89	1 938
AFR-E	Rwanda					
		0	5 851	722	2 342	
		1	430	436	3 492	5 086
		2			693	
		3		5 393	24	1 465
AFR-E	United Republic of Tanzania	0	3 104	335	1 026	
		1	642	506	2 461	2 526
		2		577	540	3
		3		2 611	2	1 500
AFR-E	Zambia					
		0	6 418	943	1 168	
		1	1 544	1 219	4 833	4 757
		2		839	1 828	15
		3		5 020	191	3 249
AFR-E	Zimbabwe					
		0	3 894	435	437	
		1	2 011	966	2 518	1 959
		2		408	2 803	11
		3		4 098	149	3 937
AMR-B	Brazil					
		0	795	4 580	769	
		1	11 813	3 002	4 254	733
		2		2 672	6 839	859
		3		2 358	747	11 020

continued

Table A.1 Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (*continued*)

Subregion	*Country*	*Variable value*	*Electricity availability*	*Urban–rural status*	*Educational level*	*Floor material*
AMR-B	Colombia					
		0	417	3821	404	
		1	11154	3280	3804	928
		2		1698	5819	422
		3		2786	1558	10230
AMR-B	Dominican Republic	0	2871	1337	793	
		1	5529	2878	4336	864
		2		911	2356	
		3		3296	935	7558
AMR-B	Mexico					
		0	1081	2411	813	
		1	8228		4352	1500
		2		1391	3504	4046
		3		5508	639	3764
AMR-B	Paraguay					
		0	2751	1561	177	
		1	3075	359	3744	2044
		2		981	1591	232
		3		2926	313	3551
AMR-B	Trinidad and Tobago	0	244		32	
		1	3558	189	1725	471
		2		1501	1972	2058
		3		2116	76	1277
AMR-D	Bolivia					
		0	2976	4846	1072	
		1	8189	1769	4023	3383
		2		807	4434	1584
		3		3765	1658	6220
AMR-D	Ecuador					
		0	908		368	
		1	3805	1920	2238	524
		2		870	1673	2065
		3		1923	434	2124
AMR-D	Nicaragua					
		0	4293	1616	2436	
		1	9323	2961	5818	6233
		2		3267	4637	616
		3		5790	743	6785
AMR-D	Peru					
		0	9538	11320	2127	
		1	19373	4105	9620	12792
		2		3709	11387	2012
		3		9817	5817	13528

Table A.1 Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (*continued*)

Subregion	*Country*	*Variable value*	*Electricity availability*	*Urban–rural status*	*Educational level*	*Floor material*
EMR-B	Tunisia					
		0	1 035		2 372	
		1	3 149	1 218	1 302	10
		2		1 244	440	359
		3		1 722	70	3 815
EMR-D	Egypt					
		0	880	1 916	6 793	
		1	13 897	2 628	3 539	5 463
		2		1 735	3 629	
		3		8 500	818	9 316
EMR-D	Morocco					
		0	4 508	1 799	5 866	
		1	4 745	1 142	1 561	2 487
		2		1 609	1 629	34
		3		4 706	200	6 735
EMR-D	Pakistan					
		0	1 753		5 055	
		1	4 838	1 820	600	3 304
		2		1 564	842	478
		3		3 227	114	2 783
EMR-D	Sudan					
		0	3 933		3 425	
		1	1 926	1 503	1 543	5 493
		2		673	821	367
		3		3 684	71	
EMR-D	Yemen					
		0	2 692	431	5 124	
		1	3 304	463	669	2 998
		2		636	153	6
		3		4 480	59	3 006
EUR-B	Kyrgyzstan					
		0	8	893	3	
		1	3 833	477	10	150
		2		115	3 097	3 154
		3		2 363	738	544
EUR-B	Turkey					
		0	*	1 376	1 590	
		1		2 854	4 455	535
		2		1 472	2 005	1 469
		3		2 874	526	6 572
EUR-B	Uzbekistan					
		0	12	828	3	
		1	4 401	711	11	463
		2		767	3 815	3 376
		3		2 109	586	576

continued

Table A.1 Distribution of DHS respondents by subregion and country, by values of the variables comprising the DHS asset score (*continued*)

Subregion	*Country*	*Variable value*	*Electricity availability*	*Urban–rural status*	*Educational level*	*Floor material*
EUR-C	Kazakhstan					
		0	143	1 850	15	
		1	4 561	434	11	50
		2		643	3 706	2 949
		3		1 873	1 068	1 801
SEAR-B	Indonesia					
		0	8 086	2 497	3 866	
		1	20 724	2 759	15 285	4 966
		2		3 416	8 646	7 748
		3		20 138	1 013	16 096
SEAR-B	Sri Lanka					
		0	4 495	*	734	
		1	1 368		1 777	2 457
		2			2 062	23
		3			1 289	3 385
SEAR-B	Thailand					
		0	1 265		599	
		1	5 509	1 248	4 984	312
		2		1 175	777	6 057
		3		4 352	415	406
SEAR-D	Bangladesh					
		0	6 795	462	4 899	
		1	2 323	65	2 530	7 887
		2		922	1 403	42
		3		7 678	295	1 198
SEAR-D	India					
		0	36 333	7 825	50 823	
		1	53 444	8 581	15 546	18 196
		2		11 128	19 281	26 507
		3		62 243	3 855	45 074
SEAR-D	Nepal					
		0	6 855	203	6 736	
		1	1 566	751	893	7 575
		2		7 475	688	193
		3			112	661
WPR-B	Philippines					
		0	3 981	954	366	
		1	9 705	4 037	4 010	1 022
		2		1 739	5 718	5 107
		3		7 253	3 889	7 854

* For Turkey and Sri Lanka, data were missing for both electricity and rural/urban status and there were no suitable substitute variables. (See the Methods section for more details.) These two countries were not included in the final asset scores.

Table A.2 Numbers of DHS observations by DHS asset score, by subregion

	Subregion										
Asset score	*AFR-D*	*AFR-E*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-B*
–1.11788	30876	22999	609	3117	6	8165			961	11202	14
–0.9019455	1795	1446	58	198		396			13	6376	2
–0.8412057	4676	19249	2496	6912		888			1798	2586	275
–0.8398404	2209	194	68	141	161	101		1	838	8931	220
–0.6860113	1451	379	15	70		281			7	503	1
–0.6252716	645	1653	165	478		105			28	764	11
–0.6239063	337	48	14	6	8	8			24	653	24
–0.5645318	999	3029	208	1479		198	15	18	324	818	158
–0.5631665	2497	188	410	746	13	19			2726	1275	1333
–0.5618012	8206	2169	222	360	398	2338			335	9518	14
–0.4700771	833	544	6	47		16				266	
–0.4305602	508	525	253	405	3	5015	1		338	3462	1
–0.4093374	491	735	45	248		100			12	103	5
–0.4079721	121	1		2	1	22			6	200	11
–0.3485976	184	370	31	212		30			15	331	14
–0.3472323	270	37	52	33	6	2			79	159	125
–0.345867	1362	448	36	58	133	161			8	491	1
–0.2878579	16	45	5	155		2		2	5	45	20
–0.2864926	562	26	119	227	1	8	2	98	394	549	727
–0.2851273	3196	3447	1526	998	75	280			1231	2233	188
–0.214626	93	166	93	190		870			20	1165	1
–0.1934032	410	1003	25	206		6			1	82	
–0.1920379	8	6	2	4						93	
–0.1538863	324	248	1035	1449		1310	2		1221	1001	111
–0.152521	805	10	218	67	127	72	2	3	672	5242	42
–0.1326635	165	99	3	161		37			1	50	13
–0.1312982	79	5	1	10		3			26	72	97
–0.1299329	1235	354	39	17	85	61			5	144	3
–0.0719237		2	1	20		1				14	4
–0.0705584	127	2	19	18		2		5	11	82	90
–0.0691931	737	870	280	125	68	32			47	209	21
–0.0098187	3			29				3	10	5	114
–0.0084534	1989	1886	324	404		77	1	6	354	1504	208
0.0013082	153	204	24	166		560			18	610	
0.0620479	73	137	330	719		267			103	379	33
0.0634132	273	2	77	35	22	61	2	1	72	974	2
0.0832707	199	210	2	244		1				48	
0.084636	37	13	6	35						30	2
0.0860013	1517	537	15	11		157			1	41	
0.1227876	165	39	306	723		454	541	27	285	708	161
0.1241529	290	8	2030	468	11	22	9	6	4584	1608	343
0.1255182	1065	515	437	199	681	3111		1	751	7685	9
0.1440104	3	1		24		3				1	2

continued

Table A.2 Numbers of DHS observations by DHS asset score, by subregion (*continued*)

	Subregion										
Asset score	*AFR-D*	*AFR-E*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-B*
0.1453757	54			21		6	2	3	5	33	82
0.146741	632	814	419	49	52	25			29	76	14
0.2061155				7						1	14
0.2074808	506	488	126	86	12	27			20	148	15
0.2172423	75	117	18	165		58			1	958	
0.2682205	57	98	42	68		3			10	53	93
0.2779821	48	164	129	791	1	362			36	281	38
0.2793474	266	2	8	76	4	208	1		98	613	6
0.3387218	66	136	303	713		155	28	1	38	376	32
0.3400871	151	8	765	266	3	27	3		785	476	67
0.3414524	1 018	568	245	179	423	1 819			113	1 323	3
0.3599446	2	5		26							
0.3613099	30	4	4	32				5		25	4
0.3626752	942	1 659	181	71		109			5	32	2
0.3994615	8	3	8	95		10	18		11	54	55
0.4008268	148	2	1 403	275		12	3 421	1 277	1 157	1 329	524
0.4021921	1 165	638	3 871	1 295	222	1 206	2		3 867	3 953	461
0.4220496				3						3	13
0.4234149	398	402	286	60	15	10		2	15	63	24
0.4841547	14	12	19	13	1	1				9	12
0.4939162	96	162	149	991		17	1		12	409	3
0.4952815	3	25	42	20		1		6	6	603	
0.554656	73	163	101	825		351	5		21	353	34
0.5560213	161	4	113	443	1	129	1		913	410	143
0.5573866	1 085	770	245	169	320	1 593		1	84	750	6
0.6153957	4	7	18	150		9	1		3	53	13
0.616761	213	16	740	488	1	27	713	351	453	762	160
0.6181263	987	751	2 155	1 250	407	824			1 079	1 207	146
0.6379838	1			6						2	
0.6393491	615	610	130	86		65			6	17	
0.6775007	1		87	69		1	345	182	151	77	193
0.678866	1 594	555	2 834	981	20	1 137	88	151	2 247	5 373	1 012
0.7000889	14	10	86	7		1			1	6	11
0.7705901	114	145	82	1 526		6	1	1	4	494	1
0.7719554	47	41	467	222		2	2	1	50	357	22
0.7733208	1 492	872	239	248		1 132		2	94	644	4
0.8313299	1	3	7	157		49		1	1	51	16
0.8326952	282	9	221	746		136	804	161	460	842	332
0.8340605	1 174	1 297	2 379	1 511	443	1 216	1		694	752	391
0.8934349	18	1	39	169		2	91	51	145	131	79
0.8948002	1 529	1 109	3 148	2 336	139	1 193	34	185	1 295	2 624	401

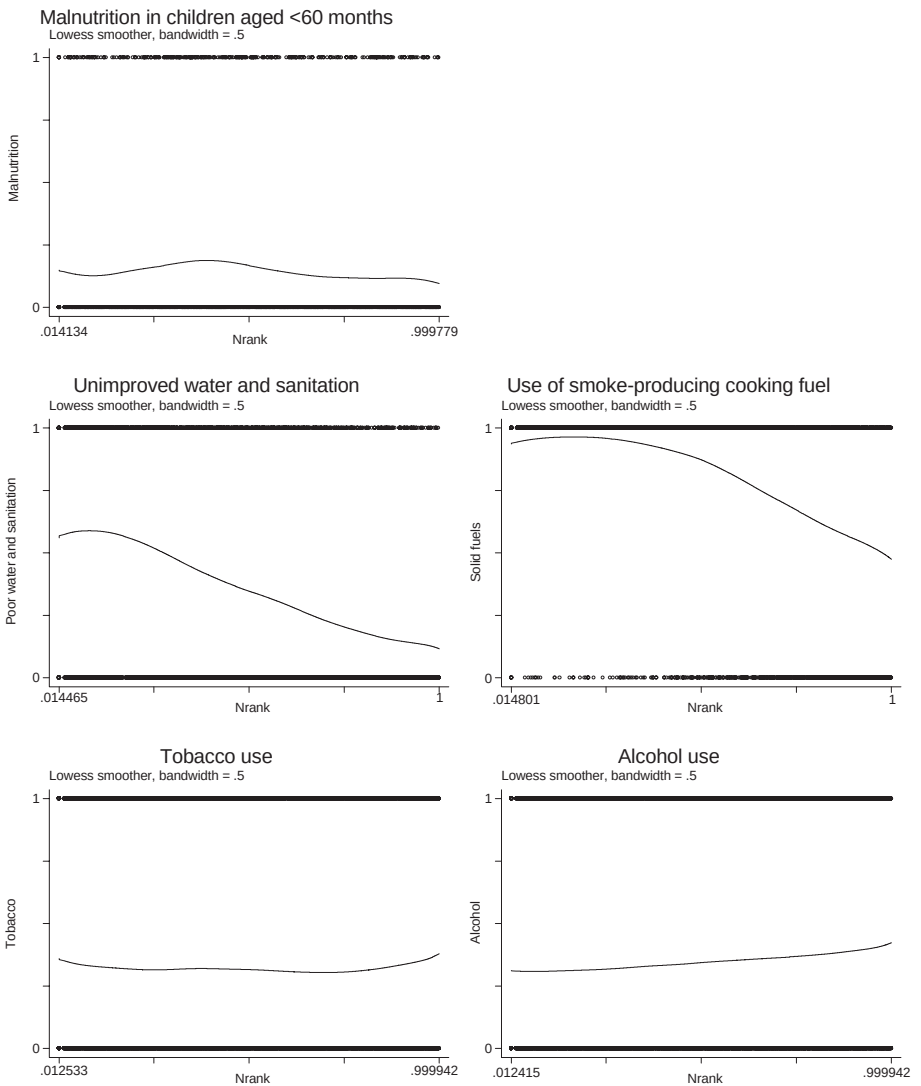
Table A.2 Numbers of DHS observations by DHS asset score, by subregion (*continued*)

	Subregion										
Asset score	*AFR-D*	*AFR-E*	*AMR-B*	*AMR-D*	*EMR-B*	*EMR-D*	*EUR-B*	*EUR-C*	*SEAR-B*	*SEAR-D*	*WPR-B*
0.916023	25	16	61	28		5					
0.9555399	163	63	455	231	4	111	19	32	220	513	812
1.047264	5	4	5	261		4				38	3
1.048629	173	86	597	744		4	670	566	43	736	69
1.049995	1983	1914	3166	2044		954		3	942	512	104
1.109369	28	2	32	304		10	178	52	94	195	226
1.110734	1651	2216	4359	3300	252	1899	141	154	1157	2286	1198
1.171474	177	94	436	845	21	189	9	41	240	775	437
1.325303	49	15	81	564			284	177	3	187	37
1.326669	3716	2824	6718	6240		1235	438	628	1118	1819	331
1.387408	219	176	1182	1550	44	487	53	56	323	974	1293
1.603342	554	323	1698	3793		374	325	443	211	1074	350
Total	94811	83252	51504	57811	4184	42443	8254	4704	35584	107044	13686

Appendix B: Loess plots for China data

Locally linear kernel regression smooth plots of the various risk factors according to normalized equivalized income ranking are shown in Figure B.1. Each sub-figure plots the proportion of individuals or households with the risk factor in question (y-axis ranging from 0 to 1) by normalized equivalized income rank for China (x-axis ranging from 0 [poorest] to 1 [richest in China]).

Figure B.1 Loess plots of the prevalence of various risk factors (y-axis) by normalized income rank (x-axis)



Appendix C: Literature review summaries

Underweight

A total of 17 journal articles on malnutrition were summarized. These studies used several measures of nutritional status, including underweight, wasting and stunting. Many studies did not report results for underweight, but the pattern of the results was comparable across the three different measures and between studies. There were negative associations between malnutrition and income, maternal education, urban residence, availability of water or sanitation, and household goods such as a radio or television.

Wagstaff and Watanabe (2001) reviewed data on malnutrition in relation to equivalized household consumption for 20 developing countries. They found that malnutrition tended to decline monotonically with rising living standards. Inequalities in underweight tended to be larger than for wasting and stunting. Between-country differences in underweight were not statistically significant.

Author/year	Anonymous (2000)
Subregion (country)	AMR-D (Bolivia)
Study type/design	Cross-sectional survey (DHS)/random sampling
Study population	Nationally representative sample of women aged 15–49 years (n = 11 187), their children aged 0–5 years and men aged 15–64 years (n = 3780)
Survey year(s)	1998
Risk factor measure(s)	Weight-for-height: ≤–2 SD below median of CDC/WHO reference population defined as moderate or severely wasted, ≤–3 SD below median defined as severely wasted Height-for-age: ≤–2 SD below median of CDC/WHO reference population defined as moderate or severely stunted, ≤–3 SD below median defined as severely stunted
Socioeconomic measure(s)	Area of residence (rural/urban) Mother's education level: a = none, b = basic, c = intermediate, d = secondary or higher
Measure(s) of association	Prevalence of malnutrition by socioeconomic measure
Bivariate summary of results (including controlling for sex and age)	*Weight-for-height* MODERATE/SEVERE WASTING (Total = 1.8%): rural = 2.4, urban = 1.3 Educational level: a = 2.9, b = 1.9, c = 2.1, d = 1.1 SEVERE WASTING (Total = 0.5%): rural = 0.6, urban = 0.4 Educational level: a = 0.3, b = 0.5, c = 1.1, d = 0.2 *Height-for-age* MODERATE/SEVERE STUNTING (Total = 25.6%): rural = 35.6, urban = 18.3 Educational level: a = 44.3, b = 33.6, c = 19.1, d = 12.6 SEVERE STUNTING (Total = 8.9%): rural = 14.0, urban = 5.1 Educational level: a = 19.1, b = 12.1, c = 6.3, d = 2.9
Multivariate summary of results (including controlling for rurality, ethnicity, etc.)	Not applicable
Comments	Weight-for-age (underweight) results not provided No further definitions of area of residence or mother's education provided

Author/year	Anonymous (1998)
Subregion (country)	AMR-B (Dominican Republic)
Study type/design	Cross-sectional survey (DHS)/random sampling
Study population	Nationally representative sample of women aged 15–49 years (n = 8422), their children aged 0–5 years and men aged 15–64 years (n = 2279)
Survey year(s)	1996
Risk factor measure(s)	Weight-for-height: ≤–2 SD below median of CDC/WHO reference population defined as moderate or severely wasted, ≤–3 SD below median defined as severely wasted. Height-for-age: ≤–2 SD below median of CDC/WHO reference population defined as moderate or severely stunted, ≤–3 SD below median defined as severely stunted
Socioeconomic measure(s)	Area of residence (rural/urban) Mother's educational level: a = none, b = primary (1–4 years), c = primary (5–8 years), d = secondary, e = superior
Measure(s) of association	Prevalence of malnutrition by socioeconomic measure
Bivariate summary of results	*Weight-for-height* SEVERE WASTING (Total = 0.2%): rural = 0.2, urban = — Educational level: a = 0.4, b = 0.3, c = 0.3, d = 0.0, e = 0.0 MODERATE/SEVERE WASTING (Total = 1.2%): rural = 1.2, urban = — Educational level: a = 1.7, b = 1.9, c = 1.0, d = 1.2, e = 0.4 *Height-for-age* SEVERE STUNTING (Total = 2.8%): rural = 4.4, urban = 1.6 Educational level: a = 7.7, b = 5.5, c = 2.4, d = 0.8, e = 0.0 MODERATE/SEVERE STUNTING (Total = 10.7%):rural = 15.2, urban = 7.3 Educational level: a = 23.1, b = 16.3, c = 11.1, d = 5.6, e = 1.6
Multivariate summary of results	Not applicable
Comments	No data on underweight (weight-for-age). No further definitions of area of residence or mother's education provided

Author/year	Anonymous (1992)
Subregion (country)	EMR-D (Pakistan)
Study type/design	Cross-sectional survey (DHS)/random sampling
Study population	Nationally representative sample of women aged 15–49 years (n = 6611), their children aged 0–5 years and subset of their husbands (n = 1354)
Survey year(s)	1990/1991
Risk factor measure(s)	Weight-for-age (underweight), height-for-age (stunting) and weight-for-height (wasting) below –2 SD and –3 SD
Socioeconomic measure(s)	Rural/urban Mother's educational level: a = no education, b = primary, c = middle, d = secondary or higher
Measure(s) of association	Prevalence of underweight, stunting and wasting below –2 SD and –3 SD by socioeconomic measure
Bivariate summary of results	BELOW –2 SD *Weight-for-age (underweight)* (Total = 40.4%): rural = 44.6, urban = 32.5 Educational level: a = 44.9, b = 37.1, c = 25.8, d = 13.0 *Height-for-age (stunting)* (Total = 50.0%): rural = 54.9, urban = 40.7 Educational level: a = 55.5, b = 43.8, c = 33.2, d = 18.2 *Weight-for-height (wasting)* (Total = 9.2%): rural = 9.8, urban = 8.1 Educational level: a = 10.3, b = 7.5, c = 5.3, d = 3.6 A similar pattern was observed for –3 SD
Multivariate summary of results	Not applicable
Comments	No detailed definition provided for residence, maternal education and measures of malnutrition

Author/year	Anonymous (1991)
Subregion (country)	AFR-E (Zimbabwe)
Study type/design	Cross-sectional survey (DHS)/random sampling
Study population	Nationally representative sample of women aged 15–49 years (n = 4201) and their children aged 3–60 months
Survey year(s)	1988
Risk factor measure(s)	Weight-for-height: ≤–2 SD below median of CDC/WHO reference population defined as moderate or severely wasted, –1 to –1.99 SD below median defined as mildly wasted Height-for-age: ≤–2 SD below median of CDC/WHO reference population defined as moderate or severely stunted, –1 to –1.99 SD below median defined as mildly stunted
Socioeconomic measure(s)	Area of residence (rural/urban) Mother's educational level: a = none, b = primary, c = secondary/higher
Measure(s) of association	Prevalence of malnutrition by socioeconomic measure
Bivariate summary of results	*Weight-for-height* SEVERE/MODERATE WASTING (Total = 1.3%): rural = 1.3, urban = 1.5 Educational level: a = 1.9, b = 1.4, c = 0.5 MILD WASTING (Total = 9.6%): rural = 10.1, urban = 8.1 Educational level: a = 12.6, b = 9.0, c = 8.7 *Height-for-age* SEVERE/MODERATE STUNTING (Total = 29.0%): rural = 33.6, urban = 14.3 Educational level: a = 37.0, b = 30.4, c = 15.6 MILD STUNTING (Total = 35.5%): rural = 36.5, urban = 32.0 Educational level: a = 35.6, b = 36.2, c = 32.7
Multivariate summary of results	Not applicable
Comments	No data on underweight (weight-for-age)

Author/year	Chen (1996)
Subregion (country)	WPR-B (China)
Study type/design	Meta-analysis using survey data from 1987 Child Survey (9 provinces), 1992 National Child Survey, 1990 Nutrition Surveillance—State Statistic Bureau (7 provinces including Beijing), 1992 Third National Nutritional Survey (all provinces), 1990 nutrition surveillance data
Study population	Nationally representative samples
Survey year(s)	1987, 1992, 1990, 1992, 1990, respectively
Risk factor measure(s)	Prevalence of underweight and stunting (neither is defined)
Socioeconomic measure(s)	Area of residence (urban, rural) Household income (groups 1–6 based on percentile groupings or 100 yuan/person per year increase); Safe drinking-water (percentage piped drinking-water coverage) Mother or father illiterate Poverty (no definition provided) *Results also given according to geographical area:* I = Beijing, Tianjin, Liaoning II = Jilin, Heilongjiang, Jiangsu, Shandong III = Shanghai, Zejiang, Guangdong IV = Hebei, Inner Mongolia, Shanxi, Anhui, Henan V = Fujian, Jiangxi, Hubei, Hunan VI = Hainan, Guizhou, Shaanxi, Gansu, Qinghai, Ningxia VII = Guangxi, Sichuan, Yunnan, Xinjiang
Measure(s) of association	Prevalence of malnutrition by area of residence Attributable risk of factors relating to stunting in rural China
Bivariate summary of results	The national prevalence of moderate underweight in children aged 0–6 years was 21.3% in 1987 and 17.9% in 1992 (figures for urban and rural areas are provided by province only—on average rural prevalences are markedly higher than urban) In 1992 for children aged 0–5 years, national prevalence of underweight (moderate plus severe) was 17.9% and that of stunting was 34.0% Prevalence of stunting in urban areas in this group was 11.4% for boys and 11.5% for girls, and in rural areas 39.3% and 40.5%, respectively
Multivariate summary of results	*Attributable risk of factors related to stunting in rural China by geographical area:* Poverty: IV = 5.7, V = 50.0, VI = 6.5, VII = 32.4 Safe drinking-water: II = 23.7, III = 35.1, IV = 15.3, V = 24.2, VII = 26.5 Illiterate mother or father: I = 31.0, II = 20.0, IV = 9.9, V = 26.5, VI = 18.7 Income (100 yuan/person per year increase): I = –1, II = –2, III = –2, IV = –3, V = –1, VI = –3, VII = –1
Comments	Difficult to interpret, broken up into regions, no raw data Few definitions given (none of stunting, underweight or poverty)

Author/year	Delpeuch et al. (2000)
Subregion (country)	AFR-E (Congo)
Study type/design	Cross-sectional survey/cluster sampling
Study population	Urban (Brazzaville) children aged 0–5 years (n = 2373)
Survey year(s)	1986, 1991
Risk factor measure(s)	Low weight-for-height: –1 SD below mean weight-for-height of CDC/WHO reference population (prevalence of wasting, defined as weight-for-height <–2 SD, was only 4.2%) Stunting: –2 SD below mean height-for-age of CDC/WHO reference population BMI (kg/m^2): thin <18.5 kg/m^2, overweight ≥25 kg/m^2
Socioeconomic measure(s)	Mother's occupation Income Mother's educational level Economic level of household (a = low, b = medium, c = high) Dwelling district
Measure(s) of association	Odds ratio Prevalence of malnutrition by socioeconomic measure
Bivariate summary of results	STUNTING Total prevalence 11.0% *Adjusted for child's age, sex, and mother's age* Economic level of household: a: odds ratio (OR) = 2.0 (95% CI 1.1–3.6); b: OR = 1.7 (95% CI 1.0–2.9); c: OR = 1.0 (P = 0.048) Maternal education: none: OR = 2.9 (95% CI 1.2–7.2); primary: OR = 2.9 (95% CI 1.3–6.7); secondary: OR = 1.4 (95% CI 0.6–2.9); higher: OR = 1.0 (P = 0.004) Dwelling district: peripheral: OR = 3.8 (95% CI 1.8–7.8); intermediate: OR = 2.1 (95% CI 1.1–4.2); central: OR = 1.0 (P = 0.0005)
Multivariate summary of results	STUNTING *Adjusted for child's age, sex, mother's age, birth weight, mother's height and BMI* Economic level of household: a: OR = 1.9 (95% CI 1.0–3.5); b: OR = 1.5 (95% CI 0.8–2.6); c: OR = 1.0 (P = 0.11) Maternal education: none: OR = 2.1 (95% CI 0.8–5.3); primary: OR = 2.6 (95% CI 1.1–6.1); secondary: OR = 1.3 (95% CI 0.6–2.8); higher: OR = 1.0 (P = 0.047) Dwelling district: peripheral: OR = 3.0 (95% CI 1.4–6.4); intermediate: OR = 1.6 (95% CI 0.8–3.3); central: OR = 1.0 (P = 0.005)
Comments	Increasing prevalence of stunting in children with increasing distance from town centre, decreasing maternal education, decreasing economic level of household. The prevalence of low weight-for-height increases when the mother is not salaried (results not listed) Congo is atypical as poverty is a more recent phenomenon

Author/year Kikafunda et al. (1998)

Subregion (country) AFR-E (Uganda)

Study type/design Cross-sectional survey/stratified multistage random sampling

Study population Children aged 0–30 months in Mubende district, central Uganda, in area with dominant use of starchy staples (potatoes, maize, green bananas, etc.) (n = 261): 183 (70%) from rural areas, 78 (30%) from semi-urban areas

Survey year(s) 1997

Risk factor measure(s) Underweight: –2 SD below median weight-for-age and sex of CDC/WHO reference population
Stunting: –2 SD below median height-for-age of CDC/WHO reference population
Wasting: –2 SD below mean weight-for-height of CDC/WHO reference population
Mid-upper arm circumference (low <135 mm)

Socioeconomic measure(s) Area of residence (R = rural, U = urban)
Economic status
Type of cooking fuel (firewood, charcoal or paraffin)
Cleanliness of household
Maternal education
Water source

Measure(s) of association Prevalence of stunting, underweight by socioeconomic measure; odds ratio

Bivariate summary of results

Total prevalences
Underweight: 63/261 (24.1%)
Stunting: 62/261 (23.8%)
Wasting: 2/261 (0.8%)

Prevalence of underweight by socioeconomic measure

Variable	*Underweight (%)*	*P value*
Area of residence	R = 27.9, U = 15.4	0.31
Economic status	Mid-upper = 13.2	0.108
	Lower = 24.3	
	Very low = 33.3	
Firewood only fuel	Yes = 27.6, No = 13.9	0.025
Charcoal	Yes = 27.6, No = 13.9	0.029
Paraffin	Yes = 14.7, No = 27.5	0.018
Cleanliness	Clean = 9.1	0.037
	Moderately clean = 24.0	
	Dirty/very dirty = 32.8	
Maternal education	None = 29.4	0.885
	Primary (incomplete) = 23.2	
	Primary (complete) = 26.9	
	Secondary and above = 22.4	

Prevalence of stunting by socioeconomic measure

Variable	*Stunting (%)*	*P value*
Area of residence	R = 23.5, U = 24.4	0.881
Economic status	Mid-upper = 7.9	0.034
	Lower = 25.4	
	Very low = 31.0	

Firewood only fuel	Yes = 25.0, No = 20.0	0.412
Charcoal main fuel	Yes = 20.8, No = 25.6	0.372
Paraffin main fuel	Yes = 13.2, No = 27.5	0.018
Cleanliness	Clean = 15.2	0.290
	Moderately clean = 23.4	
	Dirty/very dirty = 29.5	
Maternal education	None = 29.4	0.885
	Primary (incomplete) = 24.0	
	Primary (complete) = 25.0	
	Secondary and above = 20.9	

When age of child was controlled for, however, better educated mothers had less stunted children ($P = 0.045$)

Multivariate summary of results

Underweight by economic status of family by odds ratio

Economic status	*Odds ratio (95% CI)*
Mid-upper (reference)	1.00
Lower	2.57 (0.68–9.80)
Low/very low	2.62 (0.63–11.03)

Comments

Not documented how economic status measured
Positive association between maternal education and stunting has been shown in other studies in Uganda; only positive association in this study once age is controlled for
No table showing number of mothers in each educational group
Proportion underweight in this study (24%) similar to results from 1988/1989 Uganda DHS (23%)
No record of how many children had more than one type of malnutrition, although stated that both stunted and wasted children showed greater incidence of underweight ($P \leq 0.0001$ and $P = 0.012$, respectively)
Neither age, sex nor birth order had any effect on incidence of underweight, although there was higher incidence of stunting in older than younger children ($P \leq 0.0001$)

Author/year	Li et al. (1999)
Subregion (country)	WPR-B (China)
Study type/design	Cross-sectional survey/random sampling
Study population	Children aged 0–7 years from four poor rural counties in Yunnan Province (n = 2019). Sample included four minority ethnic groups (Hani, Yi, Hui, Miao) and representatives of the dominant ethnic group in the area (Han)
Survey year(s)	Not recorded
Risk factor measure(s)	Prevalence of stunting (height-for-age), underweight (weight-for-age) and wasting (weight-for-height) from –2 to –2.99 SD (moderate) or ≤–3 (severe) SD below median of NCHS/WHO reference population
Socioeconomic measure(s)	Family income per capita in previous year: low = <200 yuan; medium = 200–499 yuan; high = ≥500 yuan Ethnicity: Hani, Yi, Hui, Miao and Han (dominant in area) Drinking-water: tap water, pump water, well or spring water, rain or snow water, lake, river or pool water Lavatory: public toilet, indoor flush toilet, private pit, no toilet
Measure(s) of association	Odds ratio Prevalence of malnutrition by socioeconomic measure
Bivariate summary of results	*Overall prevalences* Moderate and severe underweight: 18.9% The prevalence of being underweight increased with age, peaking at 12–17 months then decreased steadily; boys were more likely to be underweight than girls (χ^2 = 15.58, P <0.001); most underweight children (87.1%) were also stunted Moderate and severe stunting: 51.0% The prevalence of being underweight increased with increasing age, peaking at 3 years but maintained fairly high prevalence; boys were more likely to be stunted than girls (χ^2 = 7.36, P <0.01) Moderate and severe wasting: 1.4% *Moderate and severe stunting adjusted for sex and age* Family income: low: OR = 5.2 (2.8 – 9.9); medium: OR = 2.2 (1.3–3.6); high: OR = 1 (reference) Drinking-water: well and spring: OR = 1.5 (0.9 – 2.5); lake, river, pool: OR = 4.5 (2.6 – 7.9); tap: OR = 1 (reference) Lavatory: private pit: OR = 2.3 (1.0–5.3); no toilet: OR = 3.0 (1.2–7.4); public toilet: OR = 1 (reference) Certain ethnic groups (in particular Miao) had high odds ratios for moderate/severe stunting compared with Han
Multivariate summary of results	*Moderate and severe stunting adjusted for sex and age:* Family income: low: OR = 4.1 (1.9–8.5); medium: OR = 1.2 (0.7–2.3); high: OR = 1.0 (reference) Drinking water: well and spring: OR = 1.2 (0.7–2.2); lake, river, pool: OR = 2.3 (1.0–5.1); tap: OR = 1 (reference)
Comments	Most of the analysis was performed on stunting only

Author/year	Madzingira (1995)
Subregion (country)	AFR-E (Zimbabwe)
Study type/design	Cross-sectional survey (DHS)/random sampling
Study population	Nationally representative sample of women aged 15–49 years (n = 4201) and children aged 0–5 years
Survey year(s)	1988
Risk factor measure(s)	Height-for-age, weight-for-age, weight-for-height
Socioeconomic measure(s)	Area of residence (urban/rural) Maternal education
Measure(s) of association	Odds ratio Prevalence of malnutrition by area of residence
Bivariate summary of results	*Malnutrition (undefined) in children aged 0–5 years:* rural = 40%, urban = 25%
Multivariate summary of results	*Weight-for-age* OR = 2.0 for children with mothers with no education compared with those with mothers with secondary or higher education Correlation between rural residence and malnutrition (OR = 2.5 compared with urban children) *Height-for-age* Correlation between rural residence and malnutrition (OR = 3.1 compared with urban children) *Weight-for-height* No significant findings
Comments	Maternal education is related to child's nutritional status Urban children are less likely to be malnourished owing to better living conditions and standards compared to their rural counterparts

Author/year	Matulessy et al. (1992)
Subregion (country)	SEAR-B (Jakarta, Indonesia)
Study type/design	Analysis of survey data/nationally representative routinely collected data by Central Bureau of Statistics of Indonesia and Directorate of Nutrition (1986); also National Survey of Vitamin A (1978) and other nationally collected data
Study population	Nationally representative sample as per protocols for above surveys
Survey year(s)	1986, 1978
Risk factor measure(s)	Prevalence of underweight by weight-for-age: Well nourished (good) = ≥80% weight-for-age Mild malnutrition = 70–79% weight-for-age Moderate malnutrition = 60–69% weight-for-age Severe malnutrition = <60% weight-for-age
Socioeconomic measure(s)	Area of residence (urban/rural) Maternal and paternal education
Measure(s) of association	Prevalence of malnutrition by socioeconomic measure
Bivariate summary of results	*Prevalence of moderate and severe* malnutrition *in children aged <5 years by zone in 1986 (for West Zone, n = 18 835, for East Zone, n = 4110):* Urban: West Zone: severe = 1.5, moderate + severe = 19.6 East Zone severe = 2.9, moderate + severe = 28.4 Rural: West Zone: severe = 0.7, moderate + severe = 8.1 East Zone severe = 1.3, moderate + severe = 9.7 East Zone includes Nusatenggara, Kalimantan, Maluku and Irian Jaya Also gives prevalence of moderate + severe malnutrition by maternal and paternal education by urban and rural by area. General trend is that the higher the parental education, the lower the prevalence of malnutrition
Multivariate summary of results	Not applicable
Comments	Classification of malnutrition—modification of Harvard standard

Author/year	Monteiro et al. (1992)
Subregion (country)	AMR-B (Brazil)
Study type/design	Analysis of two national nutrition surveys (both multistage cluster stratified randomized sampling): Estudo Nacional da Despesa Familiar (1975) and Pesquisa Nacional sobre Saude e Nutricao (1989)
Study population	Children aged 0–5 years for whom weight, age and sex information was provided (n = 36 407 in 1975, n = 3571 in 1989)
Survey year(s)	1975, 1989 (only results for 1989 reported here)
Risk factor measure(s)	Prevalence of malnutrition (weight-for-age indices <–2 z-scores below reference average weight expected for age and sex)
Socioeconomic measure(s)	Area of residence Quartile of monthly per capita family income
Measure(s) of association	Prevalence of malnutrition by socioeconomic measure in 1989

Bivariate summary of results

Measure	*Prevalence*
AREA OF RESIDENCE	
Urban	5.6%
Rural	10.6%
INCOME QUARTILE	
Q1 (lowest)	13.6%
Q2	9.5%
Q3	4.8%
Q4	1.4%

Multivariate summary of results	Not applicable
Comments	Nationally representative. Over time (1975–1989) the prevalence of weight-for-age malnutrition decreased, although this decrease was lowest in most vulnerable groups, i.e. those in the lowest income quartile Also geographical pattern of malnutrition

Author/year	Monteiro et al. (1995)
Subregion (country)	AMR-B (Brazil)
Study type/design	Cross-sectional anthropometric surveys/random sampling
Study population	Nationally representative sample, excluding pregnant women (in 1989: 5969 children and 23 544 adults)
Survey year(s)	1974–1975, 1989 (only results for 1989 reported here)
Risk factor measure(s)	Weight-for-height indices compared with CDC/WHO reference population used to measure children; Weight-for-age indices compared with CDC/WHO reference population used to measure children: <–2 z-scores classified as underweight; >2 z-scores classified as overweight BMI (underweight <18.5 kg/m2, obese >30.0 kg/m^2) used to classify adults
Socioeconomic measure(s)	Per capita family income (obtained through the standardized questionnaire of Instituto Brasiliero de Geografia e Estatística): poorest = lowest 30% of sample (approx. <US$ 31 per month); middle = next 40% (approx. US$ 30–91 per month); highest = top 30% (approx. >US$ 90)
Measure(s) of association	Prevalence of underweight (weight-for-age < –2 z-scores for children) by socioeconomic measure
Bivariate summary of results	Total prevalence of underweight in children (aged 1–4 years in 1989 = 7.6% *Prevalence of underweight in children (aged 1–4 years) by socioeconomic measure, 1974* Poorest 26.5% (SE 0.49) Intermediate 11.6% (SE 0.49) Richest 3.9% (SE 0.44) *Prevalence of underweight in children (aged 1–4 years) by socioeconomic measure, 1989* Poorest 12.2% (SE 0.83) Intermediate 3.8% (SE 0.63) Richest 1.4% (SE 0.59) Also results for adults by sex and BMI (sex difference among adults)
Multivariate summary of results	Not applicable
Comments	Decreasing prevalence of underweight children over time, but still a problem among the poorest

Author/year	Nube et al. (1998)
Subregion (country)	AFR-D (Ghana)
Study type/design	Analysis of household survey data (LSMS)
Study population	Adults aged 20–65 years and their children aged <5 years, excluding individuals from households with expenditure >1 million cedis/year, pregnant and lactating women and those with BMI <10 or >40 (n = 4228: 2114 males and 2114 females)
Survey year(s)	1988–1989
Risk factor measure(s)	Height-for-age z-scores (number of SDs below mean of reference data [children aged <5 years only])
Socioeconomic measure(s)	Area/type of residence (residence code): Rursavfarm = rural savannah farm household Rurforfarm = rural forest farm household Rurcoafarm = rural coastal farm household Rursavnonf = rural savannah non-farm household Rurfornonf = rural forest non-farm household Rurcoanonf = rural coastal non-farm household Urbsavunsk = urban savannah unskilled household Urbforunsk = urban forest unskilled household Urbcoasunsk = urban coastal unskilled household Urbsavskil = urban savannah skilled household Urbforskil = urban forest skilled household Urbcoaskil = urban coastal skilled household *Note*: Correlation of these codes with mean per capita expenditure, mean years of schooling of head of household, mean number of households with electricity and building material of house also provided (Urbcoaskil highest and Rursavfarm lowest scores)
Measure(s) of association	Mean height-for-age z-score
Bivariate summary of results	*Mean height-for-age z-score by residence code*

Residence code	*Mean height-for-age z-score*
Rursavfarm	−1.46
Rurforfarm	−1.65
Rurcoafarm	−1.17
Rursavnonf	−1.08
Rurfornonf	−1.15
Rurcoanonf	–0.89
Urbsavunsk	−1.69
Urbforunsk	−1.19
Urbcoasunsk	–0.79
Urbsavskil	–0.26
Urbforskil	−1.15
Urbcoaskil	–0.71

Multivariate summary of results	Not applicable
Comments	Article focuses on how strongly BMI or other nutritional measures are correlated with various measures of SES

Author/year	Quinn et al. (1995)
Subregion (country)	AFR-E (Malawi)
Study type/design	Meta-analysis of survey data/three cross-sectional surveys carried out during similar time periods using same protocol for anthropometric data collection
Study population	Children aged over 24 months. Characteristics of three surveys are (i) preschool children attending institutions known to cater to high-income families in urban centres (Blantyre, Lilongwe, Zomba) (n = 350); (ii) rural survey of villages conducted in Ntchisi district (n = 667); (iii) survey of low-income urban neighbourhoods conducted in Blantyre and Lilongwe (n = 225)
Survey year(s)	1990–1991, 1989, 1991, respectively
Risk factor measure(s)	Prevalence of stunting or wasting (height-for-age and weight-for-height, respectively, below –2 z-score) compared with mean of CDC/WHO reference population
Socioeconomic measure(s)	As per survey sample: 1 = high income urban area; 2 = low income rural area; 3 = low income urban area
Measure(s) of association	Mean height and weight by survey
Bivariate summary of results	*Prevalence of stunting or wasting (height-for-age and weight-for-height respectively below –2 z-score), all ages (24–59 months) combined*

	1. Urban rich	*2. Rural poor*	*3. Urban poor*
Stunting	7.8	83.2	69.3
Wasting	0.6	0.4	4.0

Multivariate summary of results	Not applicable
Comments	No measures of underweight

Author/year	Popkin et al. (1993)
Subregion (country)	WPR-B (China)
Study type/design	Cross-sectional survey (1989 China Health and Nutrition Survey)/multistage random cluster sampling
Study population	Nationally representative sample (n = 3800 households)
Survey year(s)	1989
Risk factor measure(s)	Prevalence of stunting and wasting (defined as less than –2 SD below median of CDC/WHO population for height-for-age and weight-for-height, respectively) for children aged 1–6 years
Socioeconomic measure(s)	Area of residence (urban, rural) Total income (including market and non market activities and non-monetary government subsidies) (low, medium, high)
Measure(s) of association	Prevalence of stunting and wasting by socioeconomic measure

Bivariate summary of results

Prevalence of stunting and wasting by area of residence and income group among children aged 1–6 years

Socioeconomic measure	*Stunting*	*Wasting*
URBAN		
Low income	26.3%	9.3%
Medium income	16.4%	6.1%
High income	17.3%	9.7%
Total	19.4%	8.4%
RURAL		
Low income	43.7%	9.6%
Medium income	34.5%	6.8%
High income	28.3%	4.5%
Total	37.8%	7.7%
URBAN AND RURAL		
Low income	40.9%	9.5%
Medium income	28.9%	6.6%
High income	24.0%	6.6%
Total	33.0%	7.9%

Multivariate summary of results	Not applicable
Comments	No further details of income are given

Author/year Radebe et al. (1996)

Subregion (country) AFR-E (Zimbabwe)

Study type/design Unmatched case–control study

Study population Children aged 6 months to 5 years in Mazowe district who attended primary health care clinics within the study period for routine immunization, weighing and treatment of minor illnesses (excluding children not brought by their mother, children of single or young (<16 years) mothers, twins, handicapped or low-birth-weight children, and HIV-positive or ill children) (n = 327: 176 cases and 151 controls)

Survey year(s) 1990

Risk factor measure(s) Well nourished (>90% expected weight-for-age) as controls
Malnourished (<80% expected weight-for-age) as cases

Socioeconomic measure(s) Live on communal or commercial farms
Mother educated ≤ grade 3
Mother employed (Yes/No)
No extra maternal income, e.g. from knitting
No support from father
Father's income <Z$ 150 per month
Family owns ≤1 acre of land (Yes/No)

Measure(s) of association Odds ratio

Bivariate summary of results *Factors associated with malnutrition*

COMMUNAL LAND

Risk factor	*OR*	*95% CI*	*P value*
Mother educated ≤ grade 3	3.01	1.25–7.39	0.011
Mother employed	4.68	2.04–10.84	0.0001
No extra maternal income	1.92	0.91–4.0	Null
No support from father	3.37	1.41–8.21	0.005
Paternal income <Z$ 150/month	12.50	2.5–100	0.001
Land ≤1 acre	2.56	1.05–6.33	0.036

COMMERCIAL FARM

Risk factor	*OR*	*95% CI*	*P value*
Mother educated ≤ grade 3	1.53	0.80–2.94	Null
Mother employed	1.44	0.77–2.70	Null
No extra maternal income	3.92	1.64–9.54	Null
No support from father	1.52	0.44–5.56	0.001
Paternal income <Z$ 150/month	5.56	1.41–25.00	Null
Land ≤1 acre	—	—	0.01

Multivariate summary of results *Factors associated with malnutrition on communal land (after correcting for confounding variables)*

Risk factor	*OR*	*95% CI*
Mother employed	5.69	2.37–13.66
Land ≤1 acre	3.66	1.40–9.53

Comments Multivariate analysis not performed on commercial farm data, as only two variables were significant. Maternal education not associated with childhood malnutrition in this study (in contrast to 1988 Zimbabwe DHS). Possible selection bias and not representative, as excludes mothers who do not go to clinics

Author/year	Ricci and Becker (1996)
Subregion (country)	WPR-B (Philippines)
Study type/design	Longitudinal cross-sectional survey/data collection performed by Office of Population Studies of the University of San Carlos as part of larger longitudinal study to evaluate health services provided by Philippines Department of Health in Metro Cebu Periodic surveys of all households in fixed samples of seven urban and 26 rural barangays as long as a child <30 months of age was resident
Study population	Children aged 0–30 months in Metro Cebu (n = 18 544 enrolled at beginning of study), from rural and urban areas, divided into three categories: 0–5 months, 6–11 months, 12–29 months
Survey year(s)	1988–1990
Risk factor measure(s)	Prevalence of stunting and wasting (defined as less than –2 SD below median of CDC/WHO reference population for height-for-age and weight-for-height, respectively)
Socioeconomic measure(s)	Area of residence (urban/rural) Television and radio presence/absence Wall material: wood, cement, all else Flooring: bamboo/all else, any wood, any cement Cooking fuel: wood, gas or oil Water source: well, purchased, pipe/pump Toilet type: none, pit, sealed Mother's employment: none, employed at home, employed away from home Formal education of father (in years) Formal education of mother (in years)
Measure(s) of association	Odds ratio
Bivariate summary of results	*Stunting* URBAN

	Age group (months)		
Variable	*0–5*	*6–11*	*12–29*
TV/Radio			
Radio only	—	—	0.75
Television	—	—	0.60
Neither (reference)	—	—	1
Water source			
Purchased	—	—	0.92*
Pump/pipe	—	—	0.97
Well (reference)	—	—	1
Toilet type			
Pit	—	—	0.95
Sealed	—	—	0.77
None (reference)	—	—	1
Flooring material			
Wood	—	0.77	0.80
Cement	—	0.66	0.74
Bamboo/all else (reference)	—	1	1
Cooking fuel			
Gas or oil	—	—	0.70
Wood (reference)	—	—	1

Father's education (years)	—	0.92	0.95
Mother's education (years)	0.91	—	0.96

RURAL

	Age group (months)		
Variable	*0–5*	*6–11*	*12–29*
TV/radio			
Radio only	0.59	0.74	0.86
Television	0.61	0.56	0.77
Neither (reference)	1	1	1
Water source			
Purchased	—	—	0.92
Pump/pipe	—	—	0.80
Well (reference)	—	—	1
Toilet type			
Pit	—	0.82	0.83
Sealed	—	0.66	0.71
None (reference)	—	1	1
Flooring material			
Wood	—	—	1.01*
Cement	—	—	0.81
Bamboo/all else (reference)	—	—	1
Cooking fuel			
Gas or oil	—	—	0.70
Wood (reference)	—	—	1
Father's education (years)	0.93	0.94	0.93
Mother's education (years)	—	—	0.95

(all $P<0.05$ except * not significant)

Wasting

URBAN

	Age group (months)		
Variable	*0–5*	*6–11*	*12–29*
TV/radio			
Radio only	—	1.05	—
Television	—	0.66	—
Neither (reference)	—	1	—
Wall material			
Cement	—	0.80	—
All else	—	0.47	—
Wood (reference)	—	1	—
Mother's employment			
At home	—	—	0.94
Away from home	—	—	0.77
None (reference)	—	—	1
Father's education (years)	—	—	0.95
Mother's education (years)	—	—	0.96

RURAL

	Age group (months)		
Variable	*0–5*	*6–11*	*12–29*
TV/radio			
Radio only	—	—	0.90

Television	—	—	0.58
Neither (reference)	—	—	1
Water source			
Purchased	—	1.59	—
Pump/pipe	—	0.95	—
Well (reference)	—	1	—
Mother's employment			
At home	—	—	1.33
Away from home	—	—	0.91
None (reference)	—	—	1

(all $P < 0.05$)

Multivariate summary of results

Model applied
Similar results as for bivariate analysis
Results not given in paper

Comments

No indication of how representative this study is for the Philippines as a whole. Weight-for-age given only at descriptive level, as investigators did not think it differentiated stunted from wasted children.
Investigators showed a clear socioeconomic difference between urban and rural households (in terms of housing characteristics such as water source, cooking fuel, ownership of television and radio, floor and wall materials, as well as parental education) but similar household size and density

Water and sanitation

Good quality water and adequate sanitation are commonly regarded as basic socioeconomic prerequisites for health, and thus their lack is associated with low SES. It is difficult to find individual-level data on water supply, as most research on water quality examines the issue at aggregate rather than individual level.

As water quality and sanitation facilities are regarded as indicators of SES, and the relationship between poverty and poor water quality and sanitation is widely accepted, it was difficult to find appropriate articles as the relationship between these two factors is rarely examined. More often, water quality and/or sanitation are used as part of socioeconomic indices. As a result, only nine journal articles were revealed by the literature review search, of which three were retrieved. On examination, two of these articles were deemed unsuitable for our purposes and were discarded.

The remaining article reported on a study carried out in rural Bangladesh. However, some of the subjects of the study were recipients of support from nongovernmental organizations (NGOs) and thus were not necessarily typical in terms of education and access to water and sanitation. In general, however, there was a strong positive relationship between SES and safe sanitation and a weaker positive relationship between SES and safe waste disposal. A number of measures for SES were used, including educational level, housing level, occupation and land ownership. A similar pattern was seen across all these measures.

Author/year	Hadi (2000)
Region (country)	SEAR-D (Bangladesh)
Study type/design	Survey/random sample of proportionally representative households selected from those covered by Watch, the demographic and health surveillance system that covers 70 villages located in 10 regions of Bangladesh (n = 1556 households)
Study population	A total of 1556 households representative of those covered by Watch Rural households only
Survey year(s)	1995
Risk factor measure(s)	Safe sanitation behaviour (Yes/No) Safe disposal of solid waste (Yes/No)
Socioeconomic measure(s)	In Bangladesh, households that own <50 decimals of land and in which the principal worker had to sell at least 100 days of labour over the past year in order to subsist are eligible for aid from NGOs Survey sample was divided into: participants (those households eligible for and receiving NGO aid); non-participants (those households eligible for but not receiving NGO aid); and not eligible (those households not eligible for NGO aid) Comparison of mean years of schooling, percentage literate, mean amount of land (in decimals), occupation and exposure to the media showed that non-participants were the most disadvantaged, and the not eligible group were the least disadvantaged Education: none, 1–5 years, ≥6 years Land ownership: none, 1–199 decimals, ≥200 decimals Occupation: labourer, agricultural worker, service/business sector Housing conditions: poor, good Exposure to the media: poor, better
Measure(s) of association	Prevalence of safe sanitation behaviour and safe disposal of solid waste by socioeconomic measure; log odds ratio (looking at variables outside of CRA)
Bivariate summary of results (including controlling for sex and age)	*Prevalence of safe sanitation behaviour (i.e. sanitary latrine use) and safe disposal of solid waste by socioeconomic measure*

Socioeconomic measure	*Safe sanitation behaviour*	*Safe disposal of solid waste*
NGO eligibility status		
Non-participants	6.7	44.0
Participants	23.0	47.0
Not eligible	35.3	45.9
Education		
None	15.7	44.7
1–5 years	30.3	44.6
≥6 years	48.8	50.7
Land ownership		
Landless	17.2	48.5
1–199 decimals	26.6	40.1
≥200 decimals	48.5	52.5

	Occupation		
	Labour	18.5	44.3
	Agriculture	28.9	46.4
	Service/business	39.2	50.4
	Housing conditions		
	Poor	19.7	41.3
	Good	41.2	60.2
	Exposure to the media		
	Poor	19.0	45.6
	Better	40.7	46.4

Looked also at intention to buy safe latrine, knowledge on how to build, knowledge of potential health benefits of safe latrine and waste disposal
Lack of knowledge even among the better-off households
Concept of community-managed or jointly owned latrines not popular, probably for social reasons

Multivariate summary of results (including controlling for rurality, ethnicity, etc.)

Not relevant to CRA

Comments

Only looked at rural areas
Potentially useful
NGO intervention could lead to confounding for some variables (i.e. literacy programmes for very poor etc.); care would have to be taken
No further definitions were provided of education, housing conditions and exposure to the media

Unsafe sex

Two journal articles on unsafe sex were retrieved of the 30 revealed by the literature search. One was unsuitable for our purposes and thus discarded. The remaining article was the result of a study on knowledge of sexually transmitted infections (particularly syphilis and gonorrhoea) among married women in rural Bangladesh. The results showed that increasing knowledge of these sexually transmitted infections correlated with increasing level of education among these women. This knowledge was also more common among women whose husbands were professionals or service-holders.

Author/year	Khan et al. (1997)
Subregion (country)	SEAR-D (Bangladesh)
Study type/design	Survey
Study population **Sample characteristics: region, urban/rural, socioeconomic strata, random, sample size**	Married women of reproductive age in four thanas (Abhoynagar, Bagherpara, Keshopur, Sirajganj) of Bangladesh (n = 8674)
Survey year(s)	1994
Risk factor measure(s)	Knowledge of which diseases are spread through sexual intercourse Knowledge of how syphilis and gonorrhoea are spread if they have heard of them Knowledge of how to prevent sexually transmitted diseases
Socioeconomic measure(s)	Subject's educational level: none, primary, above primary Husband's educational level: none, primary, above primary Husband's occupation: farmer or self-employed, labourer, trader/businessman, service-holder, professional
Measure(s) of association	Prevalence of awareness of syphilis and gonorrhoea Odds ratio
Bivariate summary of results (including controlling for sex and age)	*Prevalence of awareness of syphilis and gonorrhoea and how they are spread, by socioeconomic measure*

Socioeconomic variable	*Awareness of syphilis/gonorrhoea*
Woman's education	
None	21.5%
Primary	27.8%
Above primary	40.7%
χ^2	173.82
P value	<0.001
Husband's education	
None	21.5%
Primary	25.5%
Above primary	38.6%
χ^2	186.47
P value	<0.001
Husband's occupation	
Farmer/self employed	26.2
Labourer	23.4
Trader/business	30.2
Service-holder	45.4
Professional	53.4
χ^2	123.03
P value	<0.001

Multivariate summary of results (including controlling for rurality, ethnicity, etc.)

Odds ratios measuring association between socioeconomic measures and awareness of syphilis and gonorrhoea

Socioeconomic variable	*OR*	*95% CI*
Woman's education		
None	1.00	
Primary	1.55	1.34–1.78
Above primary	2.93	2.38–3.60
Husband's education		
None	1.00	
Primary	1.27	1.10–1.46
Above primary	1.68	1.42–1.98
Husband's occupation		
Farmer/self employed	1.00	
Labourer	1.04	0.91–1.19
Trader/business	1.12	0.94–1.36
Service-holder	1.36	1.02–1.81
Professional	1.72	1.19–2.49

Appendix references

Anonymous (1991) Zimbabwe 1988: results from the demographic and health survey. *Studies in Family Planning*, **22**:395–399.

Anonymous (1992) Pakistan 1990/91: results from the demographic and health survey. *Studies in Family Planning*, **23**:274–278.

Anonymous (1998) Dominican Republic 1996: results from the demographic and health survey. *Studies in Family Planning*, **29**:423–427.

Anonymous (2000) Bolivia 1998: Results from the demographic and health survey. *Studies in Family Planning*, **31**:257–261.

Chen CM (1996) Nutrition status of the Chinese people. *Biomedical and Environmental Sciences*, **9**:81–92.

Delpeuch F, Traissac P, Martin-Prevel Y, Massamba JP, Maire B (2000) Economic crisis and malnutrition: socioeconomic determinants of anthropometric status of preschool children and their mothers in an African urban area. *Public Health Nutrition*, **3**:39–47.

Hadi A (2000) A participatory approach to sanitation: experience of Bangladeshi NGOs. *Health Policy and Planning*, **15**:332–337.

Khan MA, Rahman M, Khanam PA, Barkat-e-Khuda, Kane TT, Ashraf A (1997) Awareness of sexually transmitted disease among women and service providers in rural Bangladesh. *International Journal of STD and AIDS*, **8**:688–696.

Kikafunda JK, Walker AF, Collett D, Tumwine JK (1998) Risk factors for early childhood malnutrition in Uganda. *Pediatrics*, **102**:E45.

Li Y, Guo G, Shi A, Anme T, Ushijima H (1999) Prevalence and correlates of malnutrition among children in rural minority areas of China. *Pediatrics International*, **41**:549–556.

Madzingira N (1995) Malnutrition in children under five in Zimbabwe: effect of socioeconomic factors and disease. *Social Biology*, **42**:239–246.

Matulessy PF, Asumi R, Thamrin MH, Husaini, Angeles TI, Kariadi D (1992) Nutrition situation in Metropolitan Jakarta. *Southeast Asian Journal of Tropical Medicine and Public Health*, **23**:S3–16.

Monteiro CA, Benicio MH, Iunes R, Gouveia NC, Taddei JA, Cardoso MA (1992) Nutritional status of Brazilian children: trends from 1975 to 1989. *Bulletin of the World Health Organization*, **70**:657–666.

Monteiro CA, Mondini L, de Souza AL, Popkin BM (1995) The nutrition transition in Brazil. *European Journal of Clinical Nutrition*, **49**:105–113.

Nube M, Asenso-Okyere WK, van den Boom GJ (1998) Body mass index as indicator of standard of living in developing countries. *European Journal of Clinical Nutrition*, **52**:136–144.

Popkin BM, Keyou G, Zhai F, Guo X, Ma H, Zohoori N (1993) The nutrition transition in China: a cross-sectional analysis. *European Journal of Clinical Nutrition*, **47**:333–346.

Quinn VJ, Chiligo-Mpoma MO, Simler K, Milner J (1995) The growth of Malawian preschool children from different socioeconomic groups. *European Journal of Clinical Nutrition*, **49**:66–72.

Radebe BZ, Brady P, Siziya S, Todd H (1996) Maternal risk factors for childhood malnutrition in the Mazowe district of Zimbabwe. *Central African Journal of Medicine*, **42**:240–244.

Ricci JA, Becker S (1996) Risk factors for wasting and stunting among children in Metro Cebu, Philippines. *American Journal of Clinical Nutrition*, **63**:966–975.

Wagstaff A, Watanabe N (2001) *Socioeconomic inequalities in child malnutrition in the developing world*. World Bank, Washington, DC.

Chapter 25

Estimating attributable burden of disease from exposure and hazard data

Stephen Vander Hoorn, Majid Ezzati, Anthony Rodgers, Alan D. Lopez and Christopher J.L. Murray

1. Estimating population attributable fractions

As described in earlier chapters, the contribution of a risk factor to disease burden (expressed as the fraction of disease or death attributable to the risk factor in a population) is given by the generalized "potential impact fraction" (PIF) in Equation 1a (Drescher and Becher 1997; Eide and Heuch 2001; Walter 1980).

$$PIF = \frac{\int_{x=0}^{m} RR(x)P(x)dx - \int_{x=0}^{m} RR(x)P'(x)dx}{\int_{x=0}^{m} RR(x)P(x)dx} \quad (1a)$$

$RR(x)$: relative risk at exposure level x

$P(x)$: population distribution of exposure

$P'(x)$: counterfactual distribution of exposure, and

m: maximum exposure level

The first and second terms in the numerator of Equation 1a represent the total exposure-weighted risk of disease or mortality in the population under current and counterfactual exposure distributions. The corresponding relationship when exposure is described as a discrete variable with n levels is given by:

$$PIF = \frac{\sum_{i=1}^{n} P_i\, RR_i - \sum_{i=1}^{n} P_i'\, RR_i}{\sum_{i=1}^{n} P_i\, RR_i} \tag{1b}$$

The PIF equation can be used to estimate the population attributable fraction (PAF), defined as the proportional reduction in disease or death that would occur if exposure to the risk factor were reduced to the counterfactual exposure distribution. The remainder of this chapter outlines how data on exposure, hazard and disease burden were combined to derive estimates of attributable disease burden, with estimation of the population attributable fraction as the intermediate step. The application of Equations 1a and 1b in the context of the comparative risk assessment (CRA) project is discussed and several issues regarding its implementation are detailed.

2. Estimating attributable mortality and burden of disease

For each risk factor–disease outcome pair, PAFs for each of the 224 age, sex, subregion[1] groups were calculated using the relationships in Equation 1, separately for mortality (PAF_M) and incidence (PAF_I) when the relative risks for mortality and incidence were different. For each of these 224 groups, the estimates of mortality (AM_{ij}) and burden of disease (AB_{ij}) from disease j attributable to risk factor i were calculated as below. Burden of disease, reported annually in the annexes of the *World Health Report*, was expressed in disability-adjusted life years (DALYs), with methods and assumptions described elsewhere (Murray and Lopez 1996). Specifically:

$$AM_{ij} = PAF_{M-ij} \times M_j$$

$$A\,YLL_{ij} = PAF_{M-ij} \times YLL_j$$

$$A\,YLD_{ij} = PAF_{I-ij} \times YLD_j$$

$$AB_{ij} = A\,YLL_{ij} + A\,YLD_{ij}$$

Where "A" indicates "attributable" and

YLL: years of life lost to premature mortality

YLD: years of life lived with disability due to disease incidence

For those risk factors with insufficient data to estimate a relative risk model (e.g. occupational or alcohol-caused injuries or the effects of lead exposure on blood pressure), disease burden or mortality was estimated using existing registers or corresponding hazard relationships. Estimates were then aggregated across age groups to obtain subregional estimates,

and across subregions to obtain global estimates. The details of this aggregation are described later in this chapter.

3. Counterfactual exposure distribution

The estimates of burden of disease and injuries due to risk factors in the CRA project are based on a counterfactual of theoretical-minimum-risk exposure distribution, defined in chapter 1 and described in individual risk factor chapters. By using the theoretical-minimum-risk exposure distribution, which by definition has a relative risk of 1.0, as the counterfactual exposure distribution or level, Equations 1a and 1b are reduced to:

$$AF = \frac{\int_{x=0}^{m} RR(x)P(x)dx - 1}{\int_{x=0}^{m} RR(x)P(x)dx} \tag{2a}$$

and

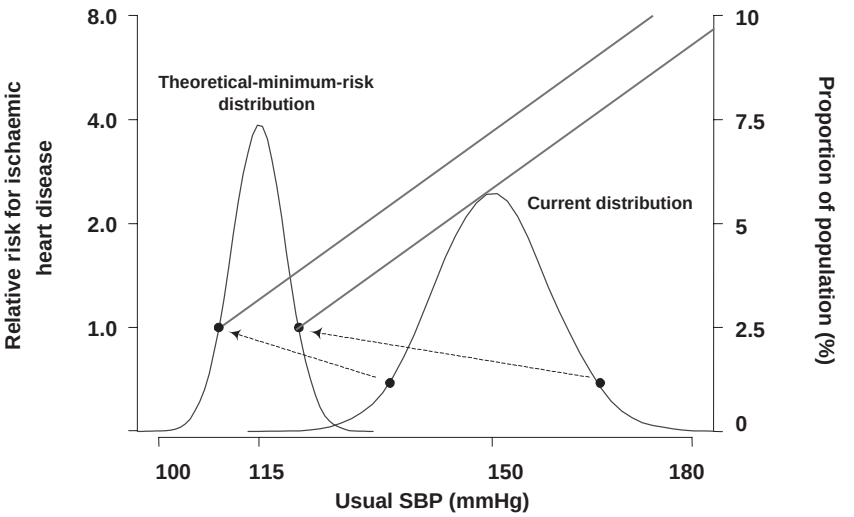
$$AF = \frac{\sum_{i=1}^{n} P_i(RR_i - 1)}{\sum_{i=1}^{n} P_i(RR_i - 1) + 1} \tag{2b}$$

3.1 Theoretical-minimum-risk exposure distribution for continuous exposure variables

The theoretical-minimum-risk exposure distribution for continuous risk factors is itself often a distribution of exposure levels, vs a constant baseline. Figure 25.1, for example, illustrates a scenario for systolic blood pressure (SBP) with typical exposure levels in an older population (mean: 150 mmHg; SD: 9 mmHg) compared with the theoretical-minimum-risk exposure distribution (mean: 115 mmHg; SD: 6 mmHg). The non-zero standard deviation of the theoretical-minimum-risk distribution reflects the reality that there always is some inter-person variability within any given population, even after hypothetical reductions such as that shown in Figure 25.1.

The optimal exposure distribution for a population would overlap precisely with the theoretical-minimum-risk exposure distribution. By definition of theoretical-minimum risk, such a population would be collectively without any increased risk and therefore with zero attributable burden due to the risk factor of interest. Any population containing individuals outside this distribution will then have a population attributable fraction greater than zero and exposure distributions converging on the

Figure 25.1 Theoretical-minimum-risk exposure distribution for continuous risk factors using systolic blood pressure (SBP) as an example



Note: Each point represents a hypothetical individual or small group of individuals in the population. The solid straight lines represent the increasing relative risk, on a log scale, for ischaemic heart disease with increasing SBP.

theoretical minimum will have attributable burden tending towards zero. The risk for any *individual* (or groups of individuals in a narrow range of exposure) in the population would be determined by the difference between her/his current exposure (SBP level) and the SBP level that s/he would have when the population distribution overlaps with the theoretical-minimum-risk exposure distribution.

Estimating the total hazard at the population level can be achieved using a micro-simulation approach in which individuals are drawn randomly from current and theoretical-minimum-risk exposure distributions. For most risk factors, however, individual exposures "track" over relatively long periods of time (Lauer and Clarke 1988; Voors et al. 1979; Wilsgaard et al. 2001). In other words, those with higher/lower exposure levels of a particular risk factor are expected to have higher/lower exposure levels within the theoretical-minimum-risk exposure distribution (see the hypothetical individuals in Figure 25.1). Random (uncorrelated) draws of individuals from current and theoretical-minimum-risk exposure distributions would be inconsistent with the empirical evidence on tracking. Consistent with this evidence, we assumed that the ordering of individuals in the exposure distribution remains unchanged (i.e.

the rank-order correlation of individual exposures equals 1) in the transition to the theoretical-minimum-risk distribution in estimating the PAF.

With correlated rank-ordering of individuals in current and theoretical-minimum-risk exposure distributions, if hazards were a linear function of exposure, then for those risks with symmetric distributions, shifting the population to the theoretical minimum distribution would be computationally equivalent to shifting everyone to the mean exposure of the theoretical-minimum-risk exposure distribution (i.e. the standard deviation of the theoretical-minimum-risk exposure distribution would not change the total hazard). This is because, with a linear hazard function, the changes in hazards for individuals above and below the mean, as a result of changing the standard deviation of the theoretical-minimum-risk exposure distribution, would fully compensate each other.

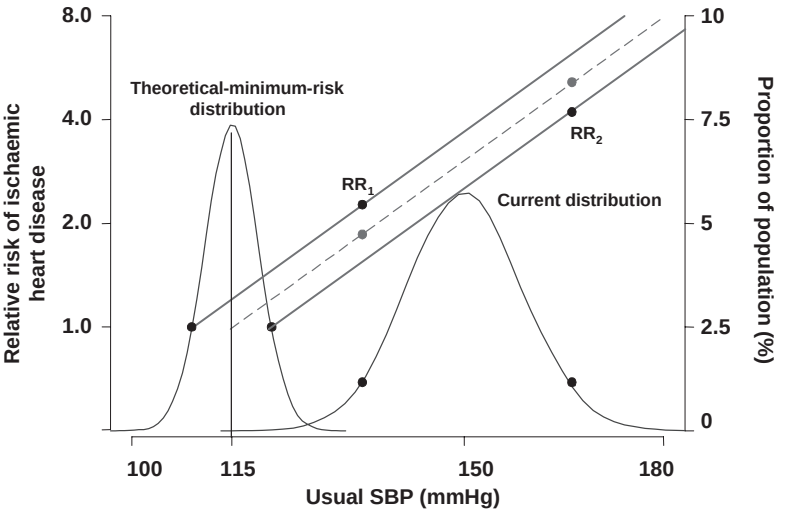
Risk is, however, an exponential function of exposure in most epidemiological models. With an exponential hazard function, when the baseline is the mean of theoretical-minimum-risk exposure distribution, the integrated risk is larger for those above the mean than those below the mean (Figure 25.2), compared to the case of treating theoretical-minimum-risk exposure as a distribution with non-zero standard deviation. The net difference would depend on both the steepness of the risk curve (i.e. increased risk per unit increase in exposure) and the standard deviations of the current and theoretical-minimum-risk exposure distributions.

For computational reasons, we estimated PAFs for continuous risk factors relative to the mean of the theoretical-minimum-risk exposure distribution. In these calculations, the relative risk for any individual in the population with an exposure below the mean of the theoretical-minimum-risk exposure distribution was set to 1.0 (e.g. in Figures 25.1 and 25.2, the lower tail of current blood pressure distribution is inside the theoretical minimum distribution with some individuals already at a level below 115 mmHg. These individuals were assigned a relative risk of 1.0). Sensitivity analysis showed that in the scenarios analysed in the CRA project, global PAFs estimated by integrating risk relative to the mean of the theoretical-minimum-risk exposure distribution were up to 2% larger than those estimated by integrating risk relative to the full distribution. As described above, this is because of the non-linear shape of most hazard functions.

4. Aggregation of attributable burden across age, sex and subregion

Within each of the 14 subregions, all-age-sex population attributable fractions ($PAF_{subregion}$) were calculated by aggregating attributable burden estimates across the 16 age-sex-specific estimates within the subregion

Figure 25.2 Effect of a non-linear hazard function and choice of baseline on total population risk



Note: With an exponential hazard function, when theoretical-minimum-risk exposure is a distribution with a non-zero standard deviation, those falling above the mean of the current distribution (e.g. 155 mmHg for SBP) contribute more to total population hazard than those below it, relative to the case when the baseline is a constant level (115 mmHg for SBP). In the figure, the solid lines represent the hazard when the theoretical-minimum-risk exposure is a distribution, and the dotted line when a constant baseline is considered. The difference between the two relative risks on the right (RR_2) is larger than those on the left (RR_1). As a result of this imbalanced contribution to hazard, using the mean of theoretical-minimum-risk exposure distribution as baseline in estimating total population hazard would result in slightly larger PAFs than using the complete theoretical minimum distribution.

and then dividing by the total subregional disease burden using the relationship in Equation 3 (the estimates could similarly be aggregated across ages separately for males and females).

$$PAF_{subregion} = \frac{\sum_{age,sex} AB_{subregion,age,sex}}{\sum_{age,sex} B_{subregion,age,sex}} \quad (3)$$

Similarly, for each age-sex group, world attributable fractions ($PAF_{age,sex}$) were calculated by aggregating attributable burden estimates across all the 14 subregion-specific estimates and then dividing by the disease burden for that age-sex group using the relationship in Equation 4.

$$PAF_{age,sex} = \frac{\sum_{subregion=1}^{14} AB_{subregion,age,sex}}{\sum_{subregion=1}^{14} B_{subregion,age,sex}} \quad (4)$$

This is shown in Tables 25.1–25.3 for the case of SBP and ischaemic heart disease (IHD). The non-italic numbers in Table 25.1 are the subregion-age-sex specific PAFs estimated using Equation 2. Next, these fractions were applied to the Global Burden of Disease (GBD) 2000 estimates of disease burden for IHD, shown in Table 25.2, producing the estimates of IHD disease burden attributable to SBP in Table 25.3 (similar estimates could be made for mortality or YLL). Dividing the total attributable burden in any subregion (e.g. 1.548 million DALYs for AMR-A in the highlighted cell) by the total IHD burden for the subregion in the GBD database (3.506 million DALYs for AMR-A in the highlighted cell) gives the all-age-sex subregional PAFs (44% for AMR-A in the highlighted cell, obtained by dividing 1.548 by 3.506).

The all-age-sex PAF estimates for the remaining 13 subregions are also shown in Table 25.1 in italics. Similarly, world PAFs were calculated within each age and sex group, as well as overall, using Equation 4, and are shown in the bottom row of Table 25.1. For example, the world PAF for 60–69 years old males was obtained by dividing the total attributable burden in that age-sex group (4.71 million DALYs) by the total world IHD burden in the GBD database (9.015 million DALYs), giving a PAF of 52%.

Computationally, aggregate PAFs (whether aggregated across age-sex groups or subregions) are equivalent to weighted averages of the subregion-age-sex specific estimates, with weights being the same as the total number of events (i.e. deaths, YLL, or DALYs) for each subregion-age-sex group. In the above example, total subregion-age-sex specific DALYs are the weighting factor. As a result, subregion and world PAFs are weighted more towards the ages and/or subregions that have higher DALYs (rather than those with larger populations). For each risk factor, these separate aggregate PAFs were estimated for deaths, YLL, and DALYs with the corresponding GBD estimates used as the denominator (or weighting factor). As a result, even when the age-sex-subregion-specific PAFs are the same for the three measures, the aggregate ones may differ.

For example, although the subregion-age-sex-specific PAFs are the same for deaths and DALYs in the case of elevated SBP and IHD, dividing the IHD deaths attributable to this risk factor in AMR-A (203 000) by the total IHD deaths in this region (618 000), gives a regional PAF for mortality of 33% in AMR-A. In fact, the smaller subregional PAF for mortality compared to that of DALYs highlights the higher weighting towards the older age PAFs (which are smaller) in the case of

Table 25.1 PAFs for IHD attributable to increased SBP (%), by age, sex and subregion

	0–4 (years)		*5–14 (years)*		*15–29 (years)*		*30–44 (years)*		*45–59 (years)*		*60–69 (years)*		*70–79 (years)*		*≥80 (years)*		
	Males	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Total*
Subregion																	
AFR-D	—	—	—	—	—	—	63	56	70	73	56	63	50	58	16	19	*57*
AFR-E	—	—	—	—	—	—	51	39	59	61	46	53	40	48	12	15	*46*
AMR-A	—	—	—	—	—	—	46	20	57	53	49	51	46	48	15	18	*44*
AMR-B	—	—	—	—	—	—	46	23	61	60	51	57	47	54	14	17	*50*
AMR-D	—	—	—	—	—	—	50	31	60	58	48	52	43	49	13	15	*45*
EMR-B	—	—	—	—	—	—	55	57	64	71	53	61	48	56	15	19	*55*
EMR-D	—	—	—	—	—	—	49	44	61	68	51	61	46	57	14	19	*51*
EUR-A	—	—	—	—	—	—	66	47	72	70	59	62	53	57	17	19	*54*
EUR-B	—	—	—	—	—	—	63	49	77	78	64	70	59	66	20	23	*64*
EUR-C	—	—	—	—	—	—	66	55	74	80	62	73	56	68	18	23	*63*
SEAR-B	—	—	—	—	—	—	43	39	60	60	51	52	46	48	14	15	*46*
SEAR-D	—	—	—	—	—	—	33	30	53	49	46	42	42	38	13	11	*41*
WPR-A	—	—	—	—	—	—	60	40	71	64	59	58	54	53	18	17	*52*
WPR-B	—	—	—	—	—	—	28	23	51	55	44	51	41	48	12	14	*41*
World	—	—	—	—	—	—	*47*	*35*	*61*	*59*	*52*	*54*	*48*	*53*	*15*	*18*	*49*

— No data.

Table 25.2 GBD 2000 estimates of total disease burden for IHD (000s of DALYs), by age, sex and subregion

	0–4 (years)		*5–14 (years)*		*15–29 (years)*		*30–44 (years)*		*45–59 (years)*		*60–69 (years)*		*70–79 (years)*		*≥80 (years)*		
	Males	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Total*
Subregion																	
AFR-D	3	4	6	5	31	48	116	88	242	281	190	189	122	166	37	48	1576
AFR-E	4	5	7	7	49	66	149	88	273	278	200	188	110	148	29	50	1653
AMR-A	1	1	1	0	15	7	214	74	709	260	514	289	470	391	219	340	3506
AMR-B	2	2	3	2	46	21	218	107	594	301	412	279	255	223	69	96	2631
AMR-D	1	2	1	1	10	7	25	11	54	33	42	30	31	24	9	11	294
EMR-B	3	5	6	4	32	17	179	59	411	143	219	128	114	99	26	29	1474
EMR-D	24	24	21	12	90	84	322	170	771	475	522	469	292	328	69	73	3746
EUR-A	1	0	1	7	15	15	181	40	657	146	717	284	671	518	239	389	3882
EUR-B	1	0	1	1	35	15	266	78	734	258	647	440	431	471	93	176	3647
EUR-C	0	0	1	1	78	15	693	127	1708	477	1496	920	847	1223	171	563	8319
SEAR-B	7	3	4	3	98	49	222	123	394	279	318	286	188	188	43	54	2259
SEAR-D	70	52	99	58	283	622	1105	900	3688	2119	2657	2361	1412	1484	278	291	17480
WPR-A	0	0	3	3	10	6	40	10	151	40	135	59	112	85	46	66	765
WPR-B	8	7	14	7	155	77	495	278	1099	665	945	783	682	776	182	340	6513
World	125	107	168	111	946	1049	4225	2154	11484	5755	9015	6704	5737	6127	1510	2526	57743

Table 25.3 Burden of IHD attributable to increased SBP (000s of DALYs), by age, sex and subregion

	0–4 (years)		*5–14 (years)*		*15–29 (years)*		*30–44 (years)*		*45–59 (years)*		*60–69 (years)*		*70–79 (years)*		*≥80 (years)*		
	Males	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Total*
Subregion																	
AFR-D	—	—	—	—	—	—	73	49	168	205	107	119	61	96	6	9	893
AFR-E	—	—	—	—	—	—	76	35	160	171	92	99	45	72	3	8	760
AMR-A	—	—	—	—	—	—	97	15	403	137	251	146	215	189	34	61	1 548
AMR-B	—	—	—	—	—	—	101	25	364	180	211	158	119	120	10	16	1 303
AMR-D	—	—	—	—	—	—	12	3	33	19	20	16	13	12	1	2	132
EMR-B	—	—	—	—	—	—	98	34	264	101	116	78	55	56	4	5	811
EMR-D	—	—	—	—	—	—	156	75	473	324	265	285	134	186	10	14	1 922
EUR-A	—	—	—	—	—	—	120	19	476	102	422	175	354	296	41	74	2 079
EUR-B	—	—	—	—	—	—	167	38	565	201	417	308	256	309	18	40	2 320
EUR-C	—	—	—	—	—	—	456	71	1 270	381	927	670	475	826	31	132	5 239
SEAR-B	—	—	—	—	—	—	95	48	237	168	161	149	87	90	6	8	1 050
SEAR-D	—	—	—	—	—	—	367	270	1 972	1 030	1 222	988	595	568	35	33	7 080
WPR-A	—	—	—	—	—	—	24	4	107	26	80	34	60	46	8	12	400
WPR-B	—	—	—	—	—	—	138	65	557	366	420	399	278	370	22	49	2 664
World	—	—	—	—	—	—	1 983	750	7 049	3 412	4 710	3 625	2 746	3 234	229	463	28 201

— No data.

mortality, because greater numbers of IHD deaths occur in these age groups. On the other hand, because deaths at younger ages contribute to larger loss of life (YLL), when DALYs are considered, the contribution of PAFs at younger ages to the all-age-sex PAF becomes greater.

5. Exceptions to the general estimation procedure

The following list briefly describes the major departures from the standard analysis framework which were required so that all risk factors could be adequately assessed within the project. Further details are provided in the individual risk factor chapters.

- Theoretical minima varied by age, sex and subregion for iron deficiency, since this was the haemoglobin distribution that would be observed if iron deficiency were eliminated from each population.
- Theoretical minima varied by age, sex and subregion for lack of contraception as a risk factor to reflect different fertility preferences across populations.
- Fruit and vegetable intakes in any population were truncated at zero. In other words, all individuals falling below zero in the distribution were allocated a value of zero. Sensitivity analyses were also performed to assess the effects of possible skewness in the distribution of fruit and vegetable intake.
- The burden of cardiovascular diseases attributable to lead exposure was estimated by assessing different scenarios of elevated blood pressure due to lead and then estimating the total mediated effect through blood pressure.

6. Other methodological issues

It has been shown that the attributable fraction estimates using the PIF relationship in Equation 1 lead to biased estimates when the relative risk has been adjusted for confounding (Greenland 1984; Greenland and Robins 1988). This bias is in fact a result of the correlation among multiple risks (the risk factor of interest and other risk factors that act as confounders), as well as the diseases affected by them (Ezzati et al. 2003). Accounting for this correlation in the estimation of attributable burden, however, would require the availability of exposure and disease data stratified by the confounding variable(s). In general, such stratified data are not available and therefore reliance on the formula with direct use of the adjusted relative risk estimates was necessary. In the case of positive correlation among risk factors, this would generally result in an underestimation of population attributable fraction.

Note

1 See preface for an explanation of this term.

References

Drescher K, Becher B (1997) Estimating the generalized impact fraction from case-control data. *Biometrics*, **53**:1170–1176.

Eide GE, Heuch I (2001) Attributable fractions: fundamental concepts and their visualization. *Statistical Methods in Medical Research*, **10**:159–193.

Ezzati M, Vander Hoorn S, Rodgers A, Lopez AD, Mathers CD, Murray CJ, Comparative Risk Assessment Collaborating Group (2003) Estimates of global and regional potential health gains from reducing multiple major risk factors. *The Lancet*, **362**:271–280.

Greenland S (1984) Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. *Statistics in Medicine*, **3**:131–141.

Greenland S, Robins JM (1988) Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology*, **128**: 1185–1197.

Lauer R, Clarke W (1988) A longitudinal view of blood pressure during childhood: the Muscatine Study. *Statistics in Medicine*, **7**:47–57.

Murray CJL, Lopez AD, eds. (1996) *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Global Burden of Disease and Injury, Vol 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Voors A, Webber L, Berenson G (1979) Time course studies of blood pressure in children—the Bogalusa Heart Study. *American Journal of Epidemiology*, **109**:320–334.

Walter SD (1980) Prevention of multifactorial disease. *American Journal of Epidemiology*, **112**:409–416.

Wilsgaard T, Jacobsen BK, Schirmer H et al. (2001) Tracking of cardiovascular risk factors: the Tromso Study, 1979–1995. *American Journal of Epidemiology*, **154**:418–426.

Chapter 26

Mortality and burden of disease attributable to individual risk factors

Majid Ezzati, Anthony Rodgers, Alan D. Lopez, Stephen Vander Hoorn and Christopher J.L. Murray

Population attributable fractions (PAFs) for mortality and burden of disease attributable to individual risk factors were calculated, as described in chapter 25, using risk factor exposure and hazard estimates provided in risk factor chapters. Mortality and burden of disease attributable to individual risk factors were then calculated by multiplying the PAFs by the estimates of total mortality and burden of disease from the Global Burden of Disease (GBD) databases in each of the 224 subregion-age-sex groups, as described in chapter 25. These results are presented in the Annex Tables for each risk factor and summarized here across risks.

1. Aggregate disease burden attributable to individual risk factors

All-cause mortality and burden of disease estimates for females and males attributable to CRA risk factors in the 14 subregions[1] are presented in Table 26.1. Figure 26.1 shows the contribution of the 20 leading global risk factors to mortality and burden of disease in the world and three broad combinations of subregions—demographically and economically developed (AMR-A, EUR and WPR-A), low-mortality developing (AMR-B, EMR-B, SEAR-B and WPR-B) and high-mortality developing (AFR, AMR-D, EMR-D and SEAR-D). Figure 26.2 presents the burden of disease due to the leading 10 risk factors for each subregional grouping, also showing the cause composition, divided into broad groups of diseases and injuries. The different ordering of risk factors in their contributions to mortality and disease burden reflects the age profile of mortality (e.g. under-five mortality for underweight has larger

Portions of this chapter have been published previously in *The Lancet*, 2002, **360**: 1347–1360, and have been reproduced with permission from Elsevier Science.

Table 26.1(a) Mortality for females and males due to selected risk factors in 14 subregions

	AFRICA Mortality stratum		*THE AMERICAS* Mortality stratum			*EASTERN MEDITERRANEAN* Mortality stratum	
	High child, high adult	*High child, very high adult*	*Very low child, very low adult*	*Low child, low adult*	*High child, high adult*	*Low child, low adult*	*High child, high adult*
	Male/Female	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*
Total population (000s)	147 133/146 945	171 600/173 915	160 494/164 689	213 309/217 623	35 471/35 759	72 156/66 903	174 275/168 301
Total mortality (000s)	2 206/2 050	3 154/3 001	1 342/1 392	1 459/1 120	290/237	409/287	1 750/1 602
Childhood and maternal undernutrition							
Childhood and maternal underweight	438/402	487/441	0/0	14/11	14/11	8/8	223/229
Iron deficiency anaemia	59/67	65/80	2/3	13/13	3/4	3/4	36/44
Vitamin A deficiency	90/112	120/151	0/0	2/3	2/2	0/0	34/53
Zinc deficiency	74/68	128/116	0/0	3/2	5/4	2/2	44/45
Other nutrition-related risk factors and physical inactivity							
High blood pressure	87/128	79/116	179/191	170/162	20/20	76/57	164/171
High cholesterol	34/52	36/53	161/189	88/79	10/9	51/31	114/101
Overweight and obesity (high BMI)	14/19	21/35	135/137	117/144	15/18	36/28	58/67
Low fruit and vegetable consumption	21/31	33/41	92/79	81/58	7/7	27/15	51/48
Physical inactivity	20/25	21/27	74/81	52/55	6/6	21/13	47/43
Addictive substances							
Smoking and oral tobacco use	43/7	84/26	352/294	163/58	5/1	43/10	114/19
Alcohol use	53/15	125/30	27/-22	207/39	22/6	6/1	8/1
Illicit drug use	5/1	1/0	10/7	7/4	1/0	5/1	18/4
Sexual and reproductive health							
Unsafe sex	198/234	805/923	8/8	22/27	17/11	0/4	33/39
Non-use and use of ineffective methods of contraception	NA/16	NA/33	NA/0	NA/5	NA/4	NA/1	NA/23
Environmental risk factors							
Unsafe water, sanitation and hygiene	129/103	207/169	0/1	16/15	13/10	9/9	117/135
Urban air pollution	11/11	5/5	14/14	16/14	3/2	5/3	28/23
Indoor air pollution from household use of solid fuels	93/80	118/101	0/0	7/9	5/5	1/1	56/60
Lead exposure	5/4	4/3	2/1	14/7	2/1	5/2	12/6
Global climate change	9/9	18/18	0/0	0/0	0/0	0/0	10/11
Occupational risk factors							
Risk factors for injury	14/1	18/1	3/0	17/1	2/0	8/0	27/2
Carcinogens	1/0	1/0	7/2	4/1	0/0	1/0	1/0
Airborne particulates	5/2	7/3	12/2	9/1	1/0	1/0	9/2
Ergonomic stressors	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Noise	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Other selected risks factors							
Contaminated injections in health care settings	10/7	27/23	0/0	1/0	1/1	0/0	24/20
Child sexual abuse	0/0	2/1	1/1	1/0	0/0	0/0	1/1

EUROPE Mortality stratum			SOUTH-EAST ASIA Mortality stratum		WESTERN PACIFIC Mortality stratum			
Very low child, very low adult	*Low child, low adult*	*Low child, high adult*	*Low child, low adult*	*High child, high adult*	*Very low child, very low adult*	*Low child, low adult*	*WORLD*	
Male/Female	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Total*
201 514/210 376	108 182/110 277	114 051/129 133	147 173/146 646	639 087/602 719	75 796/78 558	785 055/747 878	3 045 295/2 999 722	6 045 017
2 020/2 054	1 034/916	1 878/1 721	1 234/1 022	6 358/5 764	616/519	5 483/4 944	29 232/26 629	55 861
0/0	9/8	0/0	40/29	573/614	0/0	95/94	1 900/1 848	3 748
2/3	3/3	2/2	15/19	139/185	0/0	34/39	375/466	841
0/0	0/0	0/0	10/13	68/101	0/0	7/9	333/445	778
0/0	2/2	0/0	5/4	132/141	0/0	6/6	400/389	789
325/354	281/289	514/671	133/139	668/519	85/76	711/758	3 491/3 649	7 141
265/282	144/136	387/518	72/40	488/507	39/39	222/265	2 112/2 303	4 415
183/197	117/141	202/265	44/58	42/110	21/20	163/184	1 168/1 423	2 591
95/75	80/67	234/247	55/48	378/311	26/19	269/232	1 449/1 277	2 726
103/103	64/62	147/175	34/34	218/185	23/19	132/134	961/961	1 922
531/145	255/53	548/73	181/12	785/132	128/49	661/137	3 893/1 014	4 907
65/-85	100/25	338/88	51/9	148/21	23/-28	465/66	1 638/166	1 804
11/6	3/1	18/5	13/1	40/8	2/1	28/2	163/41	204
3/9	1/8	3/13	30/25	231/177	0/3	18/36	1 370/1 516	2 886
NA/0	NA/0	NA/0	NA/7	NA/56	NA/0	NA/3	NA/149	149
0/1	8/7	1/1	25/21	326/327	0/0	42/35	895/835	1 730
12/11	20/18	22/24	17/15	72/60	10/8	176/179	411/388	799
0/0	8/9	1/3	15/22	218/304	0/0	137/366	658/961	1 619
4/2	15/8	26/13	6/3	38/19	0/0	21/10	155/79	234
0/0	0/0	0/0	1/0	35/38	0/0	2/1	76/78	154
4/0	5/0	15/1	19/1	79/5	2/0	78/5	291/19	310
12/2	6/1	13/2	3/0	11/1	4/1	28/8	92/17	109
17/2	7/2	15/3	10/3	54/17	4/1	113/54	264/92	356
0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0
0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0
0/0	1/0	6/4	19/9	92/62	0/0	137/58	317/184	501
1/1	1/1	3/2	1/0	16/18	1/1	10/14	38/41	79

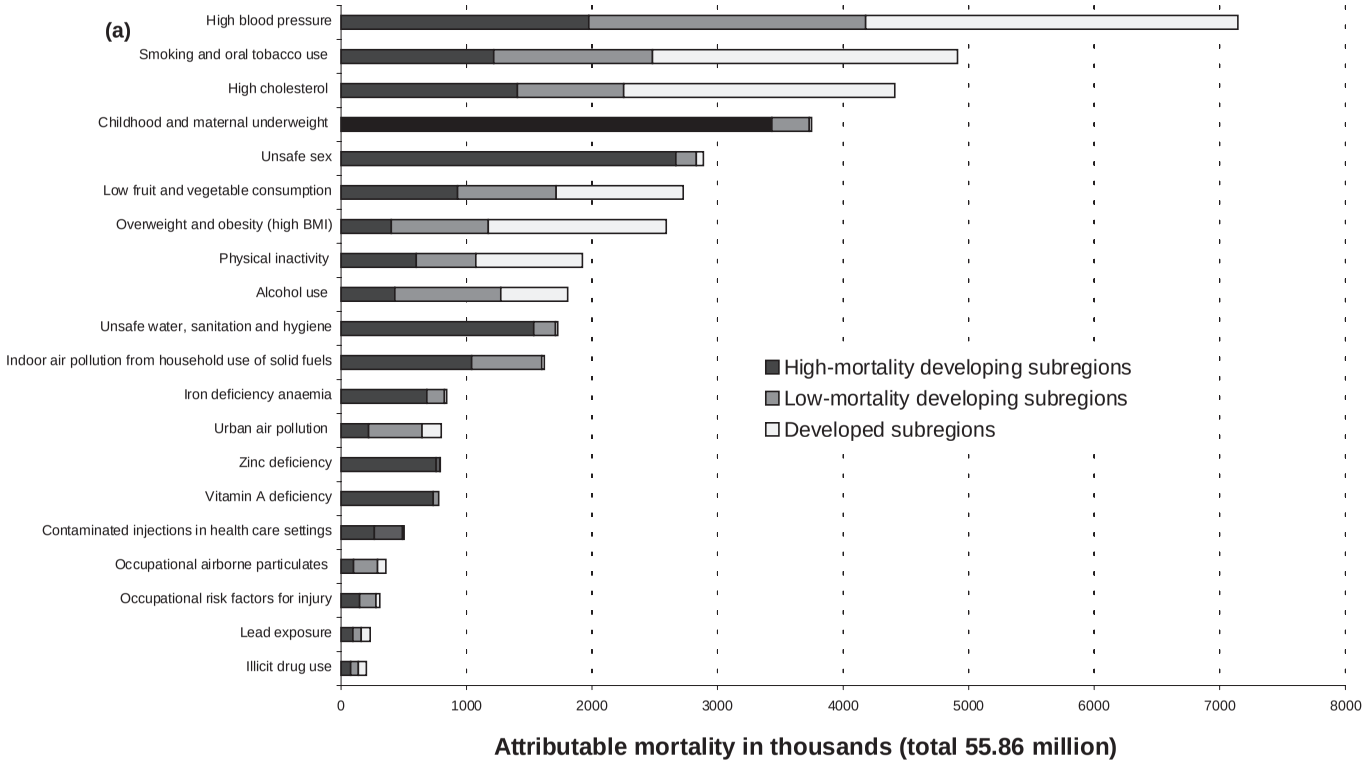
Table 26.1(b) Burden of disease for females and males due to selected risk factors in 14 subregions

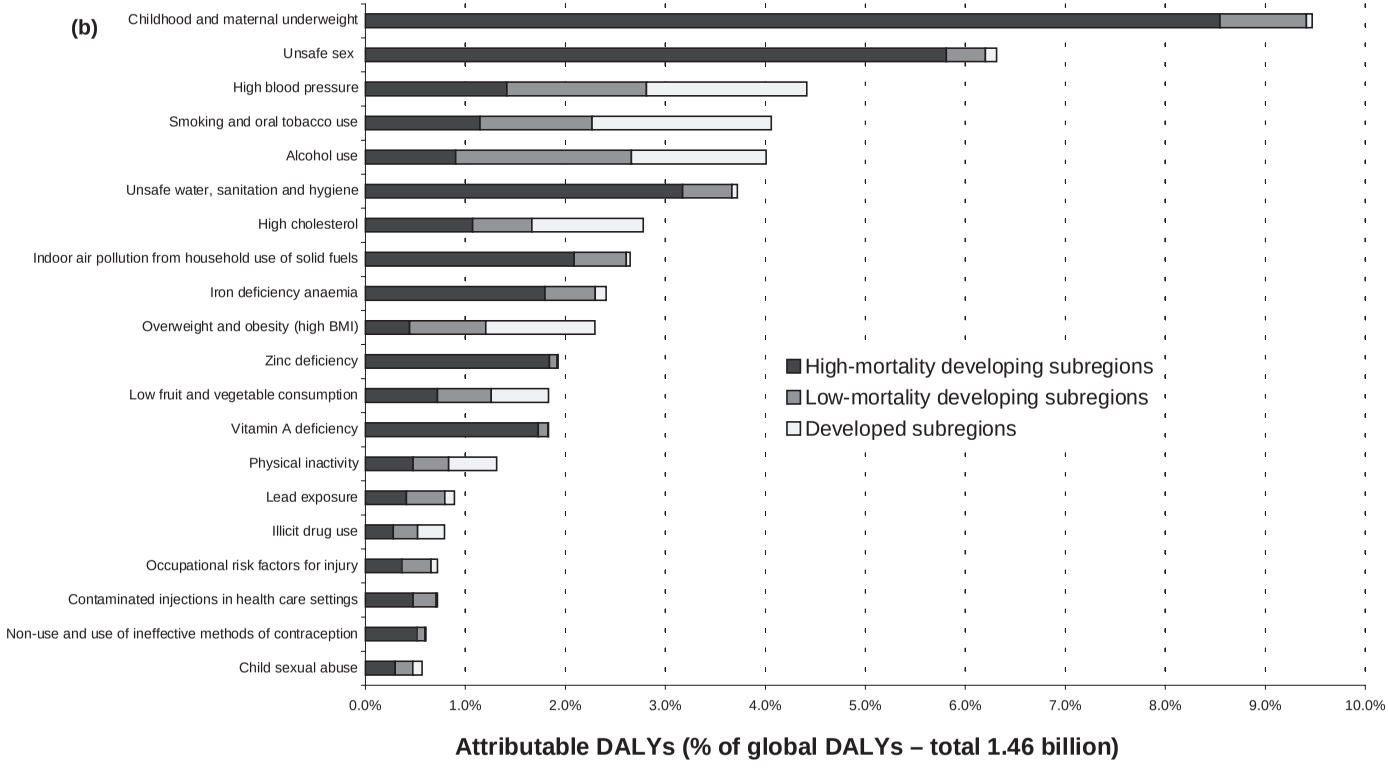
	AFRICA Mortality stratum		*THE AMERICAS* Mortality stratum			*EASTERN MEDITERRANEAN* Mortality stratum	
	High child, high adult	*High child, very high adult*	*Very low child, very low adult*	*Low child, low adult*	*High child, high adult*	*Low child, low adult*	*High child, high adult*
	Male/Female	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*
Total population (000s)	147 133/146 945	171 600/173 915	160 494/164 689	213 309/217 623	35 471/35 759	72 156/66 903	174 275/168 301
Total DALYs (000s)	73 650/70 695	103 191/101 977	24 480/21 804	45 372/35 065	9 158/7 895	12 590/10 131	55 790/54 140
Childhood and maternal undernutrition							
Childhood and maternal underweight	15 530/14 375	17 189/15 710	12/11	570/498	512/410	324/312	8 203/8 407
Iron deficiency anaemia	2 263/2 521	2 451/2 905	223/255	446/465	121/217	239/277	1 449/1 746
Vitamin A deficiency	3 178/3 856	4 208/5 167	0/0	79/103	53/68	9/8	1 159/1 758
Zinc deficiency	2 625/2 414	4 563/4 150	1/1	115/99	174/138	66/63	1 547/1 574
Other nutrition-related risk factors and physical inactivity							
High blood pressure	980/1 295	984/1 177	1 642/1 141	1 807/1 438	208/178	840/570	1 781/1 698
High cholesterol	395/563	456/578	1 451/1 012	1 070/803	109/87	605/320	1 273/1 051
Overweight and obesity (high BMI)	246/318	341/546	1 825/1 654	1 505/1 918	189/234	534/456	882/1 027
Low fruit and vegetable consumption	253/354	434/471	833/536	896/581	72/67	322/172	607/550
Physical inactivity	225/280	262/309	691/576	582/585	61/68	265/164	559/492
Addictive substances							
Smoking and oral tobacco use	591/97	1 311/367	3 567/2 606	2 190/813	51/14	593/197	1 780/379
Alcohol use	1 441/393	3 621/785	2 925/702	7 854/1 443	789/170	162/22	328/36
Illicit drug use	543/156	495/163	808/379	791/310	200/71	449/78	620/153
Sexual and reproductive health							
Unsafe sex	6 205/7 753	24 059/29 664	281/235	843/912	521/310	30/162	1 125/1 508
Non-use and use of ineffective methods of contraception	NA/997	NA/1 732	NA/2	NA/375	NA/203	NA/119	NA/1 210
Environmental risk factors							
Unsafe water, sanitation and hygiene	3 797/3 119	6 365/5 355	31/30	686/603	436/320	314/315	3 797/4 506
Urban air pollution	153/132	80/67	87/65	133/99	24/20	47/30	305/253
Indoor air pollution from household use of solid fuels	3 036/2 358	3 865/3 059	2/4	193/251	175/154	32/32	1 817/1 691
Lead exposure	512/488	460/433	68/49	907/789	140/125	238/187	606/504
Global climate change	321/305	631/636	1/2	35/36	13/10	10/10	357/391
Occupational risk factors							
Risk factors for injury	486/39	583/46	82/6	606/51	80/6	253/18	961/68
Carcinogens	9/2	13/4	56/16	38/8	3/1	12/1	18/2
Airborne particulates	106/37	141/69	184/36	213/44	21/4	37/4	148/39
Ergonomic stressors	21/16	25/20	17/10	32/15	4/2	9/3	25/16
Noise	109/49	127/60	92/31	122/43	15/6	60/21	142/88
Other selected risks factors							
Contaminated injections in health care settings	244/187	804/742	0/0	13/5	20/12	0/0	437/390
Child sexual abuse	49/102	167/238	98/320	147/118	46/27	41/83	85/225

EUROPE Mortality stratum			SOUTH-EAST ASIA Mortality stratum		WESTERN PACIFIC Mortality stratum		WORLD	
Very low child, very low adult	*Low child, low adult*	*Low child, high adult*	*Low child, low adult*	*High child, high adult*	*Very low child, very low adult*	*Low child, low adult*		
Male/Female	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Male/Female*	*Total*
201 514/210 376	108 182/110 277	114 051/129 133	147 173/146 646	639 087/602 719	75 796/78 558	785 055/747 878	3 045 295/2 999 722	6 045 017
28 006/25 314	21 304/17 689	35 099/24 144	33 585/29 302	178 923/177 345	8 780/7 591	131 634/110 818	761 562/693 911	1 455 473
10/9	367/324	32/29	1 634/1 239	21 297/22 766	6/6	4 048/3 972	69 733/68 067	137 801
87/211	166/271	110/161	681/847	5 614/6 883	31/81	1 876/2 462	15 756/19 301	35 057
0/0	1/1	0/0	347/406	2 321/3 368	0/0	241/306	11 596/15 042	26 638
0/0	65/56	5/4	197/152	4 635/4 961	0/0	208/219	14 201/13 833	28 034
2 624/1 828	2 699/2 180	5 386/4 632	1 394/1 402	7 010/5 316	781/451	6 783/6 044	34 920/29 350	64 270
2 062/1 317	1 461/996	4 109/3 211	828/412	5 562/5 528	380/227	2 376/2 195	22 136/18 301	40 437
1 922/1 735	1 420/1 445	2 578/2 684	650/818	686/1 939	334/295	2 430/2 804	15 543/17 872	33 415
785/413	777/511	2 431/1 684	614/524	4 139/3 521	237/118	2 718/2 042	15 117/11 544	26 662
852/654	636/494	1 461/1 236	414/409	2 489/2 186	228/160	1 436/1 318	10 159/8 933	19 092
4 991/1 464	3 381/715	7 230/832	2 712/180	10 474/1 621	994/325	8 313/1 296	48 177/10 904	59 081
3 103/416	2 183/446	7 543/1 570	1 793/284	4 927/675	708/43	12 020/1 941	49 397/8 926	58 323
786/344	181/81	762/223	406/121	1 386/282	231/101	1 110/259	8 769/2 719	11 488
114/202	50/240	134/295	1 009/925	7 413/6 004	12/65	804/995	42 600/49 269	91 869
NA/3	NA/83	NA/47	NA/397	NA/3 354	NA/1	NA/290	NA/8 814	8 814
33/33	287/262	64/57	734/506	8 762/9 725	14/13	2 112/1 879	27 432/26 726	54 158
73/44	170/118	191/129	154/128	718/594	53/31	1 343/1 161	3 533/2 871	6 404
0/0	233/244	18/49	458/532	6 641/7 596	0/0	2 569/3 528	19 040/19 499	38 539
75/43	304/189	424/211	379/337	1 489/1 198	15/10	1 496/1 251	7 112/5 814	12 926
1/2	5/5	2/2	19/15	1 213/1 325	0/1	92/77	2 700/2 816	5 517
130/12	203/15	410/31	577/39	2 857/184	56/5	2 495/199	9 779/718	10 496
95/13	63/7	129/16	35/5	119/12	24/4	227/87	891/179	1 070
216/43	105/32	167/43	135/47	862/315	68/18	1 726/493	4 130/1 224	5 354
21/11	18/12	21/14	26/19	111/78	9/5	146/110	485/333	818
117/47	92/50	136/92	219/185	799/303	26/22	735/365	2 788/1 362	4 151
0/0	8/5	106/59	356/156	2 341/1 759	0/0	2 028/791	6 356/4 105	10 461
61/175	72/158	132/205	42/56	1 079/2 340	29/96	888/1 158	2 934/5 302	8 235

Note: The table shows the estimated mortality and disease burden for each risk factor considered individually. These risks act in part through other risks and act jointly with other risks. Consequently, the burden due to groups of risk factors will usually be less than the sum of individual risks (see chapter 27).

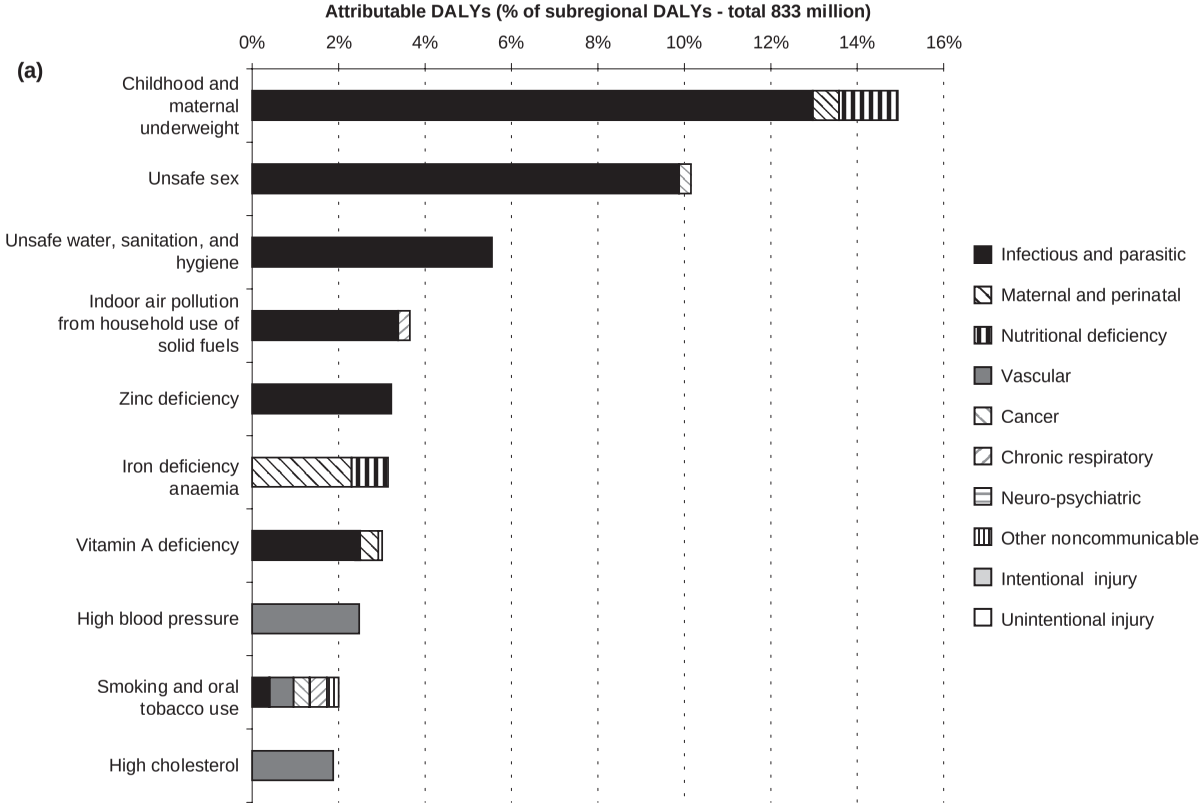
Figure 26.1 (a) Mortality and (b) burden of disease due to leading global risk factors

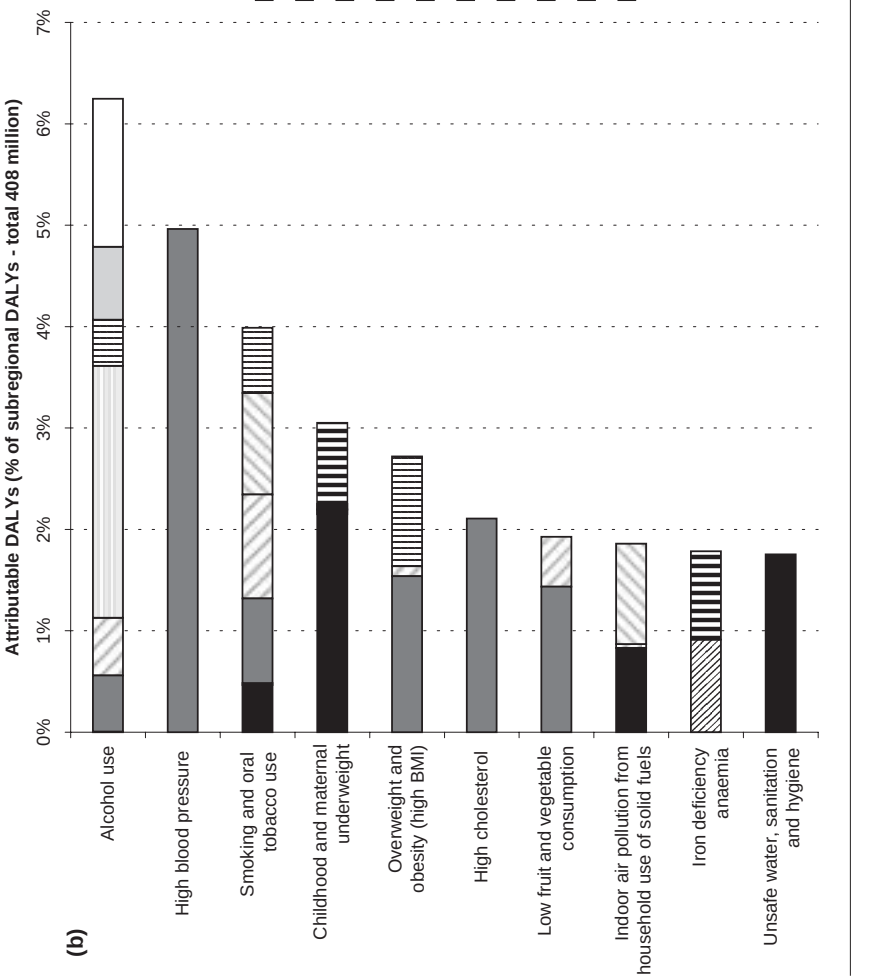




Note: High-mortality developing: AFR, AMR-D, EMR-D and SEAR-D subregions; low-mortality developing: AMR-B, EMR-B, SEAR-B and WPR-B; developed: AMR-A, EUR and WPR-A. The figure shows the estimated mortality and disease burden for each risk factor considered individually. These risks act in part through other risks and act jointly with other risks. Consequently, the burden due to groups of risk factors will usually be less than the sum of individual risks (see chapter 27).

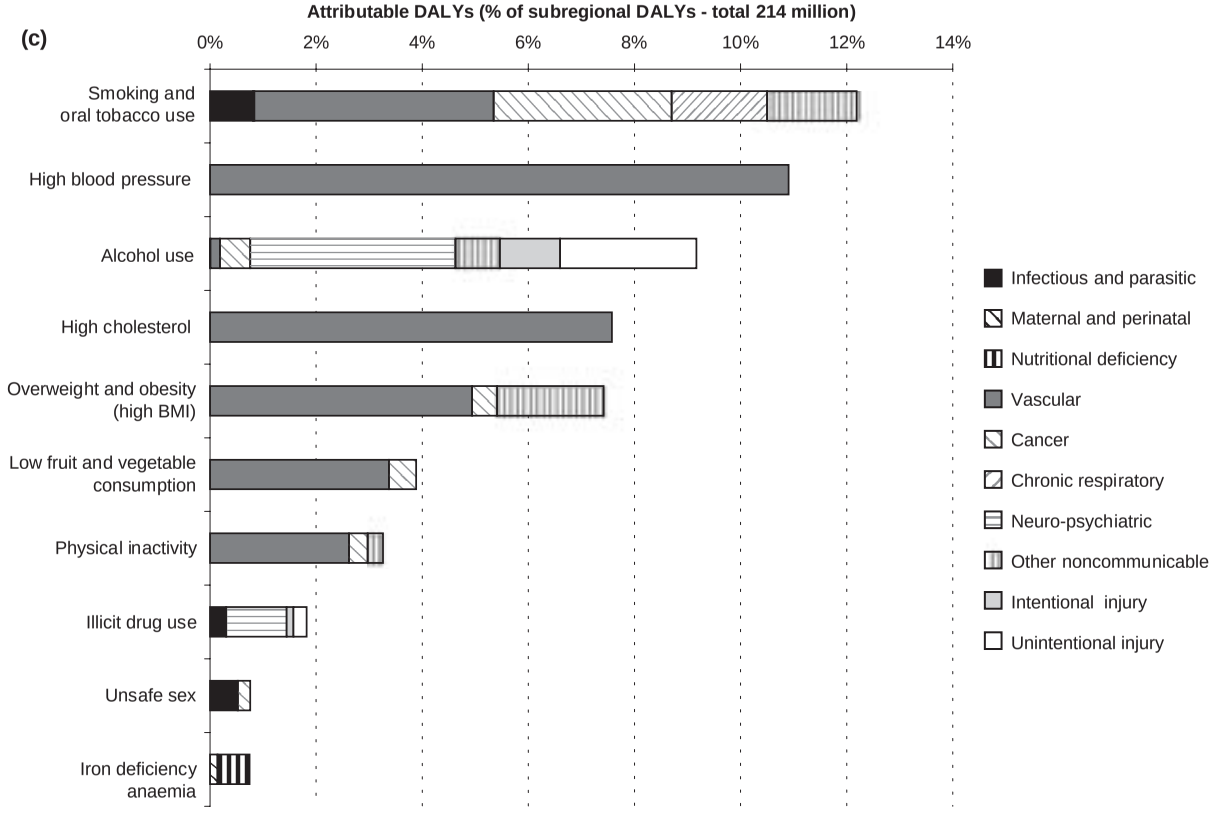
Figure 26.2 Burden of disease due to leading regional risk factors divided by disease type in (a) high-mortality developing, (b) low-mortality developing and (c) developed subregions




(b)
Attributable DALYs (% of subregional DALYs - total 408 million)
0%
1%
2%
3%
4%
5%
6%
7%
Alcohol use
High blood pressure
Smoking and oral tobacco use
Childhood and maternal underweight
Overweight and obesity (high BMI)
High cholesterol
Low fruit and vegetable consumption
Indoor air pollution from household use of solid fuels
Iron deficiency anaemia
Unsafe water, sanitation and hygiene
Infectious and parasitic
Maternal and perinatal
Nutritional deficiency
Vascular
Cancer
Chronic respiratory
Neuro-psychiatric
Other noncommunicable
Intentional injury
Unintentional injury

continued

Figure 26.2 Burden of disease due to leading regional risk factors divided by disease type in (a) high-mortalitydeveloping, (b) low-mortality developing and (c) developed subregions (*continued*)



contribution to disease burden) and the non-fatal effects (e.g. neuropsychological outcomes of alcohol).

Despite disaggregation into underweight and micronutrient deficiency (which are not additive; see chapter 27) and methodological changes, undernutrition has remained the single leading global cause of health loss with comparable contributions in 1990 (220 million DALYs, 16%, for malnutrition) (Murray and Lopez 1997) and 2000 (140 million DALYs, 9.5%, for underweight; 2.4%, 1.8%, 1.9% for iron, vitamin A and zinc deficiency respectively; 0.1% for iodine deficiency disorders). This is because while prevalence of underweight has decreased in most regions of the world in the past decade, it has increased in sub-Saharan Africa (de Onis et al. 2000) where its effects are disproportionately large due to simultaneous exposure to other childhood disease risk factors. A substantial part of the decrease in the burden of disease due to poor water, sanitation and hygiene (from 6.8% in 1990 to 3.7% in 2000) is due to a decline in global diarrhoeal disease mortality (from 2.9 million deaths in 1990 to 2.1 million in 2000), and partly a result of improved case management interventions, particularly oral rehydration therapy.

Leading causes of burden of disease in all high-mortality developing subregions were childhood and maternal undernutrition—including underweight (14.9%) and micronutrient deficiencies (3.1% for iron deficiency, 3.0% for vitamin A deficiency and 3.2% for zinc deficiency)—unsafe sex (10.2%), poor water, sanitation and hygiene (5.5%) and indoor smoke from solid fuels (3.6%). The relative contribution of unsafe sex was disproportionately larger (26%) in AFR-E, where HIV/AIDS prevalence is the highest, making it the leading cause of burden of disease in this subregion. The outcomes of these risk factors were mostly communicable, maternal, perinatal and nutritional conditions (Figure 26.2) which dominate the disease burden in high-mortality developing subregions. Despite the very large contribution of these diseases and their underlying risk factors, tobacco, blood pressure and cholesterol already resulted in significant loss of healthy life years in these subregions. For example, in SEAR-D (dominated by India in terms of population) the burden of disease attributable to tobacco, blood pressure and cholesterol was already of comparable magnitude to micronutrient deficiencies and is only marginally smaller than indoor smoke from solid fuels and poor water, sanitation and hygiene. In addition to their relative magnitude, the absolute size of the loss of healthy life years attributed to risk factors in high-mortality developing subregions was substantial. Childhood and maternal underweight and unsafe sex in these subregions alone (with 38% of global population) contributed as much (>200 million DALYs) to loss of healthy life as all diseases and injuries in developed countries (with 22% of global population).

Across developed subregions, tobacco (12.2%), high blood pressure (10.9%), alcohol (9.2%), high cholesterol (7.6%) and high BMI (7.4%) were consistently the leading causes of loss of healthy life, contributing

mainly to noncommunicable diseases and injuries. Tobacco was the leading cause of disease burden in all developed subregions, except EUR-C (dominated by Russia) where high blood pressure and alcohol resulted in slightly larger loss of healthy life. The increase in the disease burden due to blood pressure compared to 1990 (Murray and Lopez 1997) (from 3.9% in the established market economies and 5.9% in the formerly socialist economies) mainly reflects new evidence on hazard size after correction for regression dilution bias (MacMahon et al. 1990). The contributions of these risk factors are consistently larger than those of leading *diseases* of the developed subregions (i.e. ischaemic heart disease [9.4%], unipolar depressive disorders [7.2%], cerebrovascular disease [6.0%], etc.), which emphasizes the potential health gains from reducing risk factors.

The low-mortality developing subregions present possibly the most striking mixture of leading risk factors. The leading risk factors in these subregions (40% of global population) include those from both developed and high-mortality developing subregions with comparable magnitudes (e.g. underweight [3.1%] and high BMI [2.7%] had comparable contributions to the burden of disease. See also Monteiro et al. 2002). In addition, the decline in the share of burden of disease due to the risk factors in low-mortality developing subregions was less marked than that in high-mortality developing and developed subregions (e.g. the ratio of 1st to 10th leading risk factors was smaller). This lower clustering of risk factor burden further emphasizes the role of a more extended and mixed group of risk factors in low-mortality developing subregions. Alcohol was the leading cause of burden of disease in low-mortality developing subregions as a whole (6.2%) and in AMR-B and WPR-B, but made a relatively low contribution to the burden of disease in EMR-B. In general, AMR-B and EMR-B had risk factor profiles similar to the developed subregions (tobacco, blood pressure, cholesterol, BMI and alcohol), while SEAR-B and WPR-B had a more mixed risk factor profile (with the leading five risks being underweight, blood pressure, tobacco, unsafe sex and alcohol in SEAR-B; alcohol, blood pressure, tobacco, underweight and indoor smoke from solid fuels in WPR-B).

An important finding of this analysis is the key role of nutrition in health worldwide. Approximately 13% of the global disease burden can be attributed to the joint effects of childhood and maternal underweight or micronutrient deficiencies. In addition, almost as much as 7% (16% for those aged 30 years and above) can be attributed to risk factors that have substantial dietary determinants—high blood pressure, high cholesterol, high BMI and low fruit and vegetable intake. These patterns are not uniform within subregions, however, and in some countries the transition has been healthier than in others (Lee et al. 2000; Popkin et al. 2001). Further, the major nutritional risk factors show interregional heterogeneity (e.g. the relative contributions of blood pressure, cholesterol and BMI were different in AMR-A, SEAR-D and WPR-B). This het-

erogeneity further illustrates the importance of concurrent and comparable quantification of distal and proximal risk factors to provide a more complete picture of the role of various distal and proximal risk factors in reducing disease.

This analysis also provides the first quantitative evidence of the public health consequences of a number of risk factors including indoor smoke from solid fuels (2.6% of global disease burden), high BMI (2.3%) and zinc deficiency (1.9%). On the other hand, the burden of disease due to some risks (e.g. physical inactivity) was lower than expected if the methodology and results from the limited number of industrialized countries had been extrapolated (Powell and Blair 1994). This is partially because of difficulties in measuring exposure to this risk factor. A categorical exposure variable with a conservative baseline of "sufficient" (vs vigorous) activity was used. In part, it also reflects the inclusion of occupational and transportation domains of activity (that are common among rural populations of developing countries) in this analysis, above and beyond leisure-time activity which is more relevant to developed countries and urban populations (Jacobs et al. 1993; Levine et al. 2001).

2. Distributions of risk factor-attributable disease burden

An important feature of risk assessment, with implications for broad prevention policies and specific interventions and programmes, is the distribution of disease burden among population subgroups. These subgroups may be defined by factors such as age, sex, socioeconomic status or the current level of exposure to a risk factor, if exposures are defined in multiple categories or continuously. For example, reducing the large disease burden due to road traffic accidents among young adult males, largely associated with binge alcohol consumption, would require designing interventions that focus on this population subgroup and their specific drinking behaviours. On the other hand, the majority of effects from risk factors such as blood pressure have been found to occur among those at moderately elevated levels, suggesting the need for interventions beyond those intended for clinical hypertension (Cook et al. 1995; Murray et al. 2003; Rodgers et al. 2000). While the distribution of health effects by age and by exposure level has been studied in specific cohorts and for specific risk factors (Peto et al. 1992; Rodgers and MacMahon 1999; Rose 1992), there are no such estimates at the global level and for multiple risks.

The distributions of mortality and disease burden attributable to the risk factors included in this book by age and sex is shown in Table 26.2. The estimated disease burden from childhood and maternal undernutrition, unsafe water, sanitation and hygiene, and global climate change (much of whose estimated effects are mediated through nutritional and water variables) was almost exclusively among children aged <5 years.

Table 26.2 The distribution of risk factor-attributable mortality and burden of disease by age and sex

	Mortality (%)						*Disease burden (%)*					
	0–4	*5–14*	*15–59*	*≥60*	*Males*	*Females*	*0–4*	*5–14*	*15–59*	*≥60*	*Males*	*Females*
Childhood and maternal undernutrition												
Childhood and maternal underweight	100	0	0	0	51	49	100	0	0	0	51	49
Iron deficiency anaemia	72	1	22	4	45	55	62	6	30	2	45	55
Vitamin A deficiency	85	1	14	0	43	57	86	1	12	0	44	56
Zinc deficiency	100	0	0	0	51	49	100	0	0	0	51	49
Other nutrition-related risk factors and physical inactivity												
High blood pressure	0	0	19	81	49	51	0	0	43	57	54	46
High cholesterol	0	0	22	78	48	52	0	0	50	50	55	45
Overweight and obesity (high BMI)	0	0	26	74	45	55	0	0	57	43	47	53
Low fruit and vegetable consumption	0	0	23	77	53	47	0	0	49	51	57	43
Physical inactivity	0	0	21	79	50	50	0	0	48	52	53	47
Addictive substances												
Smoking and oral tobacco use	0	0	30	70	79	21	0	0	61	39	82	18
Alcohol use	1	1	65	33	91	9	1	3	87	9	85	15
Illicit drugs use	0	0	100	0	80	20	0	2	98	0	77	23
Sexual and reproductive health												
Unsafe sex	16	1	77	6	47	53	18	1	79	2	46	54
Non-use and use of ineffective methods of contraception	0	0	100	0	0	100	0	0	100	0	0	100
Environmental risk factors												
Unsafe water, sanitation and hygiene	68	5	13	14	52	48	77	8	13	3	51	49
Urban air pollution	3	0	16	81	51	49	12	0	40	49	56	44
Indoor air pollution from household use of solid fuels	56	0	5	38	41	59	83	0	8	9	49	51
Lead exposure	0	0	41	57	66	34	75	0	16	8	55	45
Global climate change	86	3	6	5	49	51	88	5	6	1	49	51
Selected occupational risk factors												
Risk factors for injuries	0	0	85	14	94	6	0	0	95	5	93	7
Carcinogens	0	0	28	72	85	15	0	0	51	49	83	17
Airborne particulates	0	0	17	83	74	26	0	0	65	35	77	23
Ergonomic stressors	0	0	0	0	0	0	0	0	95	5	59	41
Noise	0	0	0	0	0	0	0	0	89	11	67	33
Other selected risk factors												
Contaminated injections in health care settings	10	2	53	35	63	37	16	3	67	13	61	39
Child sexual abuse	0	0	81	22	48	52	0	0	96	4	36	64

For these risks, more than 85% of the total attributable burden occurred in this age group, with the exception of iron deficiency where 30% of burden was borne by women of childbearing age. The disease burden from other diet-related risks, tobacco and occupational risks (except injuries and back pain) was almost equally distributed among adults above and below the age of 60 years. For example, 43% and 61% of disease burden due to high blood pressure and tobacco respectively, occurred from adverse events in the 15–59-year age group.

More than 90% of disease burden attributable to lack of contraception, illicit drugs, occupational ergonomic stressors and risk factors for injury and child sexual abuse occurred in adults below the age of 60 years. About three-quarters (77–80%) of disease burden for alcohol, unsafe sex and contaminated injections in health care settings occurred between the ages of 15 and 59 years. Most of the risks whose burden is concentrated in younger adults are those with outcomes that include HIV/AIDS, maternal conditions, neuropsychiatric diseases and injuries. Moreover, with the exception of alcohol, which has a global presence, the majority of disease burden from these risks is concentrated in developing countries (Figures 26.1 and 26.2). This illustrates the large, and at times neglected disease burden from risks that affect young adults in developing countries, with important consequences for economic development.

Only a small fraction of disease burden from the risk factors considered occurred among 5–14-year olds. This was because some of the leading causes of ill-health of this age group (e.g. motor vehicle accidents and other injuries, depression) have complex causes that could not easily be included in the current risk-based framework. For other leading diseases at these ages (e.g. diarrhoea and lower respiratory infections), most epidemiological studies have focused on children aged <5 years and do not provide hazard estimates for older children.

The disease burden attributable to underweight and micronutrient deficiencies in children was equally distributed among males and females, but the total all-age disease burden from iron and vitamin A deficiencies was slightly greater in females due to effects on maternal conditions. Other diet-related risks, physical inactivity, environmental risks and unsafe sex contributed almost equally to disease burden in males and females. Approximately 80% of disease burden from addictive substances and 60–90% from various occupational risks occurred among men. The former reflects the social, behavioural and economic forces that have so far made addictive substances more widely used by men, especially in developing countries. The latter was partially due to the inclusion of formal employment only and partially because men tend to make up most of the workforce engaged in heavy industrial jobs and formal agriculture. Women suffered an estimated two-thirds of disease burden from childhood sexual abuse and the entire burden caused by non-use and use of ineffective methods of contraception, as defined in chapter 15.

The distributions of disease burden attributable to risk factors by exposure levels are shown in Table 26.3 for those risks quantified using categorical variables, and in Figure 26.3 for those with continuous variables. For most of these risks a substantial proportion of attributable burden occurred among those with modest elevation of risk. For example, only 35% of the disease burden from underweight, the leading global risk, occurred in severely underweight children (<–3 SD from referent group median); the rest was among those in the 1–3 SD below the median range. The large majority of the burden of disease from unsafe water, sanitation and hygiene was approximately equally distributed among three of the five exposure scenarios. This reflects the fact that the exposure categories were defined as the presence of water and sanitation technology-based interventions, and during decades of water and sanitation projects, many countries have "clustered" in a limited number of technology groups. However, there is likely to be large heterogeneity of exposure within each scenario (Curtis et al. 2000).

Figure 26.3(a) shows the distribution of the estimated cardiovascular (CVD) burden of disease (in DALYs) attributable to four major continuous risk factors, by exposure levels. Half the attributable burden occurs to the left of the solid vertical line and half occurs to the right. The dashed vertical lines indicate commonly used thresholds—140 or 160 mmHg for hypertension, 6.5 mmol/l for hypercholesterolaemia, and 30 kg/m^2 for obesity (the cut-off for hypertension is shown at 150 mmHg, between the two commonly used values of 140 and 160 mmHg). Figure 26.3(b) shows the cumulative percentage of attributable burden by exposure levels. In reality, a modest rightward skew (not modelled here) of distributions would lead to slightly more events occurring in those who were hypertensive, hypercholesterolaemic or obese.

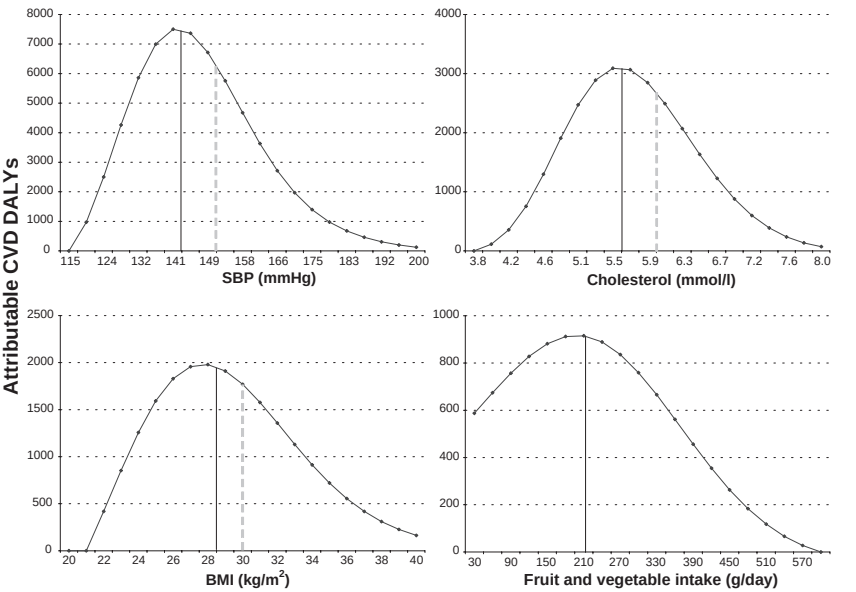
Figure 26.3 shows that a substantial proportion of the disease burden attributable to high blood pressure, cholesterol and body mass index (BMI) and inadequate fruit and vegetable intake occurred in the "mid-range" exposures. For example, the 2nd and 3rd quartiles (i.e. half of attributable burden) occurred between SBP of approximately 130 and 150 mmHg, cholesterol of 5.0 and 6.1 mmol/l and BMI of 25–32 kg/m^2 and fruit and vegetable intake of 150–300 g/day. This was similar to or greater than the amount of burden occurring among individuals with risk factor levels above the commonly used (but arbitrary) thresholds of hypertension, hypercholesterolaemia and obesity, as shown in Figure 26.3.

The distribution by levels of demographic and economic development suggest that the burden of disease due to risks such as undernutrition and unsafe water, sanitation and hygiene occurred virtually entirely in the high-mortality developing subregions of the world, whereas other risks, such as tobacco, alcohol and dietary risks had global effects (Figure 26.1). Categorization based on economic and demographic development was also a key modifier of the age distribution patterns. Most notably,

Table 26.3 Distribution by exposure level of attributable burden due to selected categorical risk factors

Risk factor	*Referent category*	*Exposure categories*				
Childhood and maternal underweight	Same fraction of children <–1 SD weight-for-age as the international reference group	<–1 to <–2 SD below the international reference group median	<–2 to <–3 SD below the international reference group median	<–3 SD below the international reference group median		
Proportion of total attributable disease burden	0	0.20	0.46	0.35		
Physical inactivity	All having at least 2.5 hours per week of moderate-intensity activity or equivalent (400 kJ/week)	Some but less than 2.5 hours per week of moderate-intensity activity	Little or no physical activity			
Proportion of total attributable disease burden	0	0.49	0.51			
Unsafe water, sanitation and hygiene	Absence of transmission of diarrhoeal disease through water, sanitation and hygiene	Regulated water supply and full sanitation coverage, with partial treatment for sewage	Improved water supply, basic sanitation, improved access to drinking water, improved personal hygiene and water disinfected at point of use	Improved water supply and basic sanitation	Basic sanitation but no improved water supply	No improved water supply and no basic sanitation
Proportion of total attributable disease burden	0	0	0.39	0.03	0.28	0.30
Child sexual abuse	No sexual abuse	Non-contact abuse	Contact abuse	Intercourse		
Proportion of total attributable disease burden	0	0.08	0.44	0.48		

Figure 26.3 Distribution by exposure level of cardiovascular disease (CVD) burden attributable to selected continuous risk factors



Note: For blood pressure and cholesterol, the plots represent the estimated usual levels (MacMahon et al. 1990), which tend to be closer to population means than levels based on one-off measurements commonly used in population surveys. For example, the distribution of usual blood pressure is approximately half as wide as the distribution of one-off blood pressure measures and so less people would be classified as hypertensive if classifications were based on usual rather than one-off blood pressure. Thus, with a population mean systolic blood pressure (SBP) of 134 mmHg, the SD of one-off measures might be 17 mmHg (with about 18% of the population having one-off SBP over 150 mmHg) and the SD of usual SBP 9 mmHg (hence about 5% of the population would have usual SBP over 150 mmHg).

disease burden due to many major risks for chronic diseases occurred in younger ages in developing regions compared to developed regions. For example, in high-mortality developing subregions, 69% of disease burden attributable to tobacco occurred in people aged 15–59 years, whereas this share was 63% for low-mortality developing subregions and 55% for developed subregions. The different age structures across major world regions, together with exposure differences, resulted in different distributions of attributable burden by region. For example, Figure 26.4 shows that disease burden attributable to elevated blood pressure, cholesterol and BMI occurred at lower levels in developing regions compared to developed regions, mainly because of lower age-specific exposure levels in those populations (see chapters 6–8).

Figure 26.4 Distribution by exposure level of attributable cardiovascular disease (CVD) burden due to selected risk factors, by age and subregional grouping (see Figure 26.3 for details)

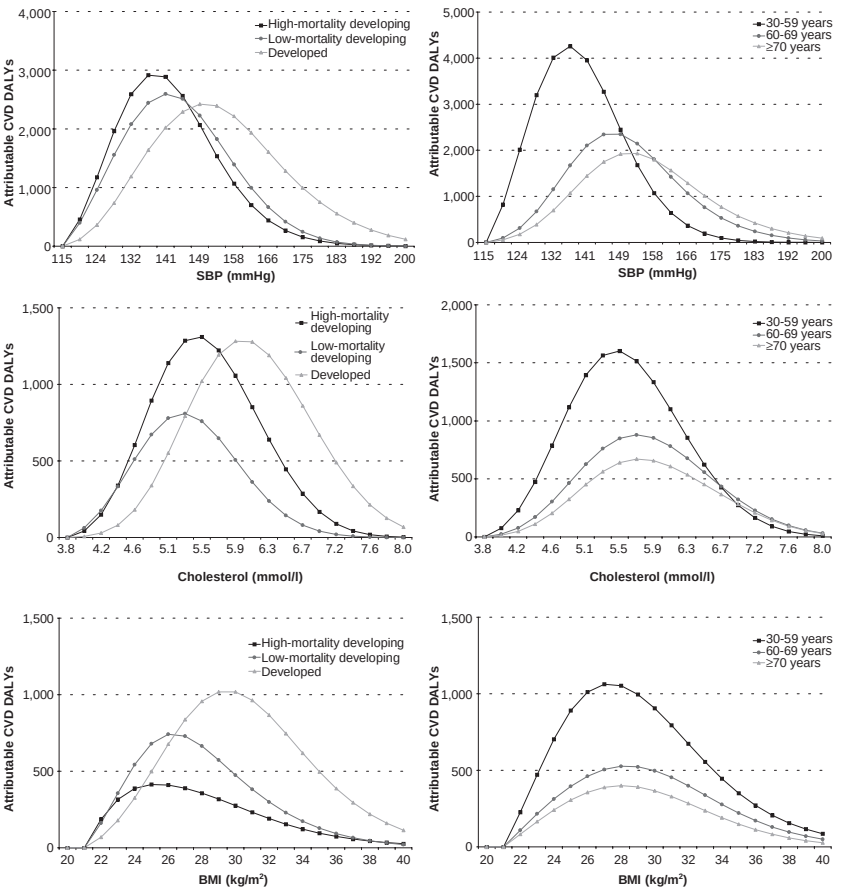


Figure 26.4 also shows that skewness of the distribution of disease burden was not substantially different across different age groups for BMI. This is because the comparatively larger relative risk per unit-BMI at younger ages (which leads to more right-hand skew) is counterbalanced by the comparatively lower BMI at younger ages (which leads to left-hand skew). This is in contrast to blood pressure, for which disease burden in younger age groups occurred at lower exposures because the age patterns of exposure and relative risk do not entirely compensate.

3. Sources of uncertainty

Broad sources of uncertainty in risk assessment were discussed in chapter 1 of this book. Uncertainty about disease causation (Evans 1978; Hill 1965) in practice was secondary to uncertainty about hazard size, because when causality was uncertain, estimates of hazard needed for risk assessment were also unknown or uncertain. For example, while there is uncertainty about whether climate change would increase incidence of certain diseases, or whether the relationships between occupational factors or physical inactivity and lower back pain are causal, in each case quantitative risk assessment would also require estimates of hazard magnitude. The collectivity of scientific knowledge from disciplines such as behavioural science, vector biology, physiology, biomechanics and epidemiology would confirm the possibility of a causal relationship in the above cases, but would shift the debate to hazard size. As a result, for some risk factors, only the contribution to a subset of disease outcomes could be quantified because epidemiological studies did not provide enough information for hazard quantification for all risk factor–disease pairs, even when the causal relationships were believed or suspected.

Estimates of hazard size in individual studies were as far as possible adjusted for confounding. Extrapolation of hazard from a limited number of studies to other populations on the other hand has received less attention. While the robustness of relative risk measures has been confirmed for more proximal factors in studies across populations (Eastern Stroke and Coronary Heart Disease Collaborative Research Group 1998; Horton 2000; Law et al. 1994), hazard extrapolation is an important source of uncertainty for more distal risks (e.g. child sexual abuse) or those whose effects are heterogeneous (e.g. alcohol and injuries vs alcohol and cancer).

Direct exposure data for many risk factors were limited due to difficulties both in their measurement and under-investment in risk factor surveillance, especially in developing countries. To allow maximum use of available data, such risk factors were represented using indirect or aggregate indicators (e.g. smoking impact ratio (*SIR*) for accumulated hazards of smoking, weight-for-age for childhood undernutrition and use of solid fuels for indoor air pollution). For some risks, multiple data sources allowed limiting the possible range of exposure estimates. For example, in the absence of alcohol surveys, total alcohol production, trade and unrecorded consumption provided upper bounds on the fraction of population that would be in the highest consumption category. Finally, some of the risk factors in this analysis were represented using continuous exposure variables (e.g. high blood pressure). Others have used categorical variables (e.g. indoor smoke from solid fuels, underweight and physical inactivity) even though the health effects occur along a continuum. This choice reflected the availability of exposure data and

hazard estimates for categories. In such cases, the contribution to disease within the categories may be under-estimated.

The findings of this work should, therefore, be considered within the context of limited available data and subject to uncertainty. This uncertainty varies across risk factors and geographical regions. Further discussion of sources and quantification of uncertainty has been provided in individual risk factor chapters.

4. Discussion

Despite inherent uncertainties, the quantification of the burden of disease attributable to selected risk factors illustrates that the loss of health in the world is dominated by those risk factors that affect the poorest regions and populations, such as undernutrition, poor water, sanitation and hygiene and indoor smoke from solid fuels. Coupled with these are hazards such as alcohol, tobacco, high blood pressure and high cholesterol that in the year 2000—even compared to a decade earlier (Murray and Lopez 1997)—are widespread or are estimated to have large health impacts. Nowhere is this picture more apparent than in the low-mortality developing regions, which account for 48% of global population, and are affected by both groups of risk factors.

Comparing the burden attributable to risk factors across the three groupings of countries in this work (Figures 26.1 and 26.2) provides a cross-sectional picture of a "risk factor transition" in which the relative contribution of adult or noncommunicable disease risk factors increases as childhood and communicable disease risk factors decrease with economic development. Analysis of previous development-based transitions, such as changes in inequality or environment with economic development, has demonstrated the role of policy in inducing or delaying, and shaping the dynamics of the transition (Bowman 1997). Examples in public health include rapid control of vector-borne diseases (Chitsulo et al. 2000), high maternal mortality where contraception and abortions are not accessible for non-economic reasons, and potential HIV epidemics in some developed countries (MacLehose et al. 2002). At the same time, at least some risk factor transitions are confirmed by the increasing role of hazards such as tobacco and obesity *over time* (Ebbeling et al. 2002; Pelletier 1998; WHO 1997). The increase in the global burden of disease due to tobacco from 2.6% in 1990 to 4.1% in 2000, while partially due to new evidence on hazard size after correction for confounding (Thun et al. 2000), mostly reflects the increased accumulated hazards, and is most noticeable in developing countries. The cross-sectional comparison demonstrates that risk factors such as alcohol and high blood pressure and cholesterol, if not increasing in absolute terms (Popkin 2002; Reddy and Yusuf 1998), are important contributors to loss of health in all regions.

The large remaining burden from childhood disease and mortality risks such as undernutrition, poor water and sanitation, and indoor smoke from solid fuels shows the continued need for developing and delivering effective interventions. At the same time, four of the five leading causes of lost healthy life affect adults (Figure 26.1). Risk factors for both adult communicable and noncommunicable diseases already make substantial contributions even in regions with low income and high infant mortality. It is imperative therefore that health programmes and policy continually reassess the appropriate balance between interventions addressing childhood disease risk factors and those that affect adult health. Dynamic and systematic policy responses can mitigate the spread of such risk factors and their more distal causes to a large extent throughout the development process, such as a healthier nutritional or environmental transitions (Arrow et al. 1995; Lee et al. 2000). Also, as illustrated by the persistence of diseases such as malaria or the large increase in the disease burden due to HIV/AIDS and its risk factors since 1990 (e.g. unsafe sex from 3.5% to 6.3%), as well as the potential for generalized HIV/AIDS epidemics in some eastern European countries (MacLehose et al. 2002) or China (Kaufman and Jing 2002), important communicable disease risk factors also need dynamic monitoring and policy responses.

There are a number of reasons why risk factors that were not among the leading global causes of disease burden should not be neglected. Most obviously, this analysis could be expanded with other risk factors that are both prevalent and hazardous. Second, although smaller than other risks, many such risk factors make non-negligible contributions to burden of disease in specific populations. For example in WPR-B (dominated by China in terms of population), where there is considerable industrial activity based on coal, ambient air pollution and lead exposure have health effects comparable to poor water, sanitation and hygiene and some micronutrient deficiencies. Similarly, lack of contraception was among the 10 leading risk factors for female burden of disease in a number of subregions.

Some risk factors with comparatively low global disease burden are highly concentrated among sectors of society (e.g. occupational exposures among mine workers) and have implications for health inequalities. This concentration may also imply that risks can be targeted more easily. For other risk factors, such as child sexual abuse, ethical considerations may outweigh direct contributions to disease burden in policy debate. Finally, while the burden of disease due to a risk factor may be comparatively small, effective or cost-effective interventions may be available. Examples include reducing the number of unnecessary medical injections coupled with the use of sterile syringes and reduction in exposure to lead or ambient air pollution in industrialized countries in the second half of the 20th century which often also led to benefits such as energy saving.

Beyond their total magnitude, this study has also provided a picture of the distribution of risk factor-attributable disease burden by age, sex and exposure. Analysis in multiple age and exposure categories, or along a continuum of exposures, suggests that globally a considerable proportion of the disease burden attributable to many major risk factors occurred among those with only moderately raised levels, not the extremes (e.g. hypertension, obesity or severe malnutrition).

For acute exposures and outcomes, the underlying relationship is more complex. For example, while in many societies the majority of alcohol-attributable injury (e.g. traffic accidents) arises among people who on average drink moderately (Kreitman 1986), these people would be at the more extreme end of the distribution in a different dimension: volume of drinking *before* the injury. This finding suggests that the shapes of both exposure distributions and risk relationships are important determinants of the distribution of disease burden. If exposure to risk factors is clustered or the risk relationship does not follow a linear pattern, high exposure groups may indeed play a disproportionately important role (Lemmens 2001; Skog 1999). Further implications of these findings for research, and for policies and programmes aimed at improving population health, are discussed in chapter 29.

Note

1 See preface for an explanation of this term.

References

Arrow K, Bolin B, Costanza R et al. (1995) Economic growth, carrying capacity, and the environment. *Science*, **168**:520–521.

Bowman KS (1997) Should the Kuznets effect be relied on to induce equalizing growth: evidence from post-1950 development. *World Development*, **25**:127–143.

Chitsulo L, Engels D, Montresor A, Savioli L (2000) The global status of schistosomiasis and its control. *Acta Tropica*, **77**:41–51.

Cook NR, Cohen J, Hebert P, Taylor JO, Hennekens CH (1995) Implications of small reductions in diastolic blood pressure for primary prevention. *Archives of Internal Medicine*, **155**:701–709.

Curtis V, Cairncross S, Yonli R (2000) Domestic hygiene and diarrhoea—pinpointing the problem. *Tropical Medicine and International Health*, **5**:22–32.

de Onis M, Frongillo E, Blössner M (2000) Is malnutrition declining? An analysis of changes in levels of child malnutrition since 1980. *Bulletin of the World Health Organization*, **78**:1222–1233.

Eastern Stroke and Coronary Heart Disease Collaborative Research Group (1998) Blood pressure, cholesterol, and stroke in Eastern Asia. *The Lancet*, **352**:1801–1807.

Ebbeling CB, Pawlak DB, Ludwig DS (2002) Childhood obesity: public health crisis, common sense cure. *The Lancet*, **360**:473–482.

Evans AS (1978) Causation and disease: a chronological journey. *American Journal of Epidemiology*, **108**:249–258.

Hill AB (1965) The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, **58**:295–300.

Horton R (2000) Common sense and figures: the rhetoric of validity in medicine (Bradford Hill Memorial Lecture 1999). *Statistics in Medicine*, **19**: 3149–3164.

Jacobs DR Jr, Ainsworth BE, Hartman TJ, Leon AS (1993) A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Medicine and Science in Sports and Exercise*, **25**:81–91.

Kaufman J, Jing J (2002) China and AIDS—the time to act is now. *Science*, **296**:2339–2340.

Kreitman N (1986) Alcohol consumption and the preventive paradox. *British Journal of Addiction*, **81**:353–363.

Law MR, Wald NJ, Thompson SG (1994) By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *British Medical Journal*, **308**:367–373.

Lee M-J, Popkin BM, Kim S (2000) The unique aspects of the nutrition transition in South Korea: the retention of healthful elements in their traditional diet. *Public Health Nutrition*, **5**:197–203.

Lemmens P (2001) Relationship of alcohol consumption and alcohol problems at the population level. In: *International handbook of alcohol dependence and problems*, Heather N, Peters TJ, Stockwell T, eds. John Wiley and Sons Ltd, Chichester.

Levine JA, Weisell R, Chevassus S, Martinez CD, Burlingame B, Coward WA (2001) The work burden of women. *Science*, **294**:812.

MacLehose L, McKee M, Weinberg J (2002) Responding to the challenge of communicable disease in Europe. *Science*, **295**:2047–2050.

MacMahon S, Peto R, Cutler J et al. (1990) Blood pressure, stroke, and coronary heart disease. Part I. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *The Lancet*, **335**:765–774.

Monteiro CA, Conde WL, Popkin BM (2002) Is obesity replacing or adding to undernutrition? Evidence from different social classes in Brazil. *Public Health Nutrition*, **5**:105–112.

Murray CJL, Lopez AD (1997) Global mortality, disability, and the contribution of risk factors: global burden of disease study. *The Lancet*, **349**:1436–1442.

Murray CJL, Lauer JA, Hutubessy RC et al. (2003) Effectiveness and costs of interventions to lower systolic blood pressure and cholesterol: a global and regional analysis on reduction of cardiovascular-disease risk. *The Lancet*, **361**:717–725.

Pelletier DL, Rahn M (1998) Trends in body mass index in developing countries. *Food and Nutrition Bulletin*, 19:223–239.

Peto R, Lopez AD, Boreham J, Thun M, Heath Jr C (1992) Mortality from tobacco in developed countries. *The Lancet*, **339**:1268–1278.

Popkin BM (2002) An overview on the nutrition transition and its health implications: the Bellagio meeting. *Public Health Nutrition*, **5**:93–103.

Popkin BM, Horton S, Kim S, Mahal A, Shuigao J (2001) Trends in diet, nutritional status and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. *Nutrition Reviews*, **59**: 379–390.

Powell K, Blair S (1994) The public health burdens of sedentary living habits: theoretical but realistic estimates. *Medicine and Science in Sports and Exercise*, **26**:851–856.

Reddy KS, Yusuf S (1998) Emerging epidemic of cardiovascular disease in developing countries. *Circulation*, **97**:596–601.

Rodgers A, Lawes C, MacMahon S (2000) The global burden of cardiovascular disease conferred by raised blood pressure. Benefits of reversal of blood pressure-related cardiovascular risk in Eastern Asia. *Journal of Hypertension*, **18**(Suppl.):S3–S5.

Rodgers A, MacMahon S (1999) Blood pressure and the global burden of cardiovascular disease. *Clinical and Experimental Hypertension*, **21**:543–552.

Rose G (1992) *The strategy of preventive medicine*. Oxford University Press, Oxford.

Skog OJ (1999) Prevention paradox revisited. *Addiction*, **94**:751–757.

Thun MJ, Apicella LF, Henley SJ (2000) Smoking vs other risk factors as the cause of smoking-attributable mortality: confounding in the courtroom. *Journal of American Medical Association*, **284**:706–712.

WHO (1997) *Tobacco or health: a global status report*. World Health Organization, Geneva.

Chapter 27

Potential health gains from reducing multiple risk factors

Majid Ezzati, Stephen Vander Hoorn, Anthony Rodgers, Alan D. Lopez, Colin D. Mathers and Christopher J.L. Murray

1. Introduction

Estimates of the burden of disease attributable to selected individual risk factors were presented in chapter 26. Diseases and injuries are, however, almost always caused by multiple risk factors (Rothman 1976; Walter 1980), motivating analysis of the health benefits of simultaneous reductions in multiple risks. Estimating the joint effects of multiple distal and proximal risks is particularly important because many factors act through other, intermediate, factors (Murray and Lopez 1999; Yerushalmy and Palmer 1959), or in combination with other factors, as we described in chapter 1 of this book. For example, education, occupation and income may affect smoking, physical activity and diet, which are risk factors for cardiovascular diseases, both directly and through further layers of intermediate factors such as body mass index (BMI), blood pressure and cholesterol. Multi-causality also means that a range of interventions can be used for disease prevention, with the specific choice determined by factors such as cost, technology availability, infrastructure and preferences.

A number of works have estimated the joint effects of two or more risk factors in specific cohorts (Hirayama 1990; Neaton and Wentworth 1992; Rothman and Keller 1972; Stampfer et al. 2000; Willet 2002), or for specific groups of diseases and risks (Doll and Peto 1981; Smith et al. 1999). Innovative models and methods have also been developed to quantify the complexity of multiple risk factor effects, especially as they interact over time (Manton et al. 1993; Robins 1999). Estimating joint risk factor effects beyond specific diseases or cohorts, however, remains relatively unexplored in epidemiology and population health. Using comprehensive reviews of data on selected major risk factors in various levels of causality, this chapter is an attempt to do so.

Portions of this chapter have been published previously in *The Lancet*, 2003, **362**:271–280, and have been reproduced with permission from Elsevier Science.

We further used the joint effects of multiple risk factors to estimate the potential gain in healthy life expectancy (HALE) from reducing these risks. Analysis of multiple risk factors, with heterogeneous contributions to disease burden in different populations, would also allow estimating how much of the cross-population health differentials (e.g. differences in HALE) are due to the selected risk factors. By estimating gains in HALE based on causes of disease, this work also contributes in a systematic way to the continued debate on the potential limits to life expectancy (Oeppen and Vaupel 2002; Riley 2001).

2. Methods

2.1 Estimating joint population attributable fractions

Methods and data sources for estimating the burden of disease attributable to individual risk factors were described in chapter 25. The contribution of a risk factor to disease or mortality relative to some alternative exposure scenario (i.e. population attributable fraction, PAF, defined as the proportional reduction in population disease or mortality that would occur if exposure to the risk factor were reduced to an alternative exposure scenario [Eide and Heuch 2001; Miettinen 1974]) is given by the generalized "potential impact fraction" in Equation 1.

$$PIF = \frac{\int_{x=0}^{m} RR(x)P(x)dx - \int_{x=0}^{m} RR(x)P'(x)dx}{\int_{x=0}^{m} RR(x)P(x)dx} \quad (1)$$

where

$RR(x)$: relative risk at exposure level x

$P(x)$: population distribution of exposure

$P'(x)$: alternative or counterfactual distribution of exposure, and

m: maximum exposure level

In equation 1, RR, P, and P' may represent joint relative risks and exposure distributions for multiple risk factors (i.e. x may be a vector of risk factors), with RR for each risk factor estimated at the appropriate level of the remaining ones (Eide and Heuch 2001). Alternatively, for n biologically independent and uncorrelated risk factors, the joint PAF is given by equation 2 (Miettinen 1974; Walter 1976). If risk factors are independent and uncorrelated, the proportion of the remaining disease which is attributed to the ith additional risk factor equals PAF_i (and hence $1 - PAF_i$ not attributable to this factor). Therefore, the second term in the right hand side of equation 2 (i.e. the product of all

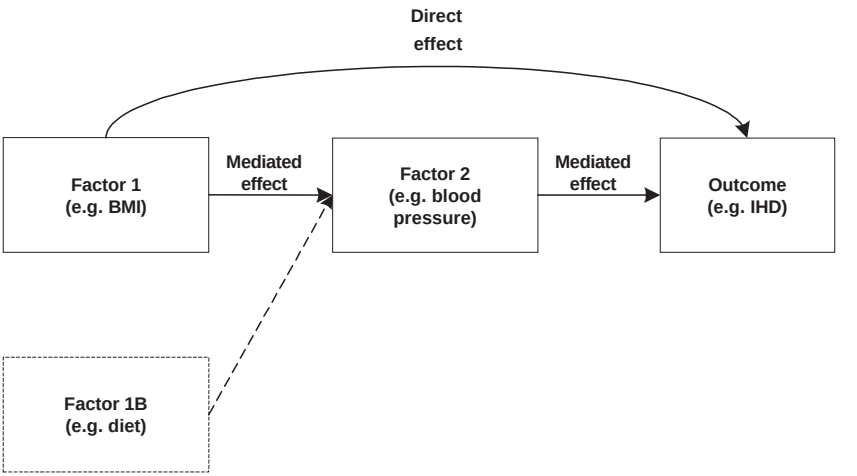
[1 – PAF_i] terms) is the fraction of disease not attributable to any of the n risk factors. One minus this term is the fraction attributable to the combined effects of the n risk factors:

$$PAF = 1 - \prod_{i=1}^{n} (1 - PAF_i) \quad (2)$$

where PAF_i is the PAF of individual risk factors

Estimating the joint effects of multiple risk factors is in practice complex for several reasons. First, some of the effects of the more distal factors (e.g. physical inactivity) are mediated through intermediate factors (e.g. high BMI itself through blood pressure) (Figure 27.1). Estimating the joint effects of distal and intermediate factors requires knowledge of independent hazards of the distal ones (vs individual risk factor effects, which are based on total hazard) (Figure 27.1). Second, the hazard due to a risk factor may depend on the presence of other risk factors (effect modification) (Koopman 1981; Rothman and Greenland 1998). Third, there may be correlation between exposure to various risk factors, because they are affected by the same distal factors and policies.

Figure 27.1 Mediated and direct effects



Note: Some of the effects of a risk factor (e.g. BMI) may be mediated through other factors (e.g. blood pressure). When estimating the total effects of individual distal factor on disease, both mediated and direct effects should be considered. This is because in the presence of mediated effects, controlling for the intermediated factor would attenuate the effects of the more distal one (Greenland 1987). When estimating the joint effects of the more distal factor (e.g. BMI) *and* the intermediate one (e.g. blood pressure), the direct and mediated effects must be separated, especially if the intermediate factor is affected by other distal factors (e.g. diet).

For example, undernutrition, poor water and sanitation and the use of solid fuels are more common among poor rural households in developing countries, or smokers generally have higher and more harmful patterns of alcohol consumption and worse diet than non-smokers.

While the current literature refers to scenarios 1 and 2 as biological interaction and to scenario 3 as statistical interaction (Miettinen 1974; Rothman and Greenland 1998; Rothman et al. 1980), this distinction is somewhat arbitrary and the three scenarios may occur simultaneously. For example, zinc deficiency affects mortality from diarrhoea directly as well as through lowering growth (weight-for-age) (scenario 1) (Brown et al. 2002; Zinc Investigators' Collaborative Group 1999), and may also be correlated with underweight, other micronutrient deficiencies, and poor water and sanitation (scenario 3). Similarly alcohol and smoking may not only be correlated (scenario 3), but also affect each other's hazard for some diseases (scenario 2) (Rothman and Keller 1972). Although the epidemiological literature has placed much emphasis on removing or minimizing the effects of confounding covariates, mediated and stratified hazards have received disproportionately little empirical attention. Therefore, we used reviews of extant literature and re-analysed existing cohort data to strengthen the empirical basis for considering interactions in sensitivity analyses.

In one set of estimates (referred to as the unadjusted scenario), we assumed no mediated effects or interactions among risk factors. We then included the mediated effects and interactions described above in a second scenario (referred to as the adjusted scenario).

Joint effects of cardiovascular disease risk factors

Epidemiological studies on the effects of high BMI, physical inactivity, and low fruit and vegetable intake on cardiovascular disease risk have illustrated some attenuation of the effects after adjustment for intermediate factors (e.g. blood pressure or cholesterol) (Berlin and Colditz 1990; Blair et al. 2001; Eaton 1992; Gaziano et al. 1995; Jarrett et al. 1982; Jousilahti et al. 1999; Khaw and Barrett-Connor 1987; Liu et al. 2001, 2000; Manson et al. 1990, 2002; Rosengren et al. 1999; Tate et al. 1998). This attenuation confirms that some of the hazard of the more distal factors is mediated through the intermediate ones (Figure 27.1). The attenuation has varied among studies but has consistently been less that one half of the excess risk of the distal factors. We used an upper bound of 50% as the proportion of the excess risk from these risk factors mediated through intermediate factors that are themselves among the selected risks.

To include effect modification, deviations from the multiplicative model of 10% for ischaemic heart disease (IHD) and 30% for ischaemic stroke were used based on existing studies (both sub-multiplicative) (Eastern Stroke and Coronary Heart Disease Collaborative Research Group 1998; Neaton and Wentworth 1992).

Joint effects of smoking and other risk factors

Liu et al. (1998) found that in China, the relative risks for mortality from lung and other cancers, respiratory diseases and cardiovascular diseases were approximately constant in different cities whose non-smoker mortality rates from these diseases varied by a factor of 4–10 (see Figures 4 and 6 in Liu et al. 1998). This finding has also been confirmed in studies which stratified hazards for serum cholesterol (Jee et al. 1999).

Joint effects of childhood undernutrition for infectious diseases

Zinc affects child growth (Brown et al. 2002) and some of its effects on infectious diseases may be mediated through growth (e.g. underweight). As no published source for these mediated effects existed, data from some of the available zinc trials (Zinc Investigators' Collaborative Group 1999) were re-analysed and an upper bound of 50% on the proportion of zinc deficiency risk mediated through underweight was used. Vitamin A deficiency, which affects some of the same diseases as underweight and zinc deficiency, has been found not to change the hazard size for the other two risk factors based on stratified results from clinical trials and recent reviews of micronutrient deficiency literature (Christian and West Jr. 1998; Ramakrishnan and Martorell 1998; Ramakrishnan et al. 1995; West et al. 1991).

Joint effects of undernutrition and environmental risk factors in childhood diseases

Anthropometric (growth) indicators of childhood nutrition (e.g. weight-for-age) are aggregate measures of multiple factors which include nutrition (e.g. protein-energy intake) and previous infection (Pelletier et al. 1993; Scrimshaw et al. 1968; UNICEF 1990). Therefore, some of the risks for indoor smoke from solid fuels and poor water, sanitation and hygiene (which result in acute lower respiratory infections [ALRI] and diarrhoea, respectively) may be mediated through underweight. In a review of existing literature, Briend (1990) concluded that attempts to disentangle direct and mediated contributions, especially over long time periods needed to affect population-level anthropometry, have not established diarrhoea as a significant cause of underweight. Other works, however, have found evidence that infection (especially diarrhoea) could result in reduced growth and increased the prevalence of underweight (Black 1991; Guerrant et al. 1992; Lutter et al. 1989, 1992; Martorell et al. 1975a, 1975b; Stephensen 1999). To account for potential mediated effects, we chose an upper bound of 50% for the proportion of the excess risks for indoor smoke from solid fuels and for poor water, sanitation and hygiene mediated through underweight in subregions[1] where underweight was a cause of disease burden.

Risk factor correlation

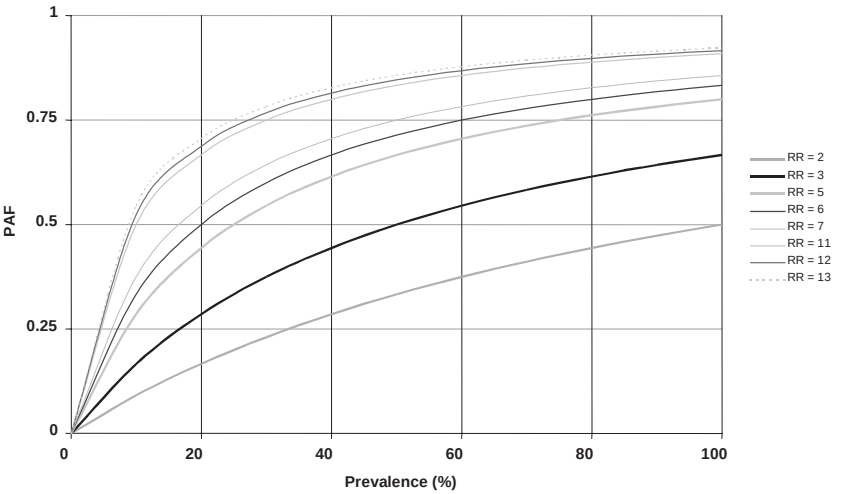
To estimate the joint effects of risk factors with a continuous exposure variable (e.g. blood pressure and cholesterol), each integral in the *PIF* relationship may be replaced with $\int_{x_1=0}^{m_1} \int_{x_2=0}^{m_2} RR_1(x_1)RR_2(x_2)P(x_1,x_2)dx_1dx_2$, where subscripts 1 and 2 denote the two risk factors and P is the joint distribution of the two exposures. If the joint *RR* were a linear function of exposure levels (x_1 and x_2), then correlation between the two risk factors would not affect total hazard. Because individual *RRs* are non-linear functions of exposure (e.g. in a logistic or Cox proportional hazard model) and joint *RRs* are the product of such terms, positive correlation between risk factors would, in general, imply a larger PAF than zero correlation, which in turn would be larger than negative correlation (sub-multiplicative effect modification could result in smaller PAF even with positive correlation for some *RR* values). Similarly, for categorical risk factors, positive correlation would in general result in larger PAF (see also Greenland 1984).

For the range of exposures and relative risks observed here, this secondary effect of risk factor correlation would be considerably smaller than the joint attributable fraction, which may be confirmed by micro-simulation of exposure distributions and relative risks. This is because the PAF relationship is an increasing concave function of individual or joint *RR* (i.e. rate of increase declines with increasing *RR* or prevalence) (Figure 27.2). Because the risk factors considered in this analysis individually accounted for large fractions of the diseases affected by them in populations where these diseases are important components of disease burden, the joint effects approached 100% asymptotically, limiting the overestimation potential (e.g. individually, underweight accounted for 60–70% of under-five diarrhoea in AFR-D, AFR-E and SEAR-D; poor water, sanitation and hygiene for approximately 90%; vitamin A deficiency for 20%-30%; and zinc deficiency for 10–17%; similarly, in various developed subregions, individually, high blood pressure accounted for 44–64% of IHD; high cholesterol for 51–68%; high BMI for 17–36%; low fruit and vegetable intake for 19–35%; and physical inactivity for 15–16%).

2.2 Gains in healthy life expectancy (HALE)

The incidence of many conditions (e.g. neuropsychiatric conditions or long-term effects of injuries) may cause ill health but not death. It is therefore important to capture both fatal and non-fatal health outcomes in describing population health. Healthy life expectancy or health-adjusted life expectancy (HALE) reduces total life expectancy into equivalent years of "full health" by taking into account the distribution and severity of health states in the population (Mathers 2002). Inputs to the calculation of HALE include the period life table (or age-sex-specific

Figure 27.2 PAF relationship as a function of prevalence and relative risk



Note: PAF relationship is an increasing concave function of both prevalence (seen in the shape of each curve) and RR (seen in the declining distance between each adjacent pair of curves). As a result, for joint risk factor PAF, errors due to deviations from a simple uncorrelated multiplicative model (due to risk factor correlation or effect modification) are secondary to the joint PAF. For example, for two risk factors with RR = 2 and RR = 3, a multiplicative model would result in a joint RR of 6. The PAF would be approximately correct even if the true joint RR were 5 or 7, as the PAF curves are close for these RR values. This phenomenon becomes increasingly dominant with increasing number of risk factors (i.e. the curves for RR = 11, 12, and 13 are even closer than those for RR = 5, 6 and 7). Further, the flattening of PAF curves at high exposures limits the error due to risk factor correlation.

mortality rates) and prevalences of health states (resulting from diseases, their sequelae, and their combinations) for each country. Methods for estimating HALE have been described in detail elsewhere (Mathers et al. 2001). Unlike estimates of the burden of disease which compare current mortality and disability to a normative survivorship function (Murray and Lopez 1996), estimates of the gain in HALE account for competing risks.

The estimates in this chapter show the improvements in HALE for the year 2000, that would have been observed if exposure to the selected risk factors had been reduced to the theoretical-minimum-risk counterfactual distribution, as described in each of the risk factor chapters. In each of the 14 subregions, joint disease-specific PAFs were estimated for all diseases affected by the 20 leading global risk factors (chapter 26; see also individual risk factor chapters), for all age and sex groups.

Mortality and incidence in the counterfactual scenario are $(1 - PAF_M)$ and $(1 - PAF_I)$ times their original values where PAF_M and PAF_I are the PAF of mortality and incidence attributable to the joint effects of the risk

factors. The age-sex-specific mortality and age-sex-cause-specific prevalence of diseases and their sequelae were adjusted to these levels to estimate the HALE gain as a result of multiple risk factor removal. Reduction in prevalence was obtained from reduction in incidence under equilibrium conditions (Kruijshaar et al. 2002). Cause-specific estimates were necessary for non-fatal conditions because of different disability weights (Murray and Lopez 1996) but not for fatal conditions.

3. Results

Table 27.1 shows the individual and joint contributions of the 20 selected risk factors for the 10 leading diseases in the world and in three broad combinations of subregions—high-mortality developing (38% of global population), lower-mortality developing (40% of global population) and demographically and economically developed (22% of global population).[2] For most diseases, the joint effects of these risk factors were substantially less than the crude sum of the individual effects (e.g. globally four separate risk factors were each responsible for 10%, 18%, 45% and 88% of diarrhoeal disease, but with a joint PAF of 92–94%), confirming that a large number of cases are caused by the joint actions of more than one of these risk factors acting as sufficient causes (Rothman 1976), or through other factors.

Globally, large fractions of diarrhoea (92–94%), ALRI (55–62%), lung cancer (72%), upper aerodigestive cancer (60%), chronic obstructive pulmonary disease (COPD) (60%), IHD (83–89%) and stroke (70–76%) were attributable to the joint effects of the risk factors considered here (see Willet 2002 and Stampfer et al. 2000 for consistent vascular disease examples from specific cohorts). The joint PAFs for cancers other than lung and upper aerodigestive (23%), perinatal conditions (23%), maternal conditions (42%), and intentional (29%) and unintentional (20%) injuries, which have more diverse risk factors, were smaller but non-negligible. Although the fraction of total malaria burden attributable to childhood undernutrition was relatively large (56–59%), this was because of the contribution of mortality at younger ages to disease burden. No adult malaria was attributed to the above risk factors because the epidemiological literature has focused on quantifying increased risk of malaria as a result of childhood undernutrition only. Finally, with the exception of alcohol and drug dependence, which were fully attributable to their specific risk factors, very small fractions or none of neuropsychiatric conditions, tuberculosis, congenital anomalies, and a number of other diseases were attributed to the risk factors considered in this book.

Figure 27.3 shows the individual and joint contributions (including overlap) of selected major risk factors to each of the following disease categories in the three subregional groups described above: I. communicable, maternal, perinatal and nutritional conditions;

Table 27.1 Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world and different subregional groupings[a]

(a) World

Disease/condition	*% global disease burden (total 1.46 billion DALYs)*	*% global mortality (total 55.9 million deaths)*	*Contributing risk factors (individual PAF for disease burden)*	*Joint PAF[b] (disease burden)*	*Joint PAF[b] (mortality)*
Lower respiratory infections	6.1	6.8	Underweight (childhood) (40%); zinc deficiency (16%); indoor smoke from solid fuels (36%); tobacco (2%)[c]	55–62%	40–45%
HIV/AIDS	5.5	4.6	Unsafe sex (94%); unsafe health care injections (5%); illicit drugs (3%)	96%	96%
Unipolar depressive disorders	4.5	0.0	Alcohol (2%); childhood sexual abuse (6%)	7%	NA[d]
Diarrhoeal diseases	4.2	3.5	Underweight (childhood) (45%); vitamin A deficiency (18%); zinc deficiency (10%); unsafe water, sanitation and hygiene (88%)	92–94%	92–94%
Ischaemic heart disease	4.0	12.6	High blood pressure (49%); high cholesterol (56%); high BMI (21%); low fruit and vegetable intake (31%); physical inactivity (22%); tobacco (12%); alcohol (2%)	83–89%	78–85%
Low birth weight	3.5	2.5	Underweight (maternal) (10%); iron deficiency (19%); alcohol (0.2%)	29%	31%
Stroke	3.1	9.6	High blood pressure (62%); high cholesterol (18%); high BMI (13%); low fruit and vegetable intake (11%); physical inactivity (7%); tobacco (12%); alcohol (4%)	70–76%	65–73%

continued

Table 27.1 Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world and different subregional groupings[a] (*continued*)

(a) World (continued)

Disease/condition	*% global disease burden (total 1.46 billion DALYs)*	*% global mortality (total 55.9 million deaths)*	*Contributing risk factors (individual PAF for disease burden)*	*Joint PAF[b] (disease burden)*	*Joint PAF[b] (mortality)*
Malaria	2.9	2.0	Underweight (childhood) (45%); vitamin A deficiency (16%); zinc deficiency (18%)	56–59%	60–62%
Road traffic accidents	2.6	2.2	Alcohol (20%); illicit drugs (2%); occupational risk factors for injuries (6%)	28%	29%
Tuberculosis	2.5	2.9	Tobacco (10%)[c]	10%	12%
Communicable, maternal, perinatal, and nutritional conditions	42.0	32.6	Multiple risks (see chapter 26)	49–50%	50–51%
Noncommunicable diseases	45.7	58.3	Multiple risks (see chapter 26)	35–36%	49–52%
Injuries	12.3	9.1	Multiple risks (see chapter 26)	22%	25%
All causes	100	100	All 20 selected risks (see chapter 26)	39–40%	47–49%

(b) High-mortality developing subregions					
Disease/condition	*% regional disease burden (total 830 million DALYs)*	*% regional mortality (total 26.4 million deaths)*	*Contributing risk factors (individual PAF for disease burden)*	*Joint PAF[b] (disease burden)*	*Joint PAF[b] (mortality)*
HIV/AIDS	9.0	9.2	Unsafe sex (97%); unsafe health care injections (5%); illicit drugs (0.3%)	97%	97%
Lower respiratory infections	8.2	9.8	Underweight (childhood) (46%); zinc deficiency (19%); indoor smoke from solid fuels (41%); tobacco (1%)[c]	62–69%	49–54%
Diarrhoeal diseases	6.3	6.6	Underweight (childhood) (49%); vitamin A deficiency (19%); zinc deficiency (11%); unsafe water, sanitation and hygiene (88%)	93–95%	93–94%
Low birth weight	5.0	4.4	Underweight (maternal) (12%); iron deficiency (22%); alcohol (0.2%)	32%	34%
Malaria	4.9	4.2	Underweight (childhood) (45%); vitamin A deficiency (17%); zinc deficiency (19%)	57–60%	60–63%
Unipolar depressive disorders	3.1	0.0	Alcohol (1%); childhood sexual abuse (8%)	9%	NA[d]
Measles	3.0	2.7	Underweight (childhood) (34%); vitamin A deficiency (15%)	42%	43%
Ischaemic heart disease	3.0	9.1	High blood pressure (44%); high cholesterol (54%); high BMI (11%); low fruit and vegetable intake (33%); physical inactivity (21%); tobacco (8%); alcohol (4%)	80–87%	77–84%
Tuberculosis	2.9	3.8	Tobacco (8%)[c]	8%	10%
Birth asphyxia and birth trauma	2.7	1.9	Iron deficiency (20%)	20%	27%

continued

Table 27.1 Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world and different subregional groupings[a] (*continued*)

(b) High-mortality developing subregions (continued)

Disease/condition	*% regional disease burden (total 830 million DALYs)*	*% regional mortality (total 26.4 million deaths)*	*Contributing risk factors (individual PAF for disease burden)*	*Joint PAF[b] (disease burden)*	*Joint PAF[b] (mortality)*
Communicable, maternal, perinatal, and nutritional conditions	58.9	54.5	Multiple risks (see chapter 26)	54–56%	56–57%
Noncommunicable diseases	30.5	37.0	Multiple risks (see chapter 26)	33–34%	47–50%
Injuries	10.6	8.4	Multiple risks (see chapter 26)	17%	19%
All causes	100	100	All 20 selected risks (see chapter 26)	44–45%	50–51%

(c) Low-mortality developing subregions

Disease/condition	*% regional disease burden (total 408 million DALYs)*	*% regional mortality (total 16.0 million deaths)*	*Contributing risk factors (individual PAF for disease burden)*	*Joint PAF[b] (disease burden)*	*Joint PAF[b] (mortality)*
Unipolar depressive disorders	5.9	0.0	Alcohol (1%); childhood sexual abuse (4%)	5%	NA[d]
Stroke	4.7	13.8	High blood pressure (58%); high cholesterol (13%); high BMI (11%); low fruit and vegetable intake (10%); physical inactivity (5%); tobacco (8%); alcohol (7%)	67–74%	62–69%

Lower respiratory infections	4.1	4.6	Underweight (childhood) (24%); zinc deficiency (5%); indoor smoke from solid fuels (20%); tobacco (3%)[c]	35–42%	25–29%
Road traffic accidents	4.1	3.4	Alcohol (20%); illicit drugs (1%); occupational risk factors for injuries (6%)	27%	27%
Chronic obstructive pulmonary disease	3.8	9.2	Indoor smoke from solid fuels (26%); tobacco (26%)	52%	55%
Ischaemic heart disease	3.2	9.3	High blood pressure (45%); high cholesterol (48%); high BMI (22%); low fruit and vegetable intake (31%); physical inactivity (22%); tobacco (8%); alcohol (3%)	79–87%	73–82%
Birth asphyxia and birth trauma	2.6	1.1	Iron deficiency (10%)	10%	17%
Tuberculosis	2.4	3.3	Tobacco (10%)	12%	13%
Alcohol use disorders	2.3	0.2	Alcohol (100%); childhood sexual abuse (5%)	100%	100%
Hearing loss	2.2	0.0	Tobacco (5%)[c]	5%	NA[d]
Communicable, maternal, perinatal, and nutritional conditions	25.0	18.1	Multiple risks (see chapter 26)	28–29%	27–28%
Noncommunicable diseases	59.8	70.5	Multiple risks (see chapter 26)	33–34%	46–48%
Injuries	15.3	11.5	Multiple risks (see chapter 26)	24%	25%
All causes	100	100	All 20 selected risks (see chapter 26)	30–31%	40–42%

continued

Table 27.1 Individual and joint contributions of 20 selected risks to 10 leading diseases and total burden of disease in the world and different subregional groupings[a] (*continued*)

(d) Developed subregions

Disease/condition	*% regional disease burden (total 214 million DALYs)*	*% regional mortality (total 13.5 million deaths)*	*Contributing risk factors (individual PAF for disease burden)*	*Joint PAF[b] (disease burden)*	*Joint PAF[b] (mortality)*
Ischaemic heart disease	9.4	23.3	High blood pressure (58%); high cholesterol (63%); high BMI (33%); low fruit and vegetable intake (28%); physical inactivity (22%); tobacco (22%); alcohol (–0.2%)	89–93%	82–87%
Unipolar depressive disorders	7.2	0.0	Alcohol (3%); childhood sexual abuse (4%)	7%	NA[d]
Stroke	6.0	13.4	High blood pressure (72%); high cholesterol (27%); high BMI (23%); low fruit and vegetable intake (12%); physical inactivity (9%); tobacco (22%); alcohol (0%)	81–86%	71–79%
Alcohol use disorders	3.5	0.2	Alcohol (100%); childhood sexual abuse (3%)	100%	100%
Alzheimer and other dementias	3.0	1.4	None of the selected risks	0%	NA[d]
Hearing loss	2.8	0.0	Tobacco (10%)[c]	10%	NA[d]
Chronic obstructive pulmonary disease	2.6	3.2	Indoor smoke from solid fuels (2%); tobacco (69%)	71%	74%

Road traffic accidents	2.5	1.4	Alcohol (38%); illicit drugs (4%); occupational risk factors for injuries (4%)	45%	44%
Osteoarthritis	2.5	0.0	High BMI (21%); Tobacco (10%)[c]	28%	NA[d]
Trachea, bronchus and lung cancers	2.4	4.5	Indoor smoke from solid fuels (coal only) (0%); tobacco (85%); low fruit and vegetable intake (11%)	86%	87%
Communicable, maternal, perinatal, and nutritional conditions	9.0	6.7	Multiple risks (see chapter 26)	24–25%	20–21%
Noncommunicable diseases	78.2	85.7	Multiple risks (see chapter 26)	41–42%	54–57%
Injuries	12.8	7.6	Multiple risks (see chapter 26)	36%	37%
All causes	100	100	All 20 selected risks (see chapter 26)	39–40%	51–53%

NA Not applicable.

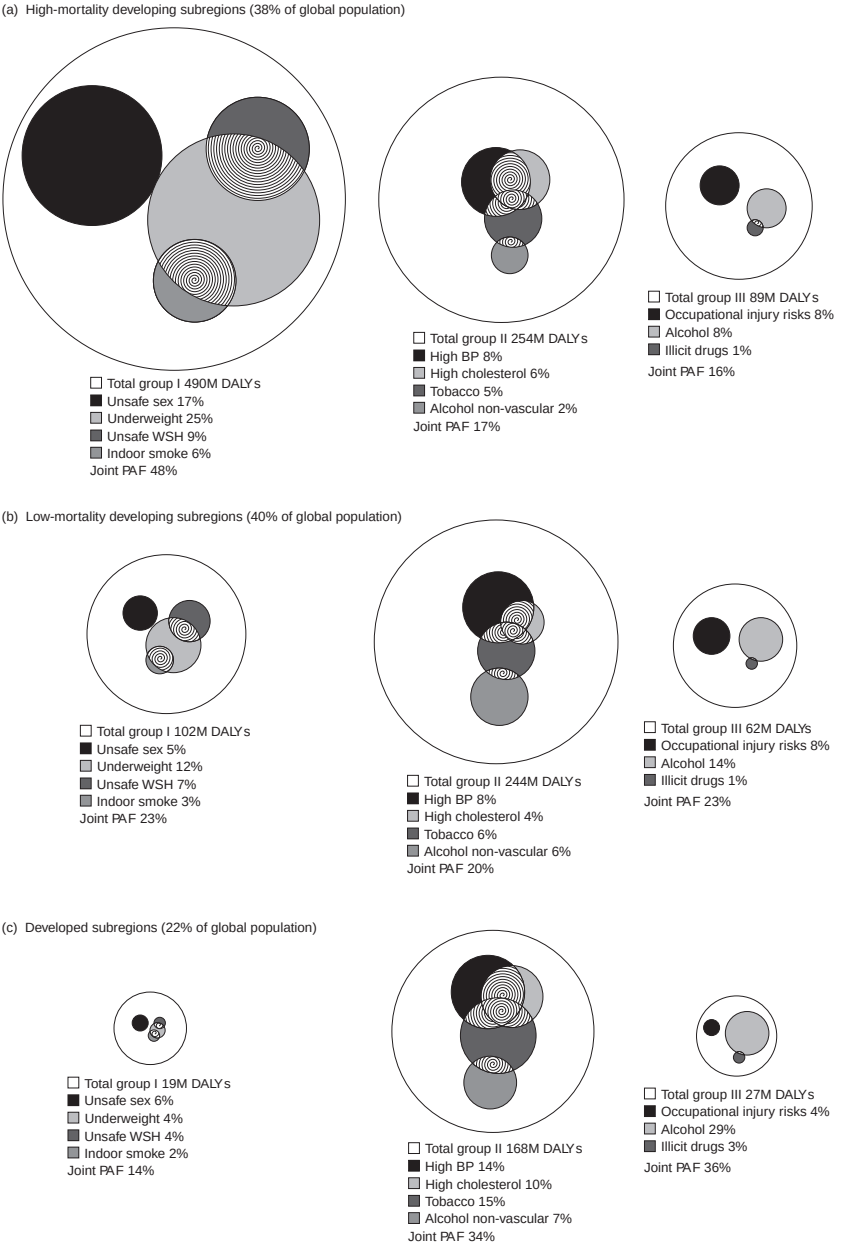
[a] The risk factors also contribute to other diseases in each subregion which are not among the leading 10.

[b] The first number is the PAF for adjusted scenario and the second for the unadjusted scenario in cases where adjustment for mediated effects and effect modification applied (see methods).

[c] Affected by tobacco in the category "other respiratory diseases" or "selected other medical causes" (Peto et al. 1992). The PAF has large uncertainty.

[d] The number of deaths coded to "hearing loss", "unipolar depressive disorders", "osteoarthritis", and "alzheimer and other dementias" is zero or very small in the GBD database, making the mortality PAF for these diseases undefined or unstable.

Figure 27.3 Individual and joint contributions (adjusted scenario as described in methods) of selected risk factors to different disease groups



continued

Notes for Figure 27.3

Key: High-mortality developing subregions: AFR, AMR-D, EMR-D and SEAR-D. Low-mortality developing subregions: AMR-B, EMR-B, SEAR-B and WPR-B. Developed subregions: AMR-A, EUR and WPR-A. Group I: communicable, maternal, perinatal and nutritional conditions; Group II: noncommunicable diseases; Group III: injuries; WSH, water, sanitation and hygiene; BP, blood pressure.

Note: The size of each circle shows the absolute size of the burden (in millions of DALYs). Numbers for individual risk factors show the total burden including those overlapping with the remaining factors shown in lined pattern. Note that each risk factor may also have contributions to other disease groups (e.g. indoor smoke also causes COPD, which is in Group II and alcohol also causes injuries, which are in Group III). In reality, there is a small overlap between underweight and unsafe sex since underweight children with HIV/AIDS are likely to survive for a shorter period (not estimated in this work).

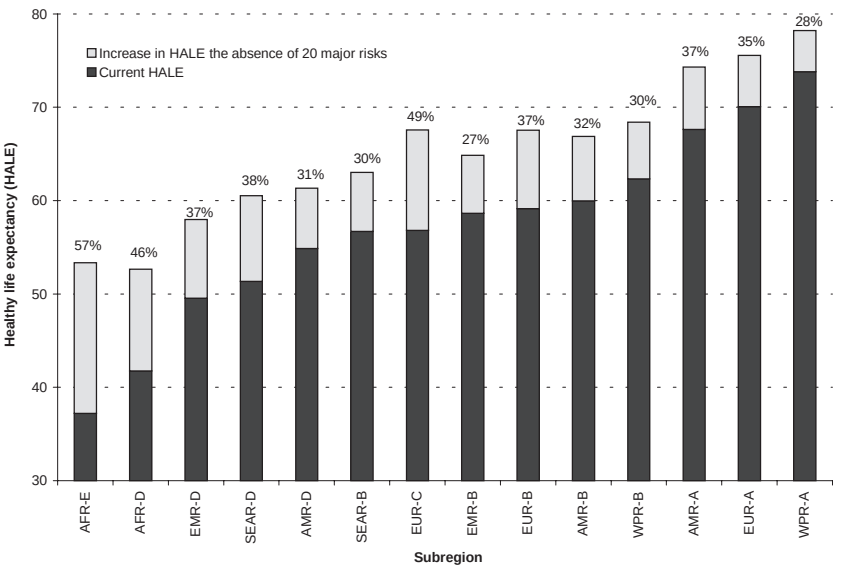
II. noncommunicable diseases; III. injuries. Communicable, maternal, perinatal and nutritional conditions, and their underlying risk factors, in high-mortality developing subregions contributed disproportionately to global loss of healthy life (i.e. large Group I disease burden and large fraction (54–56%) attributable to the selected risk factors). Noncommunicable diseases and their risk factors (54–57% attributable to the selected risk factors) dominated the burden of disease in developed subregions, although, by comparison, this was considerably smaller than total disease burden in high-mortality developing subregions. Both disease groups and their risk factors, with intermediate levels, affected low-mortality developing subregions.

Table 27.1 and Figure 27.3 also show that the selected risk factors for each disease group exhibited heterogeneous contributions to disease burden across clusters of countries and subregions. For Group I diseases, unsafe sex had the lowest proportional contribution in low-mortality developing countries but made a very large contribution in high-mortality developing countries. For Group II diseases, the relative contributions of high cholesterol and tobacco varied across the three subregional groupings, with their relative contributions reversing from one to another. Similarly, for Group III, the relative contributions of alcohol and occupational factors showed considerable heterogeneity across subregional groups.

Gains in HALE from removing these 20 selected risk factors are shown in Figure 27.4. Globally, in the year 2000, an estimated 47% of mortality and 39% of disease burden were attributable to the joint effects of 20 selected risk factors. Global HALE would increase from 56.2 to 65.5 years in the absence of these risks (adjusted scenario). The corresponding results for the unadjusted scenario were nearly identical with HALE increasing to 66.0 years.

Figure 27.4 shows that the removal of major risk factors would not only have resulted in improvements in each subregion, but also, in general, reduced the health differentials across subregions (i.e. larger gains in subregions with lower HALE) with the largest gain in health in

Figure 27.4 The joint effects of leading 20 global risk factors on HALE in year 2000, by subregion (adjusted scenario)



Note: The numbers show the fraction of the total burden of disease in each subregion attributable to the selected risk factors.

AFR-E (16.1 years) and the smallest in WPR-A (4.4 years). Important exceptions to the monotonic decreasing relationship between HALE gain and initial HALE were EUR-C and EUR-B (mainly consisting of the countries of eastern and central Europe and the former Soviet Union). In these subregions, the leading global risk factors jointly account for a disproportionately larger share of disease burden (49% in EUR-C and 37% in EUR-B) and led to substantial loss of healthy life years (10.7 in EUR-C and 8.3 in EUR-B), emphasizing the concentration of disease burden among a few important risk factors (alcohol, tobacco, high blood pressure and high cholesterol) in these two subregions.

4. Discussion

The estimates of the joint contributions of 20 selected leading global risk factors showed that these risks together were responsible for a considerable loss of healthy life in different regions of the world. In particular, for some of the leading global diseases (e.g. ALRI, diarrhoea, lung cancer, IHD and stroke), substantial proportions were attributable to these selected risk factors. Removing these 20 risk factors would not only have resulted in a 9.3-year (17%) gain in global HALE, but also would have

accounted for some of the interregional HALE differences. In fact, the analysis showed that even populations with currently high HALE (e.g. developed regions of the western Pacific and Europe) could further benefit from risk reduction. These results provide a guide to the potential gains in (healthy) life expectancy (estimated statistically from past trends [Oeppen and Vaupel 2002; Riley 2001]) through disease prevention by reducing known risks. Similar analyses for the leading 10 selected global risks suggest a gain of 8.1 years in HALE (vs 9.3 years for the leading 20). This concentration of disease burden further emphasizes the contribution of leading risks such as undernutrition, unsafe sex, high blood pressure, tobacco and alcohol to global loss of healthy life.

At the same time, the estimated joint contributions of these risk factors left an important part of the global disease burden unexplained and did not fully explain interregional HALE differentials. This was because only a small fraction of some important diseases was attributable to the selected risk factors considered here. These include diseases whose determinants: i) are diffuse among environmental and behavioural factors (e.g. some cancers, perinatal conditions, and neuropsychiatric diseases) (see Doll and Peto 1981 for examples from cancers); ii) have more complex, multi-factor etiology and often heterogeneous determinants in different populations and therefore difficult to quantify without data at very small scale (e.g. tuberculosis and injuries); iii) involve long delays; or iv) have limited quantitative research at the population level (e.g. neuropsychiatric diseases), often as a result of the above three factors as well as difficulties in measuring exposure or outcome (Evans 1978). Mitigation of many such diseases (e.g. malaria, tuberculosis or injuries) may be better guided by analyses of the effects of interventions tailored to individual settings than by risk factor analysis.

The results of this analysis changed little with plausible assumptions about mediated risks or effect modification among risk factors. An important reason for this is the concave shape of the PAF relationship (Figure 27.2). Because risk factors considered in the analysis individually accounted for large fractions of the diseases affected by them (e.g. diarrhoea and IHD), the joint affects approached 100% asymptotically limiting the sensitivity of results to assumptions about interaction. We emphasize that this does not include the considerably larger uncertainty in each of the individual PAF estimates discussed in detail in chapters 1 and 26 of this book. At the same time, since for many of the important causes of global disease burden (e.g. childhood infectious and vascular diseases), multiple important risk factors were included, the joint effects would likely remain large regardless of uncertainties in the individual PAF.

An additional important source of uncertainty, affecting both individual and joint risk factor estimates, is the concentration of *both* risks and diseases in specific subgroups (vs correlations of risks alone, discussed above). For many risk factors and diseases, exposure and outcome

are simultaneously higher in some groups (e.g. higher malnutrition, unsafe water, sanitation and hygiene, and indoor smoke in poor rural households in developing countries; unhealthy diet, and higher smoking and BMI in some groups in developed countries). In these circumstances, PAFs based on population averages would in general underestimate the effects compared to group-specific analysis, even if the relative risks are constant across groups (Greenland 1984) (also confirmed by micro-simulation). Higher concentration of disease and mortality (e.g. childhood mortality and vascular diseases) in the same groups due to factors such as limited access to health services would magnify this effect, becoming an important contributor to underestimation of the benefits of risk reduction when population level exposure and mortality data are used. In addition to risk factor analysis, estimates of HALE include large uncertainty, especially in countries with poor mortality and disease registration systems as estimated and discussed elsewhere (Mathers et al. 2001). Further implications of these findings for research, and for policies and programmes aimed at improving population health, are discussed in chapter 29.

Acknowledgements

We thank T. Armstrong, R.E. Black, F. Bull, G. Colditz (with E. Rimm and M. Stampfer), C. Lawes, K. Lock, V. Parag, J. Powles, A.J. Rice, K.P. West Jr., G. Whitlock, W. Willet and M. Woodward for discussion and references on independent and mediated effects.

Notes

1 See preface for an explanation of this term.

2 High-mortality developing subregions: AFR, AMR-D, EMR-D and SEAR-D. Low-mortality developing subregions: AMR-B, EMR-B, SEAR-B and WPR-B. Developed subregions: AMR-A, EUR and WPR-A.

References

Berlin JA, Colditz GA (1990) A meta-analysis of physical activity in the prevention of coronary heart disease. *American Journal of Epidemiology*, **132**:612–628.

Black RE (1991) Would control of childhood infectious diseases reduce malnutrition? *Acta Paediatrica Scandinavica Supplement*, **374**:133–140.

Blair SN, Cheng Y, Holder JS (2001) Is physical activity or physical fitness more important in defining health benefits? *Medicine and Science in Sports and Exercise*, **33**:S379–399.

Briend A (1990) Is diarrhoea a major cause of malnutrition among the under-fives in developing countries? A review of available evidence. *European Journal of Clinical Nutrition*, **44**:611–628.

Brown KH, Peerson JM, Rivera J, Allen LH (2002) Effect of supplemental zinc on the growth and serum zinc concentrations of prepubertal children: a meta-analysis of randomized controlled trials. *American Journal of Clinical Nutrition*, **75**:1062–1071.

Christian P, West Jr KP (1998) Interactions between zinc and vitamin A: an update. *American Journal of Clinical Nutrition*, **68**:S435–441.

Doll R, Peto R (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *Journal of the National Cancer Institute*, **66**:191–308.

Eastern Stroke and Coronary Heart Disease Collaborative Research Group (1998) Blood pressure, cholesterol, and stroke in Eastern Asia. *The Lancet*, **352**:1801–1807.

Eaton CB (1992) Relation of physical activity and cardiovascular fitness to coronary heart disease, Part I: a meta-analysis of the independent relation of physical activity and coronary heart disease. *Journal of the American Board of Family Practice*, **5**:31–42.

Eide GE, Heuch I (2001) Attributable fractions: fundamental concepts and their visualization. *Statistical Methods in Medical Research*, **10**:159–193.

Evans AS (1978) Causation and disease: a chronological journey. *American Journal of Epidemiology*, **108**:249–258.

Gaziano JM, Manson JE, Branch LG, Colditz GA, Willett WC, Buring JE (1995) A prospective study of consumption of carotenoids in fruits and vegetables and decreased cardiovascular mortality in the elderly. *Annals of Epidemiology*, **5**:255–260.

Greenland S (1984) Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. *Statistics in Medicine*, **3**:131–141.

Greenland S (1987) Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews*, **9**:1–30.

Guerrant RL, Schorling JB, McAuliffe JF, de Souza MA (1992) Diarrhea as a cause and an effect of malnutrition: diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration. *American Journal of Tropical Medicine and Hygiene*, **47**:28–35.

Hirayama T (1990) *Life-style and mortality: a large-scale census-based cohort study in Japan*. Karger, Tokyo.

Jarrett RJ, Shipley MJ, Rose G (1982) Weight and mortality in the Whitehall study. *British Medical Journal*, **285**:535–537.

Jee SH, Suh I, Kim IS, Appel LJ (1999) Smoking and atherosclerotic cardiovascular disease in men with low levels of serum cholesterol: the Korea Medical Insurance Corporation study. *Journal of the American Medical Association*, **282**:2149–2155.

Jousilahti P, Vartiainen E, Tuomilehto J, Puska P (1999) Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. *Circulation*, **99**:1165–1172.

Khaw KT, Barrett-Connor E (1987) Dietary fiber and reduced ischemic heart disease mortality rates in men and women: a 12-year prospective study. *American Journal of Epidemiology*, **126**:1093–1102.

Koopman JS (1981) Interaction between discrete causes. *American Journal of Epidemiology* **113**:716–724.

Kruijshaar ME, Barendregt JJ, Hoeymans N (2002) The use of models in the estimation of disease epidemiology. *Bulletin of the World Health Organization*, **80**:622–628.

Liu BQ, Peto R, Chen ZM et al. (1998) Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths. *British Medical Journal*, **317**:1411–1422.

Liu S, Lee IM, Ajani U, Cole SR, Buring JE, Manson JE (2001) Intake of vegetables rich in carotenoids and risk of coronary heart disease in men: the Physicians' Health Study. *International Journal of Epidemiology*, **30**: 130–135.

Liu S, Manson JE, Lee IM, Cole SR, Hennekens CH, Willett WC, Buring JE (2000) Fruit and vegetable intake and risk of cardiovascular disease: the Women's Health Study. *American Journal of Clinical Nutrition*, **72**: 922–928.

Lutter CK, Habicht JP, Rivera JA, Martorell R (1992) The relationship between energy intake and diarrhoeal disease in their effects on child growth: biological model, evidence, and implications for public health policy. *Food and Nutrition Bulletin*, **14**:36–42.

Lutter CK, Mora JO, Habicht JP et al. (1989) Nutritional supplementation: effects on child stunting because of diarrhea. *American Journal of Clinical Nutrition*, **50**:1–8.

Manson JE, Colditz GA, Stampfer MJ et al. (1990) A prospective study of obesity and risk of coronary heart disease in women. *New England Journal of Medicine*, **322**:882–889.

Manson JE, Greenland P, LaCroix AZ et al. (2002) Walking compared with vigorous exercise for the prevention of cardiovascular events in women. *New England Journal of Medicine*, **347**:755–756.

Manton KG, Singer BH, Suzman RM (1993) *Forecasting the health of elderly populations*. Springer-Verlag, New York.

Martorell R, Habicht JP, Yarbrough C, Lechtig A, Klein RE, Western KA (1975a) Acute morbidity and physical growth in rural Guatemalan children. *American Journal of Diseases of Children*, **129**:1296–1301.

Martorell R, Yarbrough C, Lechtig A, Habicht JP, Klein RE (1975b) Diarrheal diseases and growth retardation in preschool Guatemalan children. *American Journal of Physical Anthropology*, **43**:341–346.

Mathers CD (2002) Health expectancies: an overview and critical appraisal. In: *Summary measures of population health: concepts, ethics, measurement and applications*. Murray CJL, Salomon JA, Mathers CD, Lopez AD, eds. World Health Organization, Geneva.

Mathers CD, Sadana R, Salomon JA, Murray CJL, Lopez AD (2001) Healthy life expectancy in 191 countries, 1999. *The Lancet*, **357**:1685–1691.

Miettinen OS (1974) Proportion of disease caused or prevented by a given exposure, trait or intervention. *American Journal of Epidemiology*, **99**:325–332.

Murray CJL, Lopez AD, eds. (1996) *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990*. The Global Burden of Disease and Injury, Vol. 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.

Murray CJL, Lopez AD (1999) On the comparable quantification of health risks: lessons from the Global Burden of Disease. *Epidemiology*, **10**:594–605.

Neaton JD, Wentworth D (1992) Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease. Overall findings and differences by age for 316 099 white men. Multiple risk factor intervention Trial Research Group. *Archives of Internal Medicine*, **152**:56–64.

Oeppen J, Vaupel JW (2002) Broken limits to life expectancy. *Science*, **296**: 1029–1030.

Pelletier DL, Frongillo EA Jr, Habicht JP (1993) Epidemiologic evidence for a potentiating effect of malnutrition on child mortality. *American Journal of Public Health*, **83**:1130–1133.

Peto R, Lopez AD, Boreham J, Thun M, Heath Jr C (1992) Mortality from tobacco in developed countries. *The Lancet*, **339**:1268–1278.

Ramakrishnan U, Latham MC, Abel R (1995) Vitamin A supplementation does not improve growth of preschool children: a randomized, double-blind field trial in South India. *Journal of Nutrition*, **125**:202–211.

Ramakrishnan U, Martorell R (1998) The role of vitamin A in reducing child mortality and morbidity and improving growth. *Salud Publica de Mexico*, **40**:189–198.

Riley JC (2001) *Rising life expectancy: a global history*. Cambridge University Press, Cambridge.

Robins JM (1999) Marginal structural models versus structural nested models as tools for causal inference. In: *Statistical models in epidemiology.* Halloran E, ed. Springer Verlag, New York.

Rosengren A, Wedel H, Wilhelmsen L (1999) Body weight and weight gain during adult life in men in relation to coronary heart disease and mortality: a prospective population. *European Health Journal*, **20**:269–277.

Rothman KJ (1976) Causes. *American Journal of Epidemiology*, **104**:587–592.

Rothman KJ, Greenland S (1998) *Modern epidemiology*. Lippincott-Raven, Philadelphia, PA.

Rothman KJ, Greenland S, Walker AM (1980) Concepts of interaction. *American Journal of Epidemiology*, **112**:467–470.

Rothman KJ, Keller A (1972) The effect of joint exposure to alcohol and tobacco on the risk of cancer of the mouth and pharynx. *Journal of Chronic Disease*, **25**:711–716.

Scrimshaw NS, Taylor CE, Gordon JE (1968) *Interactions of nutrition and infection*. (WHO Monograph Series No. 57.) World Health Organization, Geneva.

Smith KR, Corvalan CF, Kjellstrom T (1999) How much global ill health is attributable to environmental factors. *Epidemiology*, **10**:573–584.

Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC (2000) Primary prevention of coronary heart disease in women through diet and lifestyle. *New England Journal of Medicine*, **343**:16–22.

Stephensen CB (1999) Burden of infection on growth failure. *Journal of Nutrition*, **129**:S534–538.

Tate RB, Manfreda J, Cuddy TE (1998) The effect of age on risk factors for ischemic heart disease: the Manitoba Follow-up Study, 1948–1993. *Annals of Epidemiology*, **8**:415–421.

UNICEF (1990) *Strategy to improve nutrition of children and women in developing countries: a UNICEF policy review*. United Nations Children's Fund, New York.

Walter SD (1976) The estimation and interpretation of attributable risk in health research. *Biometrics*, **32**:829–849.

Walter SD (1980) Prevention for multifactorial diseases. *American Journal of Epidemiology*, **112**:409–416.

West KP Jr, Pokhrel RP, Katz J et al. (1991) Efficacy of vitamin A in reducing preschool child mortality in Nepal. *The Lancet*, **338**:67–71.

Willet WC (2002) Balancing life-style and genomics research for disease prevention. *Science*, **296**:695–698.

Yerushalmy J, Palmer CE (1959) On the methodology of investigations of etiologic factors in chronic diseases. *Journal of Chronic Disease*, **108**:27–40.

Zinc Investigators' Collaborative Group (1999) Prevention of diarrhea and pneumonia by zinc supplementation in children in developing countries: pooled analysis of randomized controlled trials. *Journal of Pediatrics*, **135**:689–697.

Chapter 28

Effects of multiple interventions

James Robins, Miguel Hernan and Uwe Siebert

1. Introduction

The purpose of this chapter is (i) to describe some currently available analytical methods for using individual level epidemiological data to estimate the impact of multiple risk factor interventions on health and (ii) to carefully review the conditions under which these methods deliver unbiased estimates of impact. The chapter is organized as follows. In sections 2 and 3, we discuss estimation of effects of short-term, time-independent interventions. Specifically, we discuss estimating the effect of a single risk factor intervention on life expectancy or quality-adjusted life expectancy over a specified period of follow-up in a single population, when essentially ideal epidemiological data are available. That is, we assume a random sample of the population *is* randomly assigned to different levels of the risk factor and followed prospectively for a fixed time period. Second, we consider the same study design, except now we are interested in the joint effect of interventions on several risk factors. Third, we consider the problem of extrapolation of the results to longer periods of follow-up and to other populations for which no primary epidemiological data are available. Sections 2 and 3 serve to indicate the possibilities and limitations of even ideal epidemiological data for estimating the effects of multiple time-independent risk factor interventions. In sections 4 and 5 we turn to the main topic of this chapter: the estimation of the effect of multiple time-dependent interventions from observational data, possibly plagued by confounding, selection bias, measurement error, information bias and ill-defined interventions. In sections 6 to 8, we illustrate our methods by attempting to estimate the effects of various time-varying interventions on subjects entered in the Framingham Offspring cohort study. Finally, in section 9 we offer some conclusions.

2. Time-independent interventions

2.1 An ideal intervention

We suppose we have data from a randomized trial of the effect of a one-time short-term intervention (say, an anti-smoking intervention involving one week of intense study of literature on the health consequences of smoking) initiated at calendar time, say 1983. We suppose the subjects in the trial constitute a random sample of the population of a given country, say the United States of America, and these subjects are followed for twenty years. Let $S_0(t)$ and $S_1(t)$ be the survival curves of the exposed and unexposed respectively at t years from the beginning of the trial for $t \leq 20$ years. Thus $S_0(t)$ and $S_1(t)$ are both equal to one at time $t = 0$ and decrease as t increases. Suppose there is no sharing of information in the sense subjects who receive the anti-smoking intervention do not pass on (i.e. infect or contaminate) others in the trial or in the general population with their acquired knowledge. Then it is well known that the area between the survival curves is equal to the expected years of life saved over the first twenty years of the trial due to the intervention. If we are interested in quality-adjusted years of life lost, we use subject-year-specific health and interview data to associate with each year a subject lives a quality measure (taking values between 0 and 1 where 1 indicates optimal quality) and compare expected quality-adjusted years of life lost.

2.2 Multiple ideal interventions

Turn now to estimating the effect of simultaneous interventions on k time-independent risk factors $A = (A_1, \ldots, A_k)$, such as smoking, alcohol, blood pressure, cholesterol, etc., where for the moment we assume all interventions were randomly assigned at the same moment in time and only once. Here A is the k-vector of all risk factors. Each risk factor A_m has some number $|A_m|$ of possible levels $a_{m1}, \ldots, a_{m|A_m|}$. Then A has $|A| = |A_1| \times |A_2| \times \ldots \times |A_k|$ possible joint levels. Let the set $\mathcal{A} = \{a\}$ of size $|A|$ denote the set of possible values a of the vector A. Let $S_a(t)$ be the survival curve for the random sample of subjects assigned to joint level a of the various risk factors. Let $S(t)$ denote the survival curve for a random subset of the population on which no interventions were made. Then the expected years of life gained by intervening and setting each subject's level of the k risk factors to a compared to no intervention is precisely the area between the survival curves $S_a(t)$ and $S(t)$. The optimal intervention a^* is the value of a for which the area under $S_a(t)$ is largest. Further, the loss in life expectancy under intervention a compared to the optimal is simply the difference in area under $S_{a^*}(t)$ and $S_a(t)$. In principle, we need make no assumptions concerning necessary and sufficient causes, multiplicative or additive interactions, or the fraction of deaths attributable to any particular cause to order the health benefits of various joint interventions on life expectancy from such ideal

epidemiological data. All we need is a way to accurately estimate $S_a(t)$ for each joint intervention a in $\mathcal{A}$. Due to random sampling variability, in order to obtain reliable estimates of each of the $|A|$ curves $S_a(t)$ would require an inconceivably large randomized trial if $|A|$ was at all large, since a large number of subjects would need to be randomly assigned to each of the $|A|$ possible levels of A. In practice such large trials are infeasible. As a consequence, we can randomize subjects to only a subset of the possible interventions. In that case we would need to make modelling assumptions as to the nature of the interactions between the risk factors on survival (e.g. by assuming no interaction between risk factors on the mortality rate on a particular scale such as multiplicative or additive) both in order to obtain estimates of the $S_a(t)$ that are not too variable and to extrapolate to values of a outside the range of the data (i.e. to values of a to which no one was randomized). In the final analysis it is the area under the curves $S_a(t)$ that remains of interest. If our models are misspecified (e.g. we assume no interaction on an additive scale when in fact such interaction is present), the resulting estimates of $S_a(t)$ will be biased. Thus we would like to avoid use of models as much as possible. However the use of models cannot be done away with because of our inability to conduct a sufficiently large study.

In the randomized experiment of the previous paragraph, we can also estimate the effects of conditional (or dynamic) interventions. Let $L = (L^{(1)}, \ldots L^{(p)})$ denote a p-vector of measured baseline (pretreatment) covariates such as age, sex and measures of baseline health status. Let $d(l)$ be a function that assigns to each value of the vector L a value of a in the set $\mathcal{A}$ of possible joint risk factor interventions. If a regime d assigns the same value a to each L, we refer to the regime d as non-dynamic. Otherwise, we refer to d as a conditional or dynamic treatment regime, strategy or plan as it individualizes the treatment (i.e. joint risk factor intervention) a subject receives based on the subject's value of L. A wise choice of d should allow us to optimize therapy for individuals and thus should be a better strategy than even the optimal non-dynamic intervention a^* discussed above. Let $S_d(t)$ be the survival curve that would result if we randomly assigned individuals to plan d. For subjects with a given value l of L, the conditional survival curve $S_d(t|L = l)$ given $L = l$ under regime d equals $S_a(t|L = l)$ for the value $a = d(l)$ that they receive under the plan. Thus for the population as a whole $S_d(t) = \Sigma_l S_a(t|L = l)pr(L = l)$ is weighted average of $S_a(t|L = l)$ with $a = d(l)$ and weights proportional to the fraction $pr(L = l)$ of the population with $L = l$. Thus the survival curve $S_d(t)$ of a dynamic regime can be estimated from the data in the randomized trial wherein each subject is randomized to a non-dynamic regime. Define $d_{op}(l)$ to be the treatment plan that minimizes the area under $S_d(t)$ over all possible dynamic and non-dynamic treatment plans d. The loss in life expectancy under intervention d compared to the optimal is the difference in area under $S_{d_{op}}(t)$ and $S_d(t)$. In an ideal world, we would estimate $d_{op}(l)$ from a large ideal

randomized study and, after analysing the trial, treat a new individual with $L = l$ with treatments $d_{op}(l)$.

3. Some limits of ideal data

In this section, we return to the simple randomized trial of section 2.1.

3.1 Extrapolation

Although we have a precise estimate of the effect of this intervention on twenty year mortality of United States citizens in the calendar period 1983–2003, the trial provides no direct evidence concerning the following more relevant policy questions. What would be the continuing effect of this intervention on the United States population through 2013 or 2023? What would be the effect of this same intervention, if it began now in 2003 rather than in 1983? What would be the effect of a similar intervention on a population that differs from the United States population on both measured and unmeasured determinants of mortality including smoking, age, cholesterol, high blood pressure, lifestyle pattern, access to state-of-the-art health care, etc? Obviously any of these questions can only be addressed by assuming a model that extrapolates beyond the observed data. A common simple approach to extrapolation is to first statistically test from the available data whether the relative risk (equivalently, mortality ratio, hazard ratio, rate ratio) in the exposed compared to the non-exposed remains nearly constant both over the 20 years of follow-up and within levels of measured covariates such as age, ethnicity, socioeconomic status and smoking. If the test accepts the constant relative risk hypothesis then we extrapolate by assuming the same will be true if follow-up was continued past 2003, if the intervention was in 2003, and if the intervention was applied to a population with different smoking and risk factor distribution than the United States. In most studies, however, the power to detect deviations from a constant rate ratio is fairly small and there is rarely any strong biological reason to believe that rate ratios rather than other effect measures (such as rate differences) should be constant over time and location. Further, we have no way to test whether the effect of an intervention on a rate ratio scale is the same across groups that differ in unmeasured factors. Finally, even if we assume the relative risk to be constant, nevertheless, to estimate the effect of intervention on life expectancy, we still require a method to estimate covariate-calendar year-specific baseline rates in various populations in future years, since the effect on life expectancy depends both on the relative risk and on these baseline rates.

3.2 Contamination

In the discussion so far, we have assumed that the exposed do not "infect" or "contaminate" the unexposed with their exposure. This

assumption would not hold if subjects given, say, the aforementioned anti-smoking intervention distribute their anti-smoking materials to the control group. In that case, the beneficial effect of the intervention will be underestimated by the difference in the exposed and unexposed survival curves because a number of the unexposed will actually have been exposed. The difficulty is that the trial suffered from noncompliance in the sense that, contrary to our wishes, some of those assigned to no treatment actually received treatment. Certain analytic methods, referred to as instrumental variable methods, can partially adjust for this type of noncompliance if a randomly selected (possibly stratified) subsample of the trial population is interviewed to determine how much treatment they received as measured, say, by the fraction of the anti-smoking articles provided to the treatment group the controls actually read. This approach to correction for noncompliance is useful when the mechanism by which treatment exerts its effects is directly through access to the anti-smoking materials (Robins and Greenland 1996).

However, there are types of contamination that operate in quite different ways. For example, suppose that an effect of the treatment was that one of the treated subjects became so motivated that she started a public health campaign that resulted in the legal prohibition of smoking in public, leading to additional decreases in the cigarette consumption in both the exposed and unexposed. Then the difference in the treatment-arm specific survival curves would underestimate the total effect of the intervention (which should include the indirect effect through the legal prohibition) and correction by instrumental variable methods would not be possible. In this case a different design would be necessary. For example, one could use a cluster randomization design wherein different cities, counties or states are the experimental units and are randomly assigned as a whole to either treatment or control. The goal of the design is to have the experimental units sufficiently isolated from one another that one can be reasonably certain that between-unit contamination of treatment will not occur. If data from an appropriate cluster randomized design are not available, other different, and less reliable approaches to estimating the effect of the intervention on life expectancy must be used. Examples of such approaches include the following: (i) assume death rates would have remained unchanged had the intervention not occurred; (ii) specify and fit complex stochastic models for the mechanism by which the intervention reduced deaths, say, by assuming any decrease in the population exposure to cigarettes was wholly due to the intervention and modelling the effect on mortality of the observed decrease in smoking based on past or current data on changes in smoking and mortality; and (iii) create the observational equivalent of a cluster randomized design by assuming mortality and/or smoking data from other communities can be used to estimate what would have happened in the study population had no intervention taken place. A well known example of the type of contamination we are considering in this paragraph occurs in random-

ized trials studying the effect of vaccines on infectious disease and is the basis for the so-called phenomenon of "herd immunity" wherein an epidemic can be prevented in the unvaccinated (untreated) by vaccinating a sufficiently large fraction of the entire community.

The considerable difficulties caused by contamination and the need for extrapolation will not be considered further in this chapter due to space limitations and to the fact that the main goal of this chapter lies elsewhere. Manton et al. (1992) describe some models that may be used for extrapolation. We now turn to observational settings in which ideal epidemiologic data with which to estimate the effect of various treatments or risk factors even on the population under study may not be available.

4. Observational data and time-independent and time-dependent interventions

For financial, ethical or logistical reasons, randomized trial evidence concerning the effectiveness of many if not most interventions is lacking and data from observational studies must be utilized. In this section we will use a hypothetical observational study of the effect of antiretroviral therapy on the progression of HIV-related disease as a specific example. It is well understood that causal effects can be estimated from observational studies only when data on all relevant time-independent and time-dependent confounding factors have been obtained. What is less well known is that for time-varying treatments, standard approaches to confounder control can be biased, even when the causal null hypothesis of no treatment effect is true and there are no unmeasured confounding factors. Specifically, the standard approach to the estimation of the causal effect of a time-varying treatment on survival has been to model the hazard of failure at t as a function of treatment history with a time-dependent proportional hazards model. Robins et al. (1986) have shown that, even in the absence of unmeasured confounding factors or model misspecification, the usual approach may be biased even under the causal null hypothesis, whether or not one further adjusts for the past history of measured covariates in the analysis, when (i) there exists a time-dependent risk factor (say CD4 lymphocyte count and/or *Pneumocystis carinii* pneumonia [PCP] history) for survival that also predicts subsequent treatment, and (ii) past treatment history predicts subsequent risk factor level. Specifically, condition (i) implies that the analysis that does not adjust for covariates is biased due to confounding by time-dependent risk factors such as CD4 count and/or PCP. Condition (ii) implies that the analysis that includes current CD4 count and/or PCP history as a regressor is biased since it adjusts for a variable (CD4 count and/or PCP history) affected by past treatment. In contrast to standard methods, estimation methods based on marginal structural models (MSMs), the parametric g-computation formula, and structural nested models provide

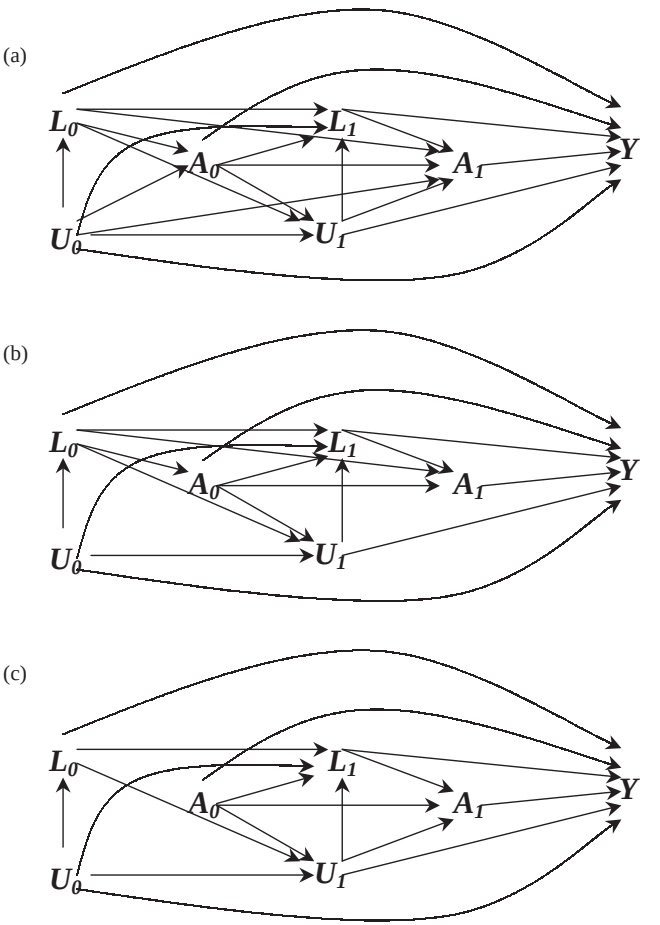
consistent estimates of causal effects whenever unmeasured confounding and model misspecification are absent (see the article, discussion and rejoinder in Robins et al. 1999, for additional details). To describe difficulties with standard approaches and the basis for the effectiveness of these three alternatives, it will be useful to informally introduce causal directed acyclic graphs (DAGs) as discussed by Spirtes et al. (1993), Pearl (1995), and Greenland et al. (1999).

A causal graph is a DAG in which the vertices (nodes) of the graph represent variables, the directed edges (arrows) represent direct causal relations between variables, and there are no directed cycles, since no variable can cause itself. For a DAG to be causal, the variables represented on the graph must include the variables and additional unmeasured variables, such that if any two measured variables on the graph have a cause in common, that common cause is itself included as a variable on the graph.

A variable is a cause of a second variable if there is a directed path from the first variable to the second consisting solely of arrows pointing towards the second variable. As an example consider a follow-up study of AIDS patients. Let $A(t) = A_t$ be the dose of the treatment or exposure of interest, say an antitetroviral therapy, at t with time measured in months since start of follow-up. The time t at which a treatment occurs will either be placed in parentheses or subscripted, depending on the context. Let Y be a dichotomous outcome of interest (e.g. $Y = 1$ if HIV RNA is not detectable in the blood and is 0 otherwise) measured at end of follow-up at time $K + 1$. Our goal is to estimate the causal effect of the time-dependent treatment $A(t)$ on the outcome Y.

Figure 28.1 is a causal graph that represents our study with $K = 1$. In Figure 28.1, $L(t) = L_t$ represents the value at time t of the vector of all measured causal risk factors for the outcome, such as CD4 count, white blood count (WBC), red blood count (RBC), the presence or absence of various opportunistic infections such as PCP, age and weight. We assume that $L(t)$ is temporally prior to $A(t)$ since physicians commonly obtained data recorded in $L(t)$ such as CD4 count before deciding on a treatment $A(t)$ to be given in month t. Similarly, $U(t) = U_t$ represents the value at time t of all unmeasured causal risk factors for Y. Figure 28.1(b) differs from Figure 28.1(a) only in that the arrows from the unmeasured risk factors into the treatment variables have been removed. When, as in Figure 28.1(b), there are no arrows from unmeasured risk factors directly into treatment variables, we say that there are no unmeasured confounders given data on the measured confounders $L(t)$. Figure 28.1(c) differs from Figures 28.1(a) and 28.1(b) in that none of the causal risk factors for Y (measured or unmeasured) has arrows into any treatment variable. Note, however, that earlier treatment $A(0)$ can causally affect later treatment $A(1)$. When, as in Figure 28.1(c), there are no arrows from any (non-treatment) risk factor into any treatment variable, there

Figure 28.1 Causal graphs for a time-dependent exposure

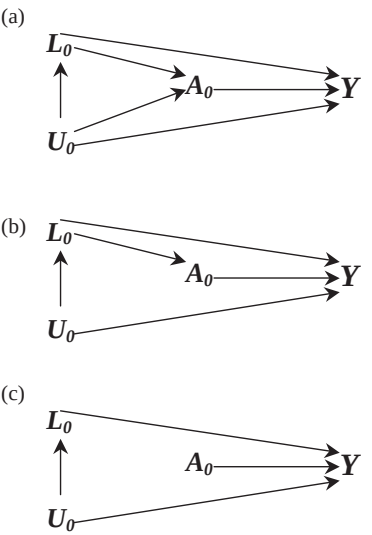


is no confounding by either measured or unmeasured factors, in which case we say that treatment is unconfounded or, equivalently, causally exogenous.

The distinctions drawn above apply equally to more familiar point treatment studies where the treatment is not time-dependent. As indicated in Figure 28.2, a point treatment study is a special case of the general set-up in which $K = 0$. Figures 28.2(a)–28.2(c) are the analogues of Figures 28.1(a)–28.1(c) for a point treatment study.

Our causal DAGs would be of no use without an assumption linking the causal structure represented by the DAG to the statistical data obtained in an epidemiological study. Recall that if a set of variables X is statistically independent of (i.e. unassociated with) another set of

Figure 28.2 Causal graphs for a point exposure A_0



variables Y conditional on a third set of variables Z, then within joint strata defined by the variables in Z, any variable in X is unassociated with any variable in Y. For example, suppose all variables are dichotomous and the set Z consists of the two variables Z_1 and Z_2. Then conditional independence implies that the population odds ratio and population risk ratio between any variable in X and any variable in Y is 1 within each of the $4 = 2^2$ strata of Z: $(Z_1, Z_2) = (0, 0)$, $(Z_1, Z_2) = (0, 1)$, $(Z_1, Z_2) = (1, 0)$ and $(Z_1, Z_2) = (1, 1)$. We use the symbol $\coprod$ to indicate statistical independence, e.g. $X \coprod Y|Z$ means X is conditionally independent of Y given Z. The following so-called causal Markov assumption links the causal structure of the DAG with the various statistical independencies.

4.1 Causal Markov assumption (CMA)

On a causal graph any variable that is not caused by a given variable V will be independent of (i.e. unassociated with) V conditional on (i.e. within joint strata defined by) V's direct causes. Thus, in Figure 28.1(c), $A(t)$ being causally exogenous implies that $A(0) \coprod L(0)$, $U(0)$ and $A(1) \coprod \{L(0), L(1), U(0), U(1)\} \,|A(0)$.

As in any observational study, we cannot determine from the observed data whether there is confounding by unmeasured risk factors. We can only hope that whatever residual confounding there may be due to the $U_1 < y$ is small. However, as discussed further below, under the untestable assumption that there is no unmeasured confounding given the L_k, we

can empirically test from the data whether treatment is causally exogenous. Specifically, a sufficient condition for treatment to be unconfounded is that, at each time k, among subjects with the same past treatment history $A_0, \ldots, A_{k-1}$, the treatment A_k is unassociated with (i.e. statistically independent of) the past history of measured covariates $L_0, \ldots, L_k$. In particular, in our point treatment study, treatment will be causally exogenous if A_0 is unassociated with L_0.

4.2 Inverse probability of treatment weighted estimation

In this section, our goal is to estimate using MSMs the causal effect of the time-dependent treatment $A(t)$ on the mean of Y, which for a dichotomous (0, 1) response is just the probability that $Y = 1$. In this section, we assume that there is no loss to follow-up so Y is observed on each study subject. For any time-dependent variable we use overbars to denote the history of that variable up to and including t. For example, $\overline{L}(t) = [L(0), L(1), L(2), \ldots, L(t)]$ is the covariate process through t. Consider first the association (i.e. regression) model that states that the mean of Y, given treatment history $\overline{A} \equiv \overline{A}(K)$, is a linear logistic function of a subject's duration of antiretroviral therapy. That is,

$$E[Y \mid \overline{A}] = g(\overline{A}; \gamma)$$

where

$$g(\overline{A}; \gamma) = \frac{\exp[\gamma_0 + \gamma_1 Dur(\overline{A})]}{1 + \exp[\gamma_0 + \gamma_1 Dur(\overline{A})]} \tag{1}$$

$Dur(\overline{A}) = \sum_{k=0}^{K} A(k)$ is the subject's duration of treatment in months, and $A(k)$ equals 1 if the subject is on treatment in month k and is 0 otherwise. That is, we are assuming a linear logistic regression model

$$\text{logit } pr\,(Y = 1 \mid \overline{A}) = \gamma_0 + \gamma_1 Dur(\overline{A}) \tag{2}$$

The logistic regression maximum likelihood estimator (MLE) of $\gamma = (\gamma_0, \gamma_1)$ that is computed by all standard packages maximizes $\prod_{i=1}^{n} Lik_i(\gamma)$ with $Lik_i(\gamma) = g(\overline{A}_i; \gamma)^{Y_i}[1 - g(\overline{A}_i; \gamma)]^{1-Y_i}$ being the likelihood contribution for a single subject and n the sample size.

Assuming the association model (1) is correct, when does γ_1 and γ_2 have a causal interpretation? To answer this question, imagine that the decision to administer treatment at each time t was made totally at random by the treating physician. In that hypothetical randomized trial, treatment at time t is not expected to be associated with the history up

to t of any measured or unmeasured prognostic factors (i.e. there is no confounding). In the absence of confounding, association implies causation and we would expect γ_1 to represent the effect of antiretroviral therapy on the mean of Y. More generally, the marginal association between treatment and response represents causation whenever the treatment is causally exogenous, that is, the conditional probability of receiving a treatment $A(t)$ at time t given past treatment and prognostic factor history for Y (measured and unmeasured) depends only on past history of treatment $\overline{A}(t-1)$ as in Figures 28.1(c) and 28.2(c). A more technical definition is provided below after we define counterfactual outcomes. It is well-recognized in the social sciences, econometrics, epidemiological, and biostatistical literature that the treatment parameters of a correctly specified association model will have a causal interpretation if treatment is causally exogenous.

To help assess whether antiretroviral therapy may be causally exogenous, we introduce the concept of "statistical exogeneity". We say that treatment $A(t)$ is a "statistically exogenous or ancillary" process if the probability of receiving treatment at time t does not depend on the history of measured time-dependent prognostic factors up to t, conditional on treatment history prior to t, i.e.

$$\overline{L}(t) \perp\!\!\!\perp A(t) \mid \overline{A}(t-1)$$

Note that a nearly necessary condition for $A(t)$ to be "causally exogenous" is for it to be "statistically exogenous". However, that a process is "statistically exogenous" does not imply it is "causally exogenous", because there may be unmeasured prognostic factors $\overline{U}(t)$ for the outcome (i.e. confounders) that predict the probability of treatment $A(t)$ at time t given past treatment history. Thus we can test from the data whether $A(t)$ is statistically exogenous as it is a relationship between observed variables, but are unable to test whether a statistically exogenous process is causally exogenous. However, as mentioned above and discussed further below, if we make the untestable assumption that there are no unmeasured confounders, then statistical exogeneity will imply causal exogeneity.

Suppose $A(t)$ is discrete and we can correctly model both the probability $f[a(t)|\bar{l}(t), \bar{a}(t-1)]$ of taking treatment $a(t)$ on day t as a function of past treatment $\bar{a}(t-1)$ and measured prognostic factor history $\bar{l}(t)$, and the probability $f[a(t)|\bar{a}(t-1)]$ of taking treatment $a(t)$ in month t as a function only of past treatment $\bar{a}(t-1)$ history. Here we use the convention that random variables (i.e. variables whose values can differ from subject to subject) are denoted by upper case letters. Lower case letters denote possible values of the corresponding random variables. Thus, for example, $f[a(t)|\bar{a}(t-1)]$ is the proportion of subjects in the target population with treatment $A(t)$ equal to $a(t)$ among subjects with past treat-

ment history $\overline{A}(t-1)$ equal to $\overline{a}(t-1)$. We could measure the degree to which the treatment process is statistically non-exogenous through time t by the time t-specific random variable

$$SW(t) = \prod_{k=0}^{t} \frac{f[A(k)|\overline{A}(k-1)]}{f[A(k)|\overline{A}(k-1), \overline{L}(k)]} \tag{3}$$

where $\prod_{k=0}^{t} z(k)$ is the product $z(0) \times z(1) \times \ldots \times z(t)$, $f[A(k)|\overline{A}(k-1), \overline{L}(k)]$ is, by definition, the conditional probability mass function $f[a(k)|\overline{a}(k-1), \overline{l}(k)]$ with $(a(k), \overline{a}(k-1), \overline{l}(k))$ evaluated at a subject's data $(A(k), \overline{A}(k-1), \overline{L}(k))$ and $f[A(k)|\overline{A}(k-1)]$.

For example, if for a given subject $A(k)$ is zero and there are 55 other subjects with the same $\overline{A}(k-1)$, $\overline{L}(k)$ history of whom 25 have $A(k)$ of zero and 70 subjects with the same $\overline{A}(k-1)$ history of whom 32 have $A(k)$ of zero, then $f[A(k)|\overline{A}(k-1), \overline{L}(k)]$ is 20/55 and $f[A(k)|\overline{A}(k-1)]$ is 32/70 for the subject. Informally, the denominator in each term in $SW(t)$ is the probability that a subject received his own observed treatment, $A(k)$, at time k given his past antiretroviral treatment and measured prognostic factor history. Informally, the numerator is the probability that a subject received his observed treatment conditional on his past antiretroviral treatment history, but not further adjusting for his past prognostic factor history. Note that the numerator and denominator of $SW(t)$ are equal for all t with probability 1 if and only if the treatment process is statistically exogenous, i.e. $\overline{L}(t) \coprod A(t)|\overline{A}(t-1)$. In practice $SW(t)$ will have to be estimated from the data but, for pedagogical purposes, assume for now that it is known.

When $A(t)$ is statistically non-exogenous, we shall consider estimating γ by a weighted logistic regression in which a subject is given the weight $SW \equiv SW(K)$. Standard software packages for logistic regression will allow the user to specify the subject-specific weight SW. The weighted logistic regression estimator, which we will refer to as an inverse-probability-of-treatment-weighted (IPTW) estimator, is the maximizer of $\prod_{i=0}^{n} [Lik_i(\gamma)]^{SW_i}$. This weighted logistic regression would agree with the usual unweighted analysis described above just in the case in which $A(t)$ were statistically exogenous. The IPTW estimator is an extension to longitudinal causal inference models of estimators proposed by Horvitz and Thompson (1952).

If the vector of prognostic factors recorded in $\overline{L}(t)$ constitutes all relevant time-dependent prognostic factors (i.e. confounders) so that there are no unmeasured confounders (as in Figures 28.1(b) or 28.2(b)), then,

whether or not the treatment process is statistically exogenous, the weighted logistic regression estimator of γ_1 will converge to a quantity β_1 that can be interpreted as the causal effect of antiretroviral therapy on the mean of Y (on the log odds ratio scale). In contrast, when $A(t)$ is statistically non exogenous, the usual unweighted logistic regression estimator will still converge to γ_1, but now γ_1 will have no causal interpretation. We now give a formal mathematical meaning to the informal concepts of the causal effect of antiretroviral therapy on the mean of Y.

4.3 Counterfactuals and marginal structural models

To formalize our results, we use counterfactual or potential outcomes. Neyman (1923) introduced counterfactual outcomes to analyse the causal effect of time-independent treatments in randomized experiments. Rubin (1978) championed Neyman's idea and emphasized the usefulness of counterfactuals in the analysis of the causal effects of time-independent treatments from observational data. Robins (1986, 1987) proposed a formal counterfactual theory of causal inference that extended Neyman's (1923) time-independent treatment theory to longitudinal studies with both direct and indirect effects and sequential time-varying treatments and confounders. In this theory, for any fixed history of antiretroviral therapy $\bar{a}$, $Y_{\bar{a}}$ is defined to be the random variable representing a subject's outcome had, possibly contrary to fact, the subject been treated with $\bar{a}$ rather than his observed treatment $\overline{A}$. Note that $\bar{a}$ is a possible value of the random variable $\overline{A}$. For each possible history $\bar{a}$ we are assuming a subject's response $Y_{\bar{a}}$ is well defined, although generally unobserved. Indeed we only observe $Y_{\bar{a}}$ for that treatment history $\bar{a}$ equal to a subject's actual treatment history $\overline{A}$, i.e. a subject's observed outcome Y equals $Y_{\overline{A}}$. This identity is the fundamental "consistency" assumption that links the counterfactual data $Y_{\bar{a}}$ to the observed data $(Y, \overline{A})$.

Note that if, at each time t, $A(t)$ can take but one of two values (0 for untreated and 1 for treated) and the study duration is K months, then there are 2^K different $Y_{\bar{a}}$ values associated with each subject as there are 2^K possible treatment patterns only one of which is actually observed for a given subject. Then, formally, the statement that the effect of treatment history on the mean of Y is a linear logistic function of duration of antiretroviral therapy is the statement that, for each $\bar{a}$,

$$E[Y_{\bar{a}}] = g[\bar{a}; \beta]$$

where

$$g[\bar{a};\beta] = \frac{\exp(\beta_0 + \beta_1 Dur(\bar{a}))}{1+\exp(\beta_0 + \beta_1 Dur(\bar{a}))} \tag{4}$$

$\beta = (\beta_0, \beta_1)$, and $Dur(\bar{a}) = \sum_{k=0}^{K} a(k)$ is the subject's duration of treatment under the treatment history $\bar{a}$. We refer to this model as a MSM for the effect of antiretroviral therapy on the mean of Y, since it is a model for the marginal distribution of counterfactual variables and, in the econometric and social science literature, causal models (i.e. models for counterfactual variables) are often referred to as structural.

The parameter β of our MSM encodes the magnitude of the average causal effects of the treatment on the outcome. By definition, the causal effect of treatment regime $\bar{a}$ on the outcome Y for a given study subject is the difference $Y_{\bar{a}} - Y_{\bar{0}}$ between his outcome $Y_{\bar{a}}$ when treated with regime $\bar{a}$ and his outcome $Y_{\bar{0}}$ when never treated. Thus the average causal effect of regime $\bar{a}$ is $E[Y_{\bar{a}} - Y_{\bar{0}}] = E[Y_{\bar{a}}] - E[Y_{\bar{0}}] = g(\bar{a};\beta) - g(\bar{0};\beta)$, which depends on β. If β_1 is zero, we say that there is no effect of treatment $\bar{a}$ on the outcome since $E[Y_{\bar{a}}] - E[Y_{\bar{0}}]$ is the same for all $\bar{a}$. In contrast, the association parameter γ_1 lacks a causal interpretation when treatment is not causally exogenous. Furthermore, the optimal non-dynamic intervention $\bar{a}^*$ is the value of $\bar{a}$ for which $E[Y_{\bar{a}}] = g(\bar{a};\beta)$ is the greatest if our goal is to maximize the probability that HIV will not be detected in the serum.

4.4 Formal definitions of causal exogeneity and no unmeasured confounders

We are now in a position to offer mathematically precise definitions of causal exogeneity and of no unmeasured confounders. *Let* $\{Y_{\bar{a}}\}$ *be the set of all counterfactual outcomes* $Y_{\bar{a}}$ *as* $\bar{a}$ *varies*. Formally, we say that the treatment process $A(t)$ is causally exogenous if,

$$\{Y_{\bar{a}}\} \perp\!\!\!\perp A(t) \,|\, \overline{A}(t-1) \tag{5}$$

which is mathematically equivalent to the statement that $\{Y_{\bar{a}}\}$ is independent of $\overline{A}$. Note that even when $A(t)$ is "causally exogenous", if the treatment has an effect on the outcome, then the observed outcome $Y = Y_{\overline{A}}$ will not be independent of $\overline{A}$, since $Y_{\overline{A}}$ is a function of a subject's observed treatment history $\overline{A}$ itself. Given the covariates recorded in $L(t)$, following Robins (1987) we say there are no unmeasured confounders for the effect of $A(t)$ on Y if

$$\{Y_{\bar{a}}\} \perp\!\!\!\perp A(t) \,|\, \overline{A}(t-1), \overline{L}(t) \tag{6}$$

We shall also refer to the assumption of no unmeasured confounders as

the assumption that treatment $A(t)$ is sequentially randomized given the past. This assumption generalizes Rosenbaum and Rubin's (1983) assumption of ignorable treatment assignment to longitudinal studies with time-varying treatments and confounders. The assumption states that, conditional on treatment history and the history of all recorded covariates up to t, treatment at t is independent of the counterfactual random variables $Y_{\bar{a}}$. This will be true if all prognostic factors for (i.e. predictors of) Y that are used by physicians to determine whether treatment is given at t are recorded in $\overline{L}(t)$ and $\overline{A}(t-1)$. That is, as in Figure 28.1(b), the causal graph generating the data has no arrows directly from any unmeasured causal risk factors for Y directly into treatment. For example, since physicians tend to administer prophylaxis to subjects with previous bouts of PCP, and in untreated subjects PCP predicts Y, the assumption of no unmeasured confounders would be false if $\overline{L}(t)$ does not contain PCP history. It is the primary goal of the epidemiologists conducting an observational study to collect data on a sufficient number of covariates to ensure that the assumption of no unmeasured confounders will be at least approximately true.

In an observational study, the assumption of no unmeasured confounders cannot be guaranteed to hold even approximately and it is not subject to empirical test. Therefore, it may be useful to investigate the sensitivity to violations of the assumption through a formal sensitivity analysis. Robins et al. (1999, rejoinder) and Robins et al. (1999) provide details.

Robins (1999) proved that when there are no unmeasured confounders, (i) statistical exogeneity implies causal exogeneity, (ii) the weighted logistic regression estimator using the weights SW converges to the parameter β of the MSM (4) for $E[Y_{\bar{a}}]$, and (iii) the probability limit γ of the usual unweighted logistic estimator generally differs from the causal parameter β of the MSM unless the treatment process is statistically exogenous. Here we provide an informal heuristic argument for (ii). View each person as a member of a pseudo-population consisting of SW copies of themselves. In this new pseudo-population, it can be shown that $\overline{L}(t)$ does not predict treatment at t given past treatment history, and thus we have created a pseudo-population in which treatment is causally exogenous. Furthermore, the causal effect of treatment on Y in the pseudo-population is the same as in the original population. That is, if $E[Y_{\bar{a}}] = g(\overline{a}; \beta)$ in the true population, the same will be true of the pseudo-population. Hence, we would like to do ordinary logistic regression in the pseudo-population. But that is what our weighted logistic regression estimator is doing, since the weights create, as required, SW copies of each subject.

We can generalize our MSM (4) slightly and model the marginal distribution of $Y_{\bar{a}}$ within levels of a subset V of the pre-treatment (baseline) covariates $L(0)$. Then, our marginal structural logistic model (4) could

be modified to

$$E[Y_{\bar{a}}|V] = \frac{\exp(\beta_0 + \beta_1 Dur(\bar{a}) + \beta_2' V + \beta_3' Dur(\bar{a})V)}{1 + \exp(\beta_0 + \beta_1 Dur(\bar{a}) + \beta_2' V + \beta_3' Dur(\bar{a})V)}$$

β'_3 measures how the magnitude of the effect of $Dur(\bar{a})$ is modified by the pretreatment covariates V. An IPTW estimator of the parameter β can be obtained by weighted logistic regression with weights SW except now the logistic model includes $Dur(\bar{A})$ and V as regressors, and $SW(t)$ is redefined to be

$$SW(t) = \prod_{k=0}^{t} \frac{f[A(k)|\bar{A}(k-1), V]}{f[A(k)|\bar{A}(k-1), \bar{L}(k)]}. \quad (7)$$

Note V is already included in the denominator, since V is a subset of the variables in $L(0)$.

Let $d(V)$ be a function (regime) that assigns to each value of the vector V a value of $\bar{a}$ in the set V of possible interventions. If a regime d assigns the same value $\bar{a}$ to each V we refer to the regime d as non-dynamic. Otherwise, we refer to d as a conditional or baseline-dynamic treatment regime, strategy or plan as it individualizes the treatment history a subject receives based on the subject's value of the baseline variables recorded in V, a subset of $L(0)$. A wise choice of d should allow us to optimize therapy for individuals and thus should be a better strategy than even the optimal non-dynamic intervention a^*. Let $E[Y_d]$ be the probability of being without HIV in the serum if all subjects followed plan d. For subjects with a given value v of V, the conditional expectation $E[Y_d|V = v]$ given $V = v$ under regime d equals $E[Y_{\bar{a}}|V = v]$ for the value $\bar{a} = d(v)$ that they receive under the plan. Thus for the population as a whole $E[Y_d] = \Sigma_v E[Y_{\bar{a}}|V = v]pr(V = v)$ is a weighted average of $E[Y_{\bar{a}}|V = v]$ with $a = d(v)$ and weights proportional to the fraction $pr(V = v)$ of the population with $V = v$. Then $d_{op}^{baseline}(v)$ is the treatment plan that maximizes $E[Y_d]$ over all possible baseline-dynamic and non-dynamic treatment plans d. Now even $d_{op}^{baseline}(v)$ only allows one to optimize treatment history $\bar{a}$ based on pretreatment (baseline) variables. However, with time-varying treatments it is usually important to dynamically choose the treatment at each time t based on a subject's entire covariate history up to time t. For example, consider drug treatment for a chronic disease. When a drug becomes toxic to a subject, the optimal strategy is to stop the drug (or reduce the dose) at least temporarily. One cannot know when to stop the drug based on baseline covariates. Rather, the optimal treatment strategy must allow treatment decisons to be based on a subject's evolving covariate history. The best methods for estimating the effects of true dynamic regimes are not based on MSMs. Further discussion is provided in section 5.

4.5 Marginal structural Cox proportional hazards model

MSMs can easily be extended to failure time outcomes by specifying a marginal structural Cox proportional hazards model such as

$$\lambda_{T\bar{a}}(t|V) = \lambda_0(t)\exp(\beta_1 a(t) + \beta_2' V + \beta_3' a(t)V), \tag{8}$$

where $T_{\bar{a}}$ is the subject's time to death if he had followed anti-retroviral therapy history $\bar{a}$, $\lambda_{T\bar{a}}(t|V)$ is the hazard (force of mortality) of $T_{\bar{a}}$ at t conditional on having pretreatment variables V, $\lambda_0(t)$ is an unspecified baseline hazard function, $\exp(\beta_1 + \beta_3' V)$ is the causal rate ratio for the effects of treatment at level V of a vector of baseline regressors including age, calendar year, CD4 count, CD8 count, WBC count, RBC count, platelets, etc. For variety, we have chosen a model which specifies that the hazard of death at time t depends on current treatment status rather than the duration of treatment. Other dose–response models could be used.

Let T be a subject's observed failure (i.e. death) time, so that $T = T_{\bar{A}}$. Arguing as above, Robins (1999) shows that, in the absence of censoring, a consistent estimator of the unknown parameter $\beta = (\beta_1, \beta_2', \beta_3')'$ is obtained by fitting the ordinary time-dependent Cox model

$$\lambda_T(t|\bar{A}(t), V) = \lambda_0(t)\exp(\gamma_1 A(t) + \gamma_2' V + \gamma_3' A(t)V) \tag{9}$$

except that the contribution of a subject to a calculation performed on a subject i at risk at time t is weighted by $SW_i(t)$, as defined in (7) with $T > k$ added to the conditioning event. Note the subject-specific weights change with time. Few standard Cox proportional hazards software programs allow for time-varying weights. To avoid this software problem one can fit a weighted pooled logistic regression treating each person-month as an observation and allowing for a time-dependent intercept. That is, one can fit, by weighted logistic regression using weights $SW(t)$, the model

$$\begin{aligned}\text{logit}\, pr[D(t) = 1|D(t-1) = 0, \bar{A}(t-1), V] \\ = \gamma_0(t) + \gamma_1 A(t-1) + \gamma_2' V + \gamma_3' A(t-1)V\end{aligned} \tag{10}$$

where $D(t) = 0$ if a subject was alive at time t and 1 if the subject died at month t, and $\gamma_0(t)$ is a time-specific intercept. This method offers the advantage of being easily programmed in any standard statistical package. Under our assumptions we thereby obtain a consistent estimator of the parameter vector β of the MSM

$$\begin{aligned}\text{logit}\, pr[D_{\bar{a}}(t) = 1|D_{\bar{a}}(t-1) = 0, V] \\ = \beta_0(t) + \beta_1 a(t-1) + \beta_2' V + \beta_3' a(t-1)V\end{aligned} \tag{11}$$

When the death rate in any given month t is small, the parameters of (11) and (8) closely approximate one another.

Because of the weights, the standard error estimates outputted by a standard logistic program are invalid and may be either too large or too small. To overcome this difficulty, model (10) should be fit using a generalized estimating equations (GEE) (Liang and Zeger 1986) program which outputs robust variance estimators. The robust variance GEE estimators provide a conservative confidence interval for the β (Robins, 1999). That is, the 95% Wald confidence interval calculated as $\beta \pm 1.96 \times$ (robust) standard error is guaranteed to cover the true β at least 95% of the time in large samples.

We now describe how to accommodate censoring in the analysis. We defined a subject as right censored at time t (i.e. $C(t) = 1$) if by time t he either dropped out of the study or reached administrative end of follow-up alive.

We say that censoring is ignorable or non-informative if the conditional cause-specific hazard of being censored at k among subjects alive and uncensored up to k does not depend on the failure times $T_{\bar{a}}$ given $\overline{A}(k-1)$, and the time-dependent covariate $\overline{L}(k-1)$ history prior to k. Under the assumptions of ignorable censoring and no unmeasured confounding, Robins (1999) shows that we still obtain from fitting (10) consistent estimators of β if we weight a subject alive and uncensored at month t by $SW(t) \times SW^{\dagger}(t)$ where (i)

$$SW^{\dagger}(t) = \prod_{k=0}^{t} \frac{pr[C(k) = 0 | \overline{C}(k-1) = 0, \overline{A}(k-1), V, T > k]}{pr[C(k) = 0 | \overline{C}(k-1) = 0, \overline{A}(k-1), \overline{L}(k), T > k]}$$

is informally the inverse of the ratio of a subject's probability of remaining uncensored up to month t divided by that probability calculated as if there had been no time-dependent determinants of censoring except past treatment history and V, and (ii) we modify our definition (7) of $SW(t)$ to add $C(k) = 0$ to the conditioning events both in the numerator and the denominator. The denominator of $SW(t) \times SW^{\dagger}(t)$ is informally the probability that a subject would have his own observed treatment and censoring history through month t.

Hernan et al. (2000) describe how to estimate the weights $SW(t) \times (SW^{\dagger}(t)$ from the data using pooled logistic regression models with treatment (censoring) at each time k as the response variable. Substituting out estimated weights into our IPTW estimators allows us to estimate, under the assumption of no unmeasured confounders, logit $pr\,[D_{\bar{a}}(t) = 1 | D_{\bar{a}}(t-1) = 0, V] = \beta_0(t) + \beta_1 a(t-1) + \beta_2' V + \beta_3' a(t-1) V$ and thus the conditional survival curves $S_{\bar{a}}(t|V) = \prod_{k=0}^{t} pr[D_{\bar{a}}(k) = 0 | D_{\bar{a}}(k-1) = 0, V]$ that would be observed if all subjects have followed regime $\bar{a}$.

Again let $d(V)$ be a function (regime) that assigns to each value of the

baseline vector V a value of $\bar{a}$ in the set V of possible interventions. Let $S_d(t)$ be the survival curve if all subjects followed plan d. For subjects with a given value v of V, the conditional survival curve $S_d(t|V = v)$ given $V = v$ under regime d is $\Sigma_v S_{\bar{a}}(t|V = v)pr(V = v)$ is a weighted average of $S_{\bar{a}}(t|V = v)$ with $a = d(v)$ and weights proportional to the fraction $pr(V = v)$ of the population with $V = v$. Then $d_{op}^{baseline}(v)$ is the treatment plan that minimizes the area under $S_d(t)$ over all possible baseline-dynamic and non-dynamic treatment plans d. Now even $d_{op}^{baseline}(v)$ only allows one to optimize treatment history $\bar{a}$ based on pretreatment (baseline) variables. The best methods for estimating the effects of general dynamic regimes are not based on MSMs. Further discussion is provided in section 5.

We have seen that under the assumption of no unmeasured confounders, IPTW estimation of a marginal structural Cox proportional hazards model can, in contrast with standard methods, be used to estimate the effect of time-varying treatments on survival.

The correctness of the resulting causal inferences is dependent on three key assumptions. First, we must assume that the covariates in $L(t)$ are sufficient to adjust for both confounding and for selection bias due to loss to follow-up. This implies that we have available, in each month, data recorded in $L(t)$ on the history of all time-dependent covariates that (i) are independent predictors of death and (ii) independently predict the probability of changes in treatment and/or of being censored in that month. As in all observational studies, this fundamental assumption cannot be empirically tested. In practice, this would never be precisely or sometimes even approximately true. As described earlier, methods have recently been developed which allow one to evaluate the sensitivity of one's estimates to increasing violation of this assumption.

Second, we must assume that our models for changes in treatment and censoring, given past covariate and treatment history, are correctly specified. Last, we need to assume that our MSM for the effect of antiretroviral therapy on mortality is correctly specified.

Even when estimating the effect of a time-independent treatment using standard statistical models, the same assumptions (no unmeasured confounders, non informative censoring, and no model misspecification) are required to endow the parameters with a causal interpretation. Furthermore, when estimating the effect of a time-varying treatment, our assumptions are less restrictive than those required by standard analyses: an approach based on IPTW estimation of MSMs does not require for validity the absence of confounding by time-dependent covariates.

5. Alternatives to MSMs

Before introducing MSMs, Robins and coauthors introduced two

other methods for estimation of the causal effect of a time-varying treatment in the presence of time-varying confounders: the parametric g-computation algorithm formula estimator (Robins 1986), and g-estimation of structural nested models (Robins et al. 1992). When (i) both treatment and the confounders are discrete variables, (ii) they are measured at only a few time points, and (iii) the sample size is large, then estimation can be carried out using fully saturated models (i.e. non-parametrically) and all three methods are precisely equivalent. They differ when, as in observational studies with sparse multivariate data, one must introduce modelling assumptions.

A major advantage of MSMs is that they resemble standard models. For example, the logistic MSM and the Cox proportional hazards MSM described above are the natural way to extend the ordinary logistic and time-dependent Cox models to allow for estimation of causal effects of time-dependent treatments. However a major advantage of the parametric g-computation algorithm formula estimator and g-estimation of structural nested models over MSMs is that these models are much more useful than MSMs for estimating both interactions between treatment and time-dependent covariates and the effect of dynamic treatment regimes (Robins 1999). Due to space limitations, we will focus in this chapter on the parametric g-computation algorithm formula estimator, and we will not consider g-estimation of structural nested models.

Now for any variable Z, let $\mathcal{Z}$ be the support (i.e. the possible values) of Z. Define a treatment regime or plan $\underline{d}$ to be a collection of $K+1$ functions $d = \{d_0, \ldots, d_K\}$ where d_m: $\bar{\mathcal{L}}_m \rightarrow \mathcal{A}_m$ maps histories $\bar{l}_m \in \bar{\mathcal{L}}_m$ into a treatment $d_m(\bar{l}_m) \in \mathcal{A}_m$. If $d_m(\bar{l}_m)$ is a constant, say a_m, not depending on $\bar{l}_m$ for each m, we say regime d is non-dynamic and write $d = \bar{a}$, $\bar{a} \equiv (a_0, \ldots, a_K)$. Otherwise, d is dynamic. We let $\mathcal{D}$ be the set of all regimes d. Let $f(o)$ and $F(o)$ represent the density and distribution function of the observed data $O = (\bar{A}_K, \bar{L}_{K+1})$.

Associated with each regime d is the distribution $F_d(o)$ with density $f_d(o)$ that represents the distribution of the observed data had, contrary to fact, all subjects in the population been treated with regime d. Suppose the assumption of no unmeasured confounders holds for the joint outcomes $\bar{L}_{\bar{a}} = \bar{L}_{\bar{a},K+1}$ [i.e. Equation 6 holds with $\{\bar{L}_{\bar{a}}\}$ replacing $\{\bar{Y}_{\bar{a}}\}$. Then given a regime $d = (d_0, d_1, \ldots, d_K)$ and the joint density

$$f(o) = f(l_0) \times f(a_0|l_0) \times \ldots \times f(a_K|\bar{l}_K, \bar{a}_{K-1}) \times f(l_{K+1}|\bar{l}_K, \bar{a}_K), \qquad (12)$$

of the observed data (written as a Markov factorization in temporal order), $f_d(o)$ is the density $f(o)$ except that, in the factorization (12), $f(a_0|l_0)$ is replaced by a degenerate distribution at $a_0 = d_0(l_0)$, $f(a_1|l_1, a_0, l_0)$ is replaced by a degenerate distribution at $a_1 = d_1(l_0, l_1)$, and, in general, $f(a_k|\bar{l}_k, \bar{a}_{k-1})$ is replaced by a degenerate distribution at $a_k = d_k(\bar{l}_k)$. That

is the density $f_d(o)$ is given by

$$f_d(o) = f(l_0) \times f_d(a_0|l_0) \times \ldots \times f_d(a_K|\bar{l}_K, \bar{a}_{K-1}) \times f(l_{K+1}|\bar{l}_K, \bar{a}_K), \qquad (13)$$

where $f_d(a_k|\bar{l}_k, \bar{a}_{k-1})$ is equal to 1 if $a_k = d_k(\bar{l}_k)$ and $f_d(a_k|\bar{l}_k, \bar{a}_{k-1})$ is equal to 0 otherwise.

Again suppose the outcome of interest is L_{K+1} which is assumed to be univariate and shall be denoted by Y. In the following, let $d(\bar{l}_k) \equiv (d_0(\bar{l}_0), \ldots, d_k(\bar{l}_k))$ and $d_k(\bar{l}_k)$ denote values of $\bar{A}_k$ and A_k respec-

tively. Then the chance of having history $(y, \bar{l}_K)$ under the distribution $F_d(o)$ is $f_d(y, \bar{l}_K) = f_d(y|\bar{l}_K, d(\bar{l}_K))\Pi_{j=0}^{K} f(l_j|\bar{l}_{j-1}, d(\bar{l}_{j-1}))$ since the terms $f_d(a_j|\bar{l}_j, \bar{a}_{j-1})$ are equal to 1 in expression (13) for $f_d(o)$. Thus the chance $f_d(y)$ of having $Y = y$ under $f_d(\cdot)$ is

$$\begin{aligned} f_d(y) &= \int f_d(y, \bar{l}_K) d\mu(\bar{l}_K) \\ &= \int \{f(y|\bar{l}_K, d(\bar{l}_K)) \times \prod_{j=0}^{K} f(l_j|\bar{l}_{j-1}, d(\bar{l}_{j-1}))\} d\mu(l_j) \end{aligned} \qquad (14a)$$

where $d\mu(\bar{l}_K)$ represents integration (or summation if $\bar{l}_K$ is discrete) over all possible $\bar{l}_K$ histories. That is if $\bar{l}_K$ is discrete we obtain

$$\begin{aligned} f_d(y) &= \sum_{\bar{l}_K} f_d(y, \bar{l}_K) \\ &= \sum_{\bar{l}_K} \{f(y|\bar{l}_K, d(\bar{l}_K)) \times \prod_{j=0}^{K} f(l_j|\bar{l}_{j-1}, d(\bar{l}_{j-1})) \end{aligned} \qquad (14b)$$

with $\bar{l}_j$ the initial segment of a given $\bar{l}_K$ history. Robins (1986) referred to this formula as the g-computation algorithm formula or functional for the effect of regime d on outcome Y. Similarly, the marginal distribution function of Y under $F_d(\cdot)$ is

$$F_d(y) = \int \ldots \int pr[Y < y|\bar{l}_K, d(\bar{l}_K)] \times \prod_{j=0}^{K} f(l_j|\bar{l}_{j-1}, d(\bar{l}_{j-1})) d\mu(l_j)$$

To estimate $f_d(y)$ from the observed data we specify parametric models $f(l_j|\bar{l}_{j-1}, \bar{a}_{j-1})$ and $f(y|\bar{l}_K, \bar{a}_K)$, fit the models to the observed data to obtain estimates $\hat{f}(l_j|\bar{l}_{j-1}, \bar{a}_{j-1})$ and $\hat{f}(y|\bar{l}_K, \bar{a}_K)$ and evaluate these estimates at $\bar{a}_{j-1} = d(\bar{l}_{j-1})$ and $\bar{a}_K = d(\bar{l}_K)$, and substitute them into the above expression for $f_d(y)$. Details are given in section 6 below where we consider

estimating the effect of treatment on a survival outcome. Under certain additonal assumptions described in section 6, one may consider the g-formula as giving the overall effect of an intervention regime d on Y by summing up both its indirect effects on Y mediated through the effect of the intervention on the non-intervened on variables L_j (which in turn affect the outcome Y) and its direct effects on Y. In contrast, MSMs always model the overall effect of the intervention on Y without any decomposition into direct and indirect effects.

6. Analysis of the framingham offspring study

In this section we describe the use of the parametric g-computation algorithm formula estimator of section 5 and IPTW MSM estimator of section 4 to estimate what the cumulative incidence of coronary heart disease mortality in the Framingham Offspring Study would have been

Content	*Time after follow-up*	*Sample size*
Exam 1	Year 0	5124
Exam 2	Year 8	3863
Exam 3	Year 12	3873
Exam 4	Year 16	4019

under various hypothetical intervention strategies. The Framingham Offspring public use file contained data on 5124 subjects (2483 male and 2641 female subjects). Subjects were examined five times over 20 years. Exam 1 occured in year 0, exam 2 in year 8, exam 3 in year 12, exam 4 in year 16 and exam 5 in year 20. The table below shows the number of subjects attending each of the first four exams. Mortality follow-up was essentially complete through exam 5 at year 20. By year 20, 891 subjects were known to have developed coronary heart disease (CHD), including those with pre-existing CHD at exam 1.

For reasons explained below, the follow-up period for our analysis began at exam 2 (which we will refer to as the baseline exam). Each subject contributed person-time from exam 2 to the date of CHD diagnosis, death from any cause, loss to follow-up, or exam 5, whichever occurred first. Subjects with pre-existing CHD at exam 2 or with missing risk factor values at exam 1 were excluded from all calculations.

In the Framingham public use file, CHD was defined as one of the following events: (i) recognized or unrecognized myocardial infarction (using diagnostic ECG or transaminase/history/autopsy evidence), (ii) angina pectoris (first episode), (iii) coronary insufficiency, or (iv) death from CHD. The following (time-dependent) risk factors were used in various analyses: age at exam (years); sex (male/female); body mass index

(kg/m^2); cigarette smoking (current smoker/non-current smoker); alcohol consumption (calculated alcohol index in ounces per week); diabetes mellitus (defined diabetes/no defined diabetes); LDL-cholesterol (mg/dl); HDL-cholesterol (mg/dl); and systolic blood pressure (mmHg).

Missing values of risk factors in later exams were carried forward from the previous exam. Fat intake was not used in our analyses, because the Framingham public use file did not contain that information. Physical activity was also not used, because it was not measured at exams 1 and 4.

6.1 Interventions

We estimated with the parametric g-computation algorithm formula estimator the cumulative incidence (i.e. risk) of CHD between exam 2 and exam 5 that would have been observed under the following hypothetical intervention strategies.

1. A random half of smokers at baseline (i.e. exam 2) quit smoking forever (no intervention on those who were not current smokers at baseline but started smoking later).
2. All smokers at baseline quit smoking forever (no intervention on those who were not current smokers at baseline but started smoking later).
3. Subjects who were drinkers at baseline were randomly assigned a non-time-varying level of alcohol consumption at baseline drawn at random from a truncated normal distribution with a mean of 20 g/day for males and 10 g/day for females, a maximum of 30 g/day for males and 15 g/day for females, and a minimum of 10 g/day for males and 5 g/day for females (with the sex-specific variance of the normal equal to the sex-specific variance of alcohol consumption at baseline among drinkers in study population). Alcohol consumption was not intervened on thereafter.
4. Body mass index (BMI) was lowered
   a) by 10% at each visit BMI exceeded 22
   b) to 22 at each visit BMI exceeded 22
5. The distribution of LDL at baseline was drawn at random from a normal distribution with mean of 90 and a SD of 30 (Chinese distribution both for males and females). LDL was not intervened on thereafter.
6. Interventions 1, 3 and 4a simultaneously.

6.2 Data analysis

Parametric g-formula

In this section we describe how we estimate the proportion of the Framingham Offspring Study population that would have developed CHD between exams 2 and 5 for each of the interventions mentioned above under the assumption of no unmeasured confounders. Specifically, we used the parametric g-formula (i.e. the g-formula evaluated at predicted values calculated under parametric regression models) to estimate the CHD risk in the Framingham Offspring data under each intervention. We remark that in this exercise we are estimating the cumulative incidence (probability of developing) CHD between exams 2 and 5 under each intervention rather than the expected life years and expected quality-adjusted life-years. Although these latter quantities are more relevant for public health decision-making, we decided to report the results in terms of cumulative incidences because this measure is more familiar to practising epidemiologists. It would be easy to use these same methods to estimate expected life measures. The g-formula to compute the CHD risk (i.e. cumulative incidences between exams 2 and 5) under a particular intervention (i.e. a possibly dynamic regime) d is one minus the probability of "not developing CHD between exams 2 and 5" which is given by the following g-formula for "survival without CHD."

$$\begin{aligned}
&\sum_{\bar{a}_4}\sum_{\bar{a}_4^*}\sum_{\bar{l}_4}\prod_{j=1}^{4}\Pr\left[CHD_{j+1}=0|\bar{l}_j,\bar{a}_j,\overline{CHD}_j=0,\overline{D}_j=0\right]^{I(j>2)}\\
&\quad\times f_d\left(a_j|\bar{l}_j,a_j^*,\bar{a}_{j-1},\overline{CHD}_j=0,\overline{C}_j=0,\overline{D}_j=0\right)\\
&\quad\times f\left(l_j|\bar{l}_{j-1},\bar{a}_{j-1},\overline{CHD}_j=0,\overline{C}_j=0,\overline{D}_j=0\right)\\
&\quad f\left(a_j^*|\bar{l}_j,\bar{a}_{j-1},\overline{CHD}_j=0,\overline{C}_j=0,\overline{D}_j=0\right)
\end{aligned} \tag{15}$$

where j denotes exam number for any $\bar{z}_j$, $\bar{z}_0$ is defined to be 0 since there is no exam 0.

CHD_{j+1} = 1 if CHD was diagnosed between exams j and j + 1, 0 otherwise;

the overbar means history, i.e. $\overline{CHD}_j = (CHD_1, \ldots, CHD_j)$;

the random variable L_j with realized values l_j is the set of risk factors not undergoing the intervention;

a_j is the intervention value of the risk factors A_j undergoing intervention at time j;

$\bar{l}_j$ is a history of (non-intervened on) risk factors through exam j compatible with a particular $\bar{l}_4$ history in the sum;

$\bar{a}_j$ is a history of (intervened on) risk factors through exam j;

a_j^* is the value of A_j that would be observed at time j if interventions were made through $j - 1$ but no intervention was made at j (see below);

the sum is overall possible $\bar{l}_4$, $\bar{a}_4$, $\bar{a}_4^*$ histories;

$\overline{C}_j = 0$ is the event that a subject remains uncensored through exam j;

$\overline{D}_j = 0$ is the event that a subject has not died from other non-CHD causes through exam j;

$f(\cdot|\cdot)$ is a conditional density function of the observed data distribution;

$f_d(\cdot|\cdot)$ is the conditional density function of the exposure A under the proposed intervention. Because we do not intervene until $j = 2$, the "intervened on" value a_1 of A_1 must equal the "non-intervened on" value a_1^*. To accomplish this, we take for $j = 1$ $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0) = 1$ if $a_j = a_j^*$ and 0 otherwise.

Formula (15) includes four generalizations of formula (14). First we allow the intervention treatment a_j at time j to depend on the value a_j^* of A_j that would be observed at time j if the planned interventions were made through $j - 1$ but no intervention was made at j. That is the intervention density $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ can be a function of a_j^*. This is meant to reflect the fact that when a subject arrives at visit j, his exposure a_j^* could be noted and if it takes certain values it can, in principle, be intervened on and instantaneously changed to a new exposure labelled a_j. We assume it is a_j that affects outcomes at the next visit $j + 1$. For example, the BMI intervention 4b was modelled in this way. This "instantaneous change" model might be a reasonable approximation to a real intervention in which BMI was reduced to 22 over 3 to 6 months, as the time between visits is four years. To determine at time j the value of a_j^* for a simulated subject, we need to generate a_j^* from the non-intervention observed data density $f(a_j^*|\bar{l}_j, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$, where it follows from the definition of a_j^* that the arguments $(\bar{l}_j, \bar{a}_{j-1})$ in the conditioning event are the values that would be seen had the planned intervention d been made through $j - 1$. Second, we now allow probabilistic (i.e. random) interventions so $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ need not take only the values 0 or 1. For example, an intervention in which a subject who is a smoker at occasion j (i.e. $a_j^* = 1$) stops smoking with probability 1/2 would have $f_d(a_j|\bar{l}_j, a_j^* = 1, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0) = 1/2$ for both $a_j = 1$ and $a_j = 0$. Third, we now allow the interventions such as the alcohol intervention 3 and LDL intervention 5 to include active intervention at some times and no further intervention at other times. For example, in intervention 5, $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ is equal to a normal density with mean 90 and standard deviation 30 at the baseline visit $j = 2$, but $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0) = 1$ if $a_j = a_j^*$ at visits $j = 3$ and $j = 4$ in addition to visit $j = 1$. Finally, the end-point is now survival to visit 5, so the g-formula multiplies the probability of not failing at visits 3, 4 and 5.

Caveats

As discussed above, the g-formula identifies the effect one would have

observed under a particular intervention under the assumption of no unmeasured confounders. But this assumption itself assumes that the counterfactual survival variables T_d (the time to CHD under intervention d) are well defined. If the intervention is on smoking, it seems only mildly philosophically problematic to assume the existence of a well-defined counterfactual time of CHD had one received a different smoking history than one's actual history. However, for interventions 4a, 4b or 5, the counterfactual T_d may be very poorly defined. For example, the effect on CHD of lowering LDL may possibly depend on the mechanism by which it is lowered. That is, the effect may be different depending on whether LDL were lowered by using drugs that decrease LDL production, drugs that increase its elimination or drugs that impair the absorption of fats from the gut. Similarly the effect on CHD of an intervention to lower BMI may be different depending on the degree of calorie restriction imposed, the foods allowed in the required diet, and the level of increased physical activity imposed. Thus for the interventions 4a, 4b and 5d that are stated solely in terms of the level of BMI or LDL to be achieved (without further specifying the precise intervention that will be used to realize the reductions), it may be that the counterfactuals T_d should not be regarded as well-defined. But if the counterfactulas are not well-defined, the assumption of no unmeasured confounders is vacuous (as the assumption is in terms of the counterfactuals). As a consequence, for such interventions, it is unclear what the counterfactual quantity one hopes the g-formula to estimate. However, if one correctly held the belief that LDL in the blood is a dominant cause of CHD and that all reasonable interventions that lowered LDL to a prespecified level would result in equal reductions in CHD risk, then T_d would be well-defined even for an intervention d that, like intervention 5, is stated only in terms of the level of LDL to be achieved. Evidence for this belief would be strenghtened if randomized experiments of various lipid lowering therapies that lowered average LDL by an identical amount showed the same improvement in CHD risk, regardless of the mechanism of action of the drug used to lower LDL.

Even when a counterfactual is well-defined, the g-formula only estimates the effect of an intervention under the assumption of no unmeasured confounders. Thus it is important for epidemiologists to try to obtain data on many potential confounders. The assumption of no unmeasured confounders will never exactly be true in observational studies and may not even be approximately true. A particular setting in which substantial unmeasured confounding may be present is when many of the confounders recorded in L_j are measured with error. In that case, even if the correctly measured confounders would have served to fully control confounding, the mismeasured confounders may not. This could be the case even when the measurement error is random and nondifferential and the null hypothesis of no treatment effect is true. This

raises an important distinction between studies with time-varying treatments and studies with time-independent treatments. With time-varying treatments, past treatment history itself can be a confounder for the effect of current treatment on the outcome. Thus even when treatment is only subject to random (non-differential) measurement error, confounding bias can exist even under the null hypothesis of no treatment effect. This is in sharp contrast with the results for time-independent treatments in which random (non-differential) subject-specific measurement error does not lead to bias under the causal null hypothesis of no treatment effect. However, there is an additional subtlety in studies with time-varying treatments. If the decision to take a treatment at time j depends on one's past measured treatment history (say as recorded in a pharmacy prescription database) and not on one's true past treatment history, then, under the null hypothesis, there is no bias introduced by the mismeasurement of treatment, even if the measurement error is differential. On the other hand, if the decision to take treatment at a given time depends on past true (but unrecorded) treatment history, then, by only controlling for past measured treatment history in the analysis, bias may exist under the null hypothesis of no treatment effect, even with random (non-differential) measurement error.

In sections 4 and 5, we assumed the intervened on risk factors A_j at time or visit j occurred temporally after the non-intervened on risk factors L_j. We clearly cannot assume this to be the case in the Framingham Offspring Study for all 6 interventions of interest. Further, because the variables are generally measured but once every four years, each risk factor may potentially causally influence the others over the four years. In this setting, if we let Z_j be the set of all measured risk factors at visit j (regardless of whether one wants to consider intervening on any subset), let D_j be the indicator variable for survival at visit j with D_j prior to Z_j, then a sufficient and nearly necessary condition for the g-functional (15) to identify the effect on the cumulative incidence of CHD under an intervention on any chosen subset of the measured time-varying risk factors is that (i) no risk factor causally influences any other risk factor over the four years between visits which we formalize as, for each visit j, no element of Z_j is a cause of any other element of Z_j on the causal DAG generating the data, (ii) the set of all measured risk factors $\overline{Z}_K$ are jointly causally exogenous for survival in the sense that on the causal DAG generating the data, for all combination of visit times s, t, j, there is no unmeasured U_s that is a common cause of both D_j and (an element of) Z_t for which there exist directed paths from U_s to D_j and U_s to Z_t whose other vertices are all unmeasured variables, and (iii) any unmeasured common causes of variables in Z_j are marginally independent of any unmeasured common cause of variables in Z_t for $j \neq s$. That is, conditions (i)–(iii) guarantee that there are no unmeasured confounders for the effect of any possible intervention on survival, even without knowledge of time order among the variables Z_t measured at

the same visit t and even though there is substantial time (four years) between the visits at which data are recorded. Conditions (i)–(iii) are so restrictive that it seems unlikely that we would ever believe that the g-formula will be exactly unbiased for the effect of all the various risk factor interventions in which we might have interest. If conditions (i)–(iii) did hold, then the g-formula gives the overall effect of an intervention regime d on survival by explicitly summing up both its indirect effects on survival mediated through the effect of the intervention on the non-intervened on variables L_j measured by the factors $f(l_j|\bar{l}_{j-1}, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ (which in turn affect the survival as measured by the dependence of the factors $\prod_{j=1}^{4} \Pr[CHD_{j+1} = 0|\bar{l}_j, \bar{a}_j, \overline{CHD}_j = 0, \overline{D}_j = 0]$ on $\bar{l}_j$) and its direct effects on survival measured by the dependence of $\prod_{j=1}^{4} \Pr[CHD_{j+1} = 0|\bar{l}_j, \bar{a}_j, \overline{CHD}_j = 0, \overline{D}_j = 0]$ on $\bar{a}_j$.

Details of estimation

Because the g-formula is a sum over all possible values of risk factor history $\bar{l}_4$ and each l_j is a high-dimensional vector of covariates, a direct calculation based on (15) is computationally infeasible. Rather, we approximate the result of the g-formula under a given intervention by Monte Carlo simulation. To see how to conduct the simulation, first note that the g-formula (15) gives the probability of developing CHD between exams 2 and 5 based on a intervention-specific joint distribution of CHD and risk factors. Under the assumption of no unmeasured confounders, this is the joint distribution had all subjects followed the intervention.

Therefore we generate, for each intervention, a simulated population in which the joint distribution of CHD and risk factors is approximately equal to the joint distribution implied by the g-formula. Then the CHD risk in the simulated population (i.e. the expected fraction of subjects in the simulated population who develop CHD between exams 2 and 5) estimates the desired probability. Note that, if we simulate under no intervention (i.e. with $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0) = 1$ if $a_j = a_j^*$ for all j rather than for just $j = 1$), the expected CHD risk in the simulated population should equal that of the actual study population, because the joint distribution implied by the g-formula for the simulated population is precisely that of the study population.

Let $z_j = (l_j, a_j^*)$. To estimate

$$f\left(z_j|\bar{l}_{j-1},\bar{a}_{j-1},\overline{CHD}_j=0,\overline{C}_j=0,\overline{D}_j=0\right)$$
$$=f\left(a_j^*|\bar{l}_j,\bar{a}_{j-1},\overline{CHD}_j=0,\overline{C}_j=0,\overline{D}_j=0\right)$$
$$f\left(l_j|\bar{l}_{j-1},\bar{a}_{j-1},\overline{CHD}_j=0,\overline{C}_j=0,\overline{D}_j=0\right)$$

where l_j includes all non-intervened on risk factors at exam j, we chose an arbitrary ordering of risk factors at exam j: such as (i) body mass index, (ii) cigarette smoking, (iii) alcohol consumption, (iv) diabetes mellitus, (v) LDL, (vi) HDL and (vii) systolic blood pressure. The density (15) is invariant to the ordering of the variables in z_j. This invariance to ordering is one reason why (15) can only have a causal interpretation for all regimes d under that assumption that no risk factor measured at j causes any other risk measure at j, as in the above caveats. We then estimate (i) the conditional probability of BMI at j given the past variables through $j - 1$, (ii) the conditional probability of smoking given BMI at j and past variables through $j - 1$, (iii) the conditional probability of alcohol at j given BMI, smoking at j and past variables through $j - 1$, and so on. We estimate each of these conditional densities by maximum likelihood from the observed data. Finally, we estimate $f(z_j|\bar{l}_{j-1}, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ as the product of these estimated conditional densities.

In detail, the algorithm used to simulate the simulated population exposed to an intervention was as follows:

Part A: Modelling

1. We fit a pooled (over persons and time) linear logistic regression model to predict the risk of CHD given risk factor history. The outcome was CHD diagnosis between exams j and $j + 1$, and the covariates in the model were risk factors at exams j and $j - 1$. The model was restricted to those with no diagnosis of CHD at or before exam j. The parameters of this model define the estimated conditional probability of CHD risk given the entire past, and thus implicitly assume that risk factors more than 2 time periods previously do not predict CHD risk given risk factors in the past 2 periods.

2. For $j = 3$ and $j = 4$, we fit pooled regression models to predict each risk factor given past risk factor history, among those with no prior diagnosis of CHD. Risk factors at exam j were the "outcome" in models that include other risk factors at exam j (according to the arbitrary ordering explained above) plus all risk factors at exams $j - 1$ and $j - 2$ as covariates. We used linear regression for continuous risk factors, and logistic regression for dichotomous risk factors. Continuous risk factor variables with skewed distributions were log transformed. The parameters of these models define the estimated conditional distributions (Bernoulli or Normal or Log normal) of each risk factor.

3. Follow-up started at exam 2 in our analyses because two prior exams are used to predict CHD risk between exams j and $j + 1$. Thus, the CHD risk we estimate refers to the 12-year period between exam 2 (year 8) and exam 5 (year 20).

Part B: Data generation

4. We simulated a sample of 10 000 individuals by sampling with replacement from the study population.
5. The risk factor values at exams 1 and 2 of these 10 000 individuals were those actually observed (as interventions beginning at exam 2 could not affect these distributions so we were able to use the empirical distribution of the data).
6. The risk factor values $z_j = (l_j, a_j^*)$ at exams 3 and 4 were generated by sampling a value from the conditional distributions estimated in step 2 above.
7. a_j was set equal to a_j^* for each j.
8. The 12-year probability $1-\prod_{j=1}^{4}\Pr\left[CHD_{j+1}=0|\bar{l}_j,\bar{a}_j,\overline{CHD}_j=0,\ \overline{D}_j=0\right]$ of CHD for each simulated individual was estimated, based on his/her simulated risk factor values, using the conditional distribution estimated in step 1 above. The population CHD risk is the average of each person's estimated risk.
9. The above procedure simulates the observed study population under no intervention.
10. To simulate a counterfactual population subject to a given intervention d under the assumption of no unmeasured confounders, step 7 had to be modified by drawing a_j for $j = 2, 3, 4$ from the intervention density $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ rather than setting a_j to a_j^* as under no intervention.

    We used nonparametric bootstrap methods (sampling the observed study population with replacement 100 times) to estimate approximate confidence intervals (CI) of the counterfactual CHD risks and risk ratios. The size of each bootstrap sample was that of the original sample. To obtain standard errors and thus Wald-type confidence intervals, we re-applied all of steps 1–8 to each of the 100 bootstrap samples. All analyses were performed separately by sex.

Marginal structural model

We also analysed the non-dynamic interventions on cigarette smoking using MSMs and compared the results to those obtained using the g-formula. Specifically our goal was (i) to estimate the (potentially time-

varying) hazard of CHD under pre-specified (i.e. non-dynamic) regimes or interventions, and (ii) to use this hazard to compute the 12-year CHD risk.

To estimate the hazard, we fit the discrete-time marginal structural Cox model

$$\text{logit}\lambda_{T_{\bar{a}}}(t) = \text{logit}\lambda_0(t) + \beta a(t)$$

where $\lambda_{T\bar{a}}(t) = \Pr[CHD_{\bar{a}}(t+1) = 1 | CHD_{\bar{a}}(t) = 0]$ is the discrete hazard of CHD between t and $t+1$ under intervention $\bar{a} = [a(2), a(3), a(4)]$ at visits 2, 3 and 4, $\lambda_0(t)$ is the baseline hazard, $a(t)$ is smoking status (smoker or non-smoker) at study visit t under regime $\bar{a}$, and, for dichotomous smoking exposure, β is the log relative risk (i.e. odds ratio) for always exposed vs never exposed. Under the assumption of no unmeasured confounders, the parameters of this model can be estimated using inverse-probability-of-treatment weighting. The model used to predict smoking status was the same as the one used for the parametric g-formula. We fit the model by using a weighted pooled logistic model with a time-varying intercept.

To estimate the 12-year CHD risk when everybody quits smoking at baseline (and nobody initiates smoking thereafter), we could have computed one minus the survival probability

$$[1 - \lambda_{T_{\bar{a}}}(2)][1 - \lambda_{T_{\bar{a}}}(3)][1 - \lambda_{T_{\bar{a}}}(4)]$$

with $\bar{a} = [0, 0, 0]$. However, this simple way of estimating the risk is potentially very inefficient (large variance) because the inverse-probability-of-treatment weights can be unstable. Therefore, to improve efficiency we used inverse-probability-of-treatment weighting to estimate the parameters of the model

$$\text{logit}\lambda_{T_{\bar{a}}}(t) = \text{logit}\lambda_0(t) + \alpha^T V + \beta a(t)$$

where V is a vector of baseline (pre-intervention) covariates. The inclusion of V allows us to use V-stabilized weights which result in much more efficient estimation.

We then simulated a sample of 10 000 individuals by sampling with replacement from the study population and used the parameter estimates of the MSM above to compute the CHD risk between exams 2 and 5 for each simulated individual at each time based on the observed values of V and $\overline{A}(t)$. We then computed the 12-year CHD risk for each individual and averaged over all individuals to obtain an estimate of the risk under no intervention.

To estimate the 12-year CHD risk when 50% of the subjects quit smoking at baseline (and nobody initiates or quits smoking thereafter),

we first assigned a random half of the smokers in our simulated sample to quitting smoking and then proceeded as above to compute each subject's 12-year CHD risk under his/her smoking history (always smoking or never smoking).

7. Results

After exclusions, our analyses included 2230 men (47.8%) and 2440 (52.2%) women with 189 and 68 CHD events, respectively. Thus, the observed 12-year risk of CHD in the study population was 8.48% (95% CI 7.37%–9.73%) for males and 2.79% (95% CI 2.19%–3.54%) for females.

The estimated 12-year risks of CHD (and 95% CI) based on the parametric g-formula under several public health interventions are shown in Table 28.1. The simulated 12-year risk of CHD under no intervention was 8.46% for males (95% CI 5.61%–11.30%) and 2.82% for females (95% CI 2.15%–3.49%). The risk ratios (and 95% CIs) for each intervention compared with no intervention are also shown in Table 28.1. For example, the estimated risk ratio among men is 0.81 (95% CI 0.73–0.90) for all smokers quitting smoking at baseline vs no intervention. Note in row 6, we obtain the overall effect of combined interventions without explicitly imposing any particular functional form (e.g. multiplicative or additive) for the interaction of the individual interventions 1, 3 and 4a on the overall risk of CHD.

Table 28.2 displays the results when only one repeated measure of each risk factor (the most recent one) was included in the model for CHD risk. Results in Table 28.1 come from a model that includes the two most recent measures of each factor. The results in the two tables are compared and contrasted in the next section. Results of MSM analyses for the smoking intervention are given in Table 28.3 which displays the risk and risk ratio estimates from a MSM for two smoking interventions similar to interventions 1 and 2, respectively.

8. Discussion

We have presented two methods—parametric g-formula and MSMs—to estimate the causal effect of hypothetical public health interventions in the Framingham Offspring Study. In the absence of unmeasured confounders and model misspecification, these methods provide causal estimates from observational data that mimic the results of randomized experiments.

The dataset we used has two major limitations: (i) not all relevant confounders are available and (ii) the number of events, especially for women, is small. Because our methods are only valid when the assumption of no unmeasured confounders holds, we would not expect our estimates to be necessarily close to those of a randomized experiment.

Table 28.1 G-formula estimates (reference analysis)

Intervention	*Risk (%)*	*95% CI*	*Risk ratio*	*95% CI*	*BS SE*	*BS av. risk*
Male, *n* = 2230						
0) No intervention (simulated results)	8.46	5.61–11.30	1.00		1.45	8.72
1) 50% quit smoking at baseline	7.65	5.02–10.27	0.90	0.86–0.95	1.34	7.86
2) All quit smoking at baseline	6.82	4.35–9.29	0.81	0.73–0.90	1.26	7.00
3) Alcohol intake to specified distribution	8.41	5.52–11.30	0.99	0.97–1.02	1.47	8.70
4a) BMI lowered by 10% when BMI >22	8.09	4.57–11.61	0.96	0.74–1.24	1.80	8.39
4b) BMI lowered to 22 when BMI >22	7.78	3.84–11.72	0.92	0.65–1.30	2.01	8.25
5) LDL shifts to Chinese distribution	5.80	3.35–8.25	0.69	0.57–0.82	1.25	6.17
6) Combined interventions 1, 3, and 4a	7.26	4.00–10.53	0.86	0.66–1.12	1.67	7.54
Female, *n* = 2440						
0) No intervention (simulated results)	2.82	2.15–3.49	1.00		0.34	2.79
1) 50% quit smoking at baseline	2.63	1.96–3.29	0.93	0.84–1.04	0.34	2.62
2) All quit smoking at baseline	2.44	1.65–3.22	0.86	0.68–1.09	0.40	2.46
3) Alcohol intake to specified distribution	2.83	2.15–3.51	1.00	0.95–1.06	0.35	2.81
4a) BMI lowered by 10% when BMI >22	3.11	2.20–4.02	1.10	0.90–1.34	0.46	3.06
4b) BMI lowered to 22 when BMI >22	3.23	1.94–4.51	1.14	0.82–1.59	0.66	3.23
5) LDL shifts to Chinese distribution	1.44	0.72–2.15	0.51	0.34–0.77	0.37	1.51
6) Combined interventions 1, 3, and 4a	2.91	1.95–3.87	1.03	0.80–1.33	0.49	2.91

Key: BS SE, bootstrap standard error; BS av. risk, bootstrap average risk.

However, our methods gave coherent results, i.e. when applied to similar interventions, both methods yielded similar estimates. Such coherent results are some evidence against model misspecification but constitute no evidence against confounding by unmeasured factors as even when such unmeasured confounders exist both the parametric g-formula and MSM approaches are estimating the same association parameter. We now discuss the relative advantages and disadvantages of the parametric g-formula and MSMs.

8.1 Model misspecification

When using the parametric g-formula, gross model misspecification can be detected by comparing the observed risk and the estimated risk under no intervention. In our example, both risks are similar (8.48% vs 8.46% in men, 2.79% vs 2.82% in women). Note that, whereas dissimilar risks indicate model misspecification, similar risks cannot rule out the existence of model misspecification.

Even in the absence of unmeasured confounding, (i) valid estimation of the effect of the smoking intervention (1) with the parametric g-

Table 28.2 G-formula estimates (risk factor history lagged 1 visit only)

Intervention	*Risk (%)*	*95% CI*	*Risk ratio*	*95% CI*	*BS SE*	*BS av. risk*
Male, *n* = 2230						
0) No intervention (simulated results)	8.46	5.55–11.36	1.00		1.48	8.72
1) 50% quit smoking at baseline	7.63	4.92–10.34	0.90	0.86–0.95	1.38	7.85
2) All quit smoking at baseline	6.79	4.20–9.37	0.80	0.72–0.90	1.32	6.98
3) Alcohol intake to specified distribution	8.42	5.48–11.36	1.00	0.97–1.02	1.50	8.69
4a) BMI lowered by 10% when BMI > 22	7.74	4.49–10.99	0.92	0.78–1.08	1.66	8.06
4b) BMI lowered to 22 when BMI > 22	7.40	3.86–10.94	0.88	0.69–1.11	1.80	7.75
5) LDL shifts to Chinese distribution	5.80	3.24–8.36	0.69	0.57–0.82	1.31	6.11
6) Combined interventions 1, 3, and 4a	6.94	3.87–10.01	0.82	0.69–0.98	1.57	7.22
Female, *n* = 2440						
0) No intervention (simulated results)	2.80	2.15–3.45	1.00		0.33	2.74
1) 50% quit smoking at baseline	2.57	1.94–3.21	0.92	0.83–1.02	0.32	2.54
2) All quit smoking at baseline	2.36	1.64–3.08	0.84	0.68–1.05	0.37	2.34
3) Alcohol intake to specified distribution	2.79	2.14–3.43	1.00	0.96–1.03	0.33	2.72
4a) BMI lowered by 10% when BMI >22	2.76	2.05–3.48	0.99	0.83–1.18	0.36	2.69
4b) BMI lowered to 22 when BMI >22	2.73	1.91–3.55	0.98	0.76–1.26	0.42	2.65
5) LDL shifts to Chinese distribution	1.36	0.82–1.91	0.49	0.35–0.68	0.28	1.36
6) Combined interventions 1, 3, and 4a	2.53	1.82–3.24	0.90	0.73–1.12	0.36	2.48

Key: BS SE, bootstrap standard error; BS av. risk, bootstrap average risk.

Table 28.3 Estimates from marginal structural model

Intervention	*Risk (%)*	*95% CI*	*Risk ratio*	*95% CI*	*BS SE*	*BS av. risk*
Male, *n* = 2230						
0) No intervention (simulated results)	8.61	6.97–10.24	1.00	0.83	8.55	
1) 50% quit smoking at baseline	7.93	6.61–9.25	0.92	0.85–1.00	0.67	7.86
2) All quit smoking at baseline	6.93	5.52–8.34	0.80	0.65–1.00	0.72	6.89
Female, *n* = 2440						
0) No intervention (simulated results)	2.57	1.96–3.19	1.00		0.32	2.41
1) 50% quit smoking at baseline	2.37	1.78–2.96	0.92	0.78–1.09	0.30	2.26
2) All quit smoking at baseline	2.10	1.29–2.91	0.82	0.55–1.22	0.42	2.06

Key: BS SE, bootstrap standard error; BS av. risk, bootstrap average risk.

formula requires that one can correctly model both (a) the conditional probability of developing CHD conditional on risk factor history and (b) the conditional probability of current non-intervened on risk factors given past risk factor history; while (ii) valid estimation with MSMs requires that one (a) correctly model the conditional probability of the intervened on risk factor (smoking) given risk factor history and (b) the structural discrete time proportional hazard model. The parametric g-formula does not require that one correctly models the conditional probability of the intervened-on risk factor (smoking) given risk factor history because for the smoking intervention (1) the density $f_d(a_j|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ does not actually depend on a_j^*; so no model for $f(a_j^*|\bar{l}_j, a_j^*, \bar{a}_{j-1}, \overline{CHD}_j = 0, \overline{C}_j = 0, \overline{D}_j = 0)$ is required in (15) because the a_j^* simply get summed out. Thus the models required for valid estimation of the effect of the smoking intervention (1) with the parametric g-formula have no overlap with the models required for valid estimation with MSMs. As a result, if estimates of the effect of the intervention agree under the two approaches, one can be somewhat confident that major model misspecification is absent. This represents a great advantage of having two independent methods of estimating the same intervention effect; each method can either reinforce or call into question the results obtained under the alternative method.

A subtle form of model misspecification when using the parametric g-formula is due to collinearity among the repeated measures of the risk factors. Collinearity is not a problem for the prediction of CHD risk under no intervention but it may be a problem for predictions of CHD risk under certain interventions. For example, suppose that

1. in the logistic model for CHD risk at time 3, the odds ratio for log BMI at time 2 is 0.2 and the odds ratio for log BMI at time 1 is 5, and because of collinearity among the repeated measures of BMI, neither estimate is significantly different from zero, but a chi-squared test on two degrees of freedom of the hypothesis that both coefficients are zero strongly rejects.

2. in the observed data, a subject has a quite elevated BMI of 29 at exams 1 and 2. Nonetheless, the predicted effect of the subjects BMI at exams 1 and 2 on CHD risk at exam 3 is null in the sense that the odds ratio comparing any level of BMI that is the same at the two exams with any other level of BMI also constant at the two exams on CHD risk at exam 3 is $0.2 \times 5 = 1.0$.

Under 1 and 2, the prediction of CHD risk for the subject under no intervention is quite stable (i.e. the confidence interval for the predicted probability is narrow). Now suppose we intervene to set BMI equal to 22 as soon as it is greater than 22 for times greater or equal than 2 (intervention 4b). The overall effect of BMI reduction is predicted to be

harmful since the odds ratio is $5^{\ln 29 - \ln 22} = 1.56$ compared to a subject with an observed BMI of 29 at both exams. However, in any given bootstrap sample, the odds ratios may be reversed due to sampling variability (because of the high correlation between BMI values over time). Therefore, in some bootstrap samples BMI would appear harmful, whereas in others BMI would appear to have a protective effect. The net result is the large variance we found for the risk (and risk ratio) of intervention 4b and, to a lesser extent, of intervention 4a (Table 28.1). The variance of the risk estimate for intervention 4b was 39% greater than that of the risk estimate under no intervention.

To further assess issues of model misspecification, we conducted two separate sensitivity analyses: (i) we added a quadratic term to the linear term for log BMI in the CHD model, and (ii) we used cubic splines with 3 knots for log BMI in the CHD model. Both of these strategies slightly increased the variance of the estimates for intervention 4a or 4b (data not shown) and the risk ratio estimates did not change significantly.

The high variability arises because the correlation structure of the data is destroyed under the intervention. That is, there is no subject in the sample who has a BMI of 29 at occasion 1 and of 22 at occasion 2. Thus, it is not surprising and wholly appropriate that we are uncertain of what the result would be of an intervention that would create such subjects. Indeed, because there is no subject in the sample who has a BMI of 29 at occasion 1 and of 22 at occasion 2, the uncertainty we see in Table 28.1 for intervention 4b is really an underestimate; it would have been more appropriate to admit that there was no data evidence with which to estimate the effect of such an intervention. Our apparent ability to estimate the intervention effect at all (albeit with great uncertainty) was wholly based on extrapolation under the, possibly incorrect, modelling assumption that the effect of BMI at the last 2 visits on CHD risk is linear on a logistic scale.

Some investigators, when faced with the problem of highly collinear repeated measures of exposures and the associated high variance of predicted interventions, decide to enter only one of the measures in their model, incorrectly arguing that the collinearity obscures the true exposure effect. To empirically examine some of the consequences of this flawed logic, we included only the most recent measure of each risk factor in the CHD model in Table 28.2 rather than the most recent two measures as in Table 28.1. The variance of the risk estimate for intervention 4b was now only 24% greater than that of the risk estimate under no intervention. The difficulty, however, is that this estimate of the intervention effect can be badly biased if, in fact, the second most recent measure of BMI is in truth a risk factor for CHD controlling for the most recent BMI measure.

The lesson to be learned from the above discussion is that interventions that propose big and abrupt changes in the value of a risk factor that otherwise shows little variation over time (under no intervention)

will yield causal estimates with large variances and even these large variances will underestimate the actual uncertainty. This will be true regardless of the methodology used (i.e. MSMs or the parametric g-formula). Intervention involving these risk factors should be formulated in ways that do not imply big and abrupt changes. For example, reducing BMI by 10% (intervention 4a) resulted in a smaller variance than reducing BMI to the value 22 (intervention 4b).

8.2 Efficiency

The parametric g-formula yielded narrower confidence intervals for the risk ratios of interest than the MSM. This was expected based on underlying statistical theory.

8.3 Types of interventions

Marginal structural models are not useful to estimate risks under dynamic interventions (i.e. interventions that depend on the evolving values of time varying risk factors), such as interventions 4a and 4b. In those cases, the g-formula or structural nested models are needed.

A related subtle limitation of MSMs is that they cannot estimate the risk under no intervention in the presence of effect modification by time dependent non-intervened-on risk factors. This is so because IPTW estimation of a MSM effectively creates a pseudo-population in which the probability of receiving exposure does not depend on one's history of time dependent non-intervened-on risk factors. To clarify this point, we consider the following two scenarios:

1. Exposure is assigned to those who actually received it in the observed data.
2. Exposure is assigned to random subjects (in the same proportion as in the original population).

The proportion of CHD in both scenarios will be equal only in the absence of effect modification. A MSM provides estimates of CHD risk for scenario 2 but not for scenario 1. One consequence of this limitation is that, when using MSMs, we lack a way to detect gross model misspecification by comparing observed and estimated risk under no intervention. In our example, the observed and estimated risks under no intervention were close, so we assumed that effect modification was not an important issue and the relative risks in Table 28.3 are good approximations. Note that the no-intervention MSM results represent scenario 2 while the empirical CHD risk in the cohort represents scenario 1.

Because of these same reasons, we needed to slightly modify the meaning of interventions 1 and 2 when using a MSM. Rather than allowing non smokers to follow their own observed history after baseline (as was done in Tables 28.1 and 28.2), we forced them to remain as non-

smokers. This made little substantive difference as very few subjects initiated smoking after baseline.

8.4 Censoring

We defined as censored all subjects that were either lost to follow-up before the last study visit or died from competing causes (i.e. not CHD). Our causal estimates can be interpreted as the CHD risk under each intervention had censoring been abolished (i.e. had nobody been lost to follow-up or died from competing causes), provided there is no unmeasured confounding and censoring is ignorable in the sense that the conditional cause-specific hazard of being censored at k among subjects alive and uncensored up to k does not depend on the time T to CHD given both $\overline{A}(k-1)$ and the time-dependent covariate $\overline{L}(k-1)$ history prior to k. However, one might be interested in the CHD risk had censoring by loss to follow-up, but not by competing causes of death, been abolished, especially because one may feel that the time T to CHD were deaths from competing causes abolished is not a well-defined counterfactual quantity. For example, imagine a cohort of 100 people with 10 CHD cases, 20 deaths from other causes, and 70 subjects alive at the end of follow-up. If censoring from other causes was completely independent of failure and censoring by loss to follow-up, our CHD risk estimate would be 10/80 = 12.5% whereas the risk estimate had other causes of death not been abolished is 10/100 = 10%. Of course, these two risks would only differ substantially if a large proportion of people died from competing causes, which is not the case in our data. We could easily adapt our methods to estimate the effect of interventions when deaths from other causes are included rather than eliminated.

In summary, estimation of non-dynamic interventions should be done both with MSMs and with the parametric g-formula, as agreement between the results is important evidence of lack of major model misspecification. MSMs are only useful to estimate the effect of non-dynamic interventions. The g-formula and structural nested models should be used to estimate the causal effect of dynamic interventions. In future work, we plan to compare g-formula and structural nested models based estimates of the effect of dynamic interventions in a manner similar to the comparisons of g-formula- and MSM-based estimates of non-dynamic regimes reported here.

9. Conclusions

We have tried to show why it is that even with the high quality longitudinal data collected in the Framingham Offspring Study, valid estimation from observational data of the effect of considered interventions requires one to make a large number of unverifiable assumptions that will never hold exactly and may not even hold approximately. In most of the world and for many interventions of interest, relevant high quality

longitudinal data are unavailable, compounding the difficulties. Even if we succeeded in validly estimating the effect of simultaneous interventions on smoking, LDL, BMI and alcohol among the subjects in the Framingham Offspring Study, the question of how to extrapolate such results to other populations would remain unresolved. In this chapter we have stressed the large number of untestable assumptions that must be fulfilled to obtain valid estimates of causal effects and the improbability that all of them hold. We have not done so to discourage attempts to prioritize among potential interventions. Such prioritizing is, and will continue to be, done implicitly or explicitly every day. Our goal was simply to try to help the process by highlighting some of the central issues and potential biases that need to be carefully considered. It is a fact of life that deciding among potential interventions must be done—and must be done under great uncertainty. Methods for making decisions under uncertainty are well established, but the usefulness of such methods ultimately rests on an honest and comprehensive assessment of the uncertainties. It is our hope that this chapter will help in making such assessments.

References

Greenland S, Pearl J, Robins JM (1999) Causal diagrams for epidemiologic research. *Epidemiology*, **10**:37–48.

Hernán MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, **11**:561–570.

Hernán MA, Brumback B, Robins JM (2001) Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association—Applications and Case Studies*, **96**: 440–448.

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association*, **47**:663–685.

Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**:13–22.

Manton KG, Stallard E, Singer B (1992) Projecting the future size and health status of the US elderly population. *International Journal of Forecasting*, 8:433–458.

Neyman J (1923) On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, **5**:465–480.

Pearl J (1995) Causal diagrams for experimental research. *Biometrika*, **82**: 669–710.

Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker

survivor effect. *Mathematical Modelling*, **7**:1393–1512, with 1987 Errata in *Computers and Mathematics with Applications*, **14**:917–921.

Robins JM (1987) Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, **14**:923–945, with 1987 Errata in *Computers and Mathematics with Applications*, **18**:477.

Robins JM (1999) Marginal structural models versus structural nested models as tools for causal inference. *Statistical models in epidemiology: the environment and clinical trials*. IMA Volume 116. Halloran ME, Berry D, eds. Springer-Verlag, New York.

Robins JM, Blevins D, Ritter G, Wulfsohn M (1992) G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, **3**:319–336. See also the following link for the Errata to the article: Robins JM, Blevins D, Ritter G, Wulfsohn M (1993) Errata to G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, **4**:189. Reproduced with permission of the publisher, Lippincott, Williams and Wilkins.

Robins JM, Greenland S, Hu F-C (1999) Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association*, **94**:687–700.

Robins JM, Rotnitzky A, Scharfstein D (1999) Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical models in epidemiology: the environment and clinical trials*. Halloran ME, Berry D, eds. IMA Volume 116. Springer-Verlag, New York.

Rubin D (1978) Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, **6**:34–58.

Spirtes P, Glymour C, Scheines R, eds (1993) *Causation, prediction and search*. Springer-Verlag, New York.

Chapter 29

Conclusions and directions for future research

Alan D. Lopez, Majid Ezzati, Anthony Rodgers, Stephen Vander Hoorn and Christopher J.L. Murray

The analyses of risk factors within a common analytical framework and using comparable methods as outlined in these volumes has ensured greater consistency and comparability in evaluating and using scientific evidence on major risks to health. At the same time, data and knowledge gaps identified in the analyses of these risks illustrate key areas of scientific enquiry necessary to better inform policies and programs that aim to prevent disease by reducing risk factor exposure. The principal findings were discussed in individual risk factor chapters, as well as in those that presented summary results for individual risk factors (chapter 26) and for the joint effects of multiple risks (chapter 27). In this chapter, we use these findings to describe broadly how the analyses might affect pubic health practice as well as research on risk factors.

The analyses of the selected risk factors considered in this work, based on comprehensive reviews of available evidence on exposure and hazards, suggest that a small number of risks, such as childhood and maternal underweight and unsafe sex, accounted for a very large contribution to global loss of healthy life. Further, several risks, such as high blood pressure, tobacco and alcohol, have relative prominence in regions at all stages of development. While reducing all of the above risks to their theoretical minima may not be possible using current interventions, the results illustrate that disease prevention by addressing known distal and proximal risk factors can provide substantial, and under-appreciated, public health gains.

Treatment of established disease will always have a role in public health, especially in the case of diseases such as tuberculosis where treatment contributes to prevention. At the same time, the current devotion of a disproportionately small share of resources to prevention by reducing exposure to major known risk factors, through personal and non-personal interventions, should be reconsidered in a more systematic way in the light of this evidence.

Beyond their total hazard, the distributions of risks in a population have major implications for prevention strategies. Risk typically increases along a continuum of exposures. Dichotomous labels such as "hypertensive" and "normotensive" are therefore not a description of the health consequences of risks, but rather an operational convenience. In fact, the "deviant minority" (e.g. hypertensives) who are considered to be at high risk are only part of a risk continuum, rather than a distinct group, leading to one of the most fundamental axioms in disease prevention across risk factors: "a large number of people exposed to a small risk may generate many more cases than a small number exposed to high risk". Rose (1992) pointed out that wherever this axiom applies (see chapter 26 for possible exceptions), a preventive strategy focusing on high-risk individuals will deal only with the margin of the problem and will not have any impact on the large proportion of disease occurring in the large proportion of people who are at moderate risk.

While a high-risk approach may be more appropriate to the individuals and their physicians at any point in time, treating prevention as managing individual, high-risk crises can only have a limited effect at the population level and over long time periods. This is particularly relevant in the context of efforts to improve global health by addressing multiple diseases and risk factors, many of which exhibit continuous associations with disease outcomes. Focusing on high-risk individuals does not alter the underlying causes of illness, relies on having adequate discriminative ability to predict future disease, and requires continued and expensive screening for new high-risk individuals. In contrast, population-based strategies that seek to shift the whole distribution of risk factors have the potential to control population incidence. Such strategies aim to make healthy behaviours and reduced exposures the social norm and thus lower risk in the entire population.

Our exploration of the joint contributions of multiple risk factors suggests that 20 leading risks contributed to considerable loss of healthy life in different regions of the world. In particular, for some of the leading global diseases (e.g. acute lower respiratory infections, diarrhoea, lung cancer, ischaemic heart disease and stroke), substantial proportions were attributable to these selected risk factors. Removing the leading 20 risk factors from among those studied here would not only have resulted in a 9.3-year (17%) gain in global healthy life expectancy in 2000, but also accounted for some of the interregional healthy life expectancy (HALE) differences. The analysis showed that even populations with currently high healthy life expectancy (e.g. developed regions of the Western Pacific and Europe) could further benefit considerably from risk reduction. These results provide a guide for achieving potential gains in (healthy) life expectancy that have been estimated statistically from past trends (Oeppen and Vaupel 2002; Riley 2001) through disease prevention by reducing known risks. The results for multiple risk factors further emphasize that for more effective and affordable implementation of a

prevention paradigm, policies, programmes and scientific research should acknowledge and take advantage of the interactive and correlated role of major risks to health, across and within causality layers. This could be an important step in addressing health inequalities, many of which may arise from concentration of major risks among specific socioeconomic groups.

Health research has at times focused on topics which, while scientifically intriguing, have not always been motivated by broader population health consequences (Anonymous 2001; Gross et al. 1999; Horton 2003; Willet 2002). The collation of evidence on exposure and hazard for different risks in this book, and the existing data gaps have illustrated data and monitoring needs for better quantification of important risks.

Research needs include more detailed and better quality data on exposure to most risks, using exposure variables that capture the full distribution of hazards in the population. Important examples include detailed data on alcohol consumption volume and patterns, dietary and biological markers for micronutrients, and better indicators for physical activity, indoor air pollution and occupational risks, all of which were quantified using indirect measures with limited resolution.

Assumptions and extrapolations were also needed in quantifying risk factor–disease relationships because of gaps in knowledge about the impact of some important global risk factors, particularly in developing countries. Examples include limited quantitative assessment of the hazards of specific sexual behaviours for HIV/AIDS or other sexually transmitted diseases, alcohol drinking patterns (Puddey et al. 1999) or exposure to indoor smoke from solid fuels (Ezzati and Kammen 2002). Equally important are detailed exposure data for risks that have been traditionally studied in developed countries, but have global importance and require more detailed data and hazard quantification in developing regions (e.g. alcohol and obesity).

The limited evidence on the effects of multiple risk factors and risk factor interactions also points to important gaps in research on multirisk and stratified hazards as also discussed in chapters 27 and 28. Including multiple causes in epidemiological research and risk assessment would allow estimating the benefits of reduction in combinations of distal and proximal exposures using multiple interventions (e.g. using education and economic tools to: i) promote physical activity or healthier diet coupled with screening and lowering cholesterol; and ii) address overall childhood nutrition and environment instead of a focus on individual components). In such research, risk factor groups should be selected based on both biological relationships and socioeconomic factors that affect multiple diseases, as discussed in chapter 27. Examples include those risk factors that are affected by the same policies and distal socioeconomic factors (e.g. malnutrition, unsafe water, sanitation and hygiene, indoor smoke from solid fuels and rural development policies) or affect the same group of diseases (e.g. all of the above for child-

hood infectious diseases; smoking, diet and physical activity for cardiovascular diseases). Once risk factors have been selected, the emphasis on reducing confounding should be matched by equally important enquiry into independent and mediated hazard sizes that are stratified based on the levels of other risks.

This is a substantial research agenda, and one for which some progress has been made over the past decades. Yet public health policy needs demand that much greater priority is given to research that more reliably and relevantly identifies the potential for prevention in all countries, including information on exposures and risks among subpopulations, especially among the least well off. Such research will undoubtedly provide a more compelling basis for the massive increase in preventive efforts worldwide that is required if the potential for health gains identified in this book is to be realized.

References

Anonymous (2001) The human genome, in proportion. *The Lancet*, **357**:489.

Ezzati M, Kammen DM (2002) The health impacts of exposure to indoor air pollution from solid fuels in developing countries: knowledge, gaps, and data needs. *Environmental Health Perspectives*, **110**:1057–1068.

Gross CP, Anserson GF, Powe NR (1999) The relation between funding by the national institutes of health and the burden of disease. *New England Journal of Medicine*, **340**:1881–1887.

Horton R (2003) Medical journals: evidence of bias against the diseases of poverty. *The Lancet*, **361**:712–713.

Oeppen J, Vaupel JW (2002) Broken limits to life expectancy. *Science*, **296**: 1029–1030.

Puddey IB, Rakic V, Dimmitt SB, Beilin LJ (1999) Influence of pattern of drinking on cardiovascular disease and cardiovascular risk factors—a review. *Addiction*, **94**:649–663.

Riley JC (2001) *Rising life expectancy: a global history*. Cambridge University Press, Cambridge.

Rose G (1992) *The strategy of preventive medicine*. Oxford University Press, Oxford.

Willet WC (2002) Balancing life-style and genomics research for disease prevention. *Science*, **296**:695–698.

Index

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES