CHAPTER TWO

Defining and Assessing Risks to Health

This chapter offers a detailed explanation of the report's approach to health risks. It argues that while much scientific effort and most health resources today are directed towards treating disease, rather than preventing it, focusing on risks to health is the key to prevention. Such risks do not occur in isolation, so both proximal and distal causes of adverse health outcomes need to be considered. Population-based strategies aim to make healthy behaviour a social norm, thus lowering risk in the entire population. Small shifts in some risks in the population can translate into major public health benefits. Therefore this chapter strongly advocates the assessment of population-wide risks as well as high-risk individuals in strategies for risk reduction. The key challenge is to find the right balance between the two approaches. Risk assessment has emerged in recent years from its roots in the study of environmental problems, and the steps generally involved in environmental risk assessment can be adapted to apply more specifically to the analysis of health risks. This chapter explains the benefits of comparing different risks to health and defines and explains risk assessment.

2

Defining and Assessing Risks to Health

What are risks to health?

*R*isk can mean different things to different people, as summarized in Box 2.1. The two most common meanings will be used in this report – risk as a probability of an adverse outcome, or a factor that raises this probability.

Why focus on risks to health?

Focusing on risks to health is key to preventing disease and injury. The most emotive and tangible images in health are of people suffering from disease, but preventing disease and injury occurring in the first place requires systematic assessment and reduction of their causes. Much scientific effort and most health resources are directed towards treating disease – the "rule of rescue" still dominates *(3)*. Data on disease or injury outcomes, such as death or hospitalization, tend to focus on the need for palliative or curative services. In contrast, assessments of burden resulting from risk factors will estimate the potential of prevention. One notable exception concerns communicable diseases, since treating infected individuals can prevent further spread of infection, and hence treatment can be a method of prevention in itself.

Even when the focus is on causes as well as disease outcomes, much scientific activity has been directed to assessing whether a risk exists at all. Does electromagnetic frequency radiation cause leukaemia? Do certain infections predispose to heart attacks? These assessments are usually accompanied by estimates of how much higher the risk is in individuals who are exposed compared with those who are not. It has been much less common to assess impact at a population level by asking "of all the disease burden in this population, how much could be caused by this risk?"

Many factors are relevant in prioritizing strategies to reduce risks to health: the extent of the threat posed by different risk factors, the availability of cost-effective interventions, and societal values and preferences are particularly important. These factors are also key for research priorities – if major threats exist without cost-effective solutions, then these must be placed high on the agenda for research. Governments are also likely to place particular value on ensuring their main efforts focus on the largest threats to health in their countries. Reliable, comparable and locally relevant information on the size of different risks to health is therefore crucial to prioritization, especially for governments setting broad directions for health policy and research. However, such information has typically been very limited, cre-

ating a gap in which interest groups may seek either to downplay or to overestimate some risks. In addition, there is an inherent imbalance in media information about risks: common, major threats to health are usually not reported because they are already known, whereas rare or unusual threats to health are highly newsworthy.

Stewardship is one of the key functions of government, necessitating a broad overview, a long-term horizon and an evidence-based approach, and requiring information from reliable, comparable assessments of the magnitude of different major risks to health. This report helps to redress the dearth of such information. The report recognizes that risk analysis is a political enterprise as well as a scientific one, and that public perception of risk also plays a role in risk analysis, bringing issues of values, process, power and trust into the picture. The roles and contributions of risk assessment, communication, risk management, cost-effectiveness and policy development form the focus of the report.

DEVELOPMENT OF RISK ASSESSMENT

People have been interested in risks to health throughout history. During the past several decades, this interest has intensified and has also begun to include many new perspectives. The field of risk analysis has grown rapidly, focusing on the identification, quantification and characterization of threats to human health and the environment – a set of activities broadly called risk assessment.

While clearly there has been very long interest in comparing risks posed by different threats to health, formal frameworks have been developed only relatively recently. Risk assessment has its roots in the environmental sector, where it was developed as a systematic way of comparing environmental problems that pose different types and degrees of health risk. Such environmental risk assessment exercises generally comprise four elements.

- *Hazard identification* identifies the types of health effect that can be caused, based on toxicological data from laboratory or epidemiological studies: for example, chemical X causes liver damage.
- *Exposure assessment* combines data on the distribution and concentrations of pollution in the environment with information on behaviour and physiology to estimate the amount of pollutant to which humans are exposed. Biomarkers have been used to gauge levels of some exposures, such as lead and dioxin.
- *Dose–response assessment* relates the probability of a health effect to the dose of pollutant or amount of exposure.
- *Risk characterization* combines the exposure and dose–response assessments to calculate the estimated health risks, such as the number of people predicted to experience a particular disease, for a particular population. This typically includes estimation and communication of uncertainties.

Environmental risk assessments of likely health effects, together with consideration of costs, technical feasibility and other factors, can be used to set priorities for environmental management. Environmental risk assessment has analogies to the strategies developed in epidemiology for assessing population attributable risks, that is, the proportion of disease in a population that results from a particular hazard. A more general approach based on these frameworks can be extended to many other areas. A key part of this report outlines such methods and provides an illustrative analysis of burden caused by a variety of different risks to health.

Risk assessment can be defined here as a systematic approach to estimating and comparing the burden of disease and injury resulting from different risks. The work pre-

Box 2.1 What does risk mean?

- Risk can mean a probability, for example, the answer to the question: "What is the risk of getting HIV/AIDS from an infected needle?"
- Risk can mean a factor that raises the probability of an adverse outcome. For example, major risks to child health include malnutrition, unsafe water and indoor air pollution.
- Risk can mean a consequence. For example, what is the risk from driving while drunk? (answer: being in a car crash).
- Risk can mean a potential adversity or threat. For example, is there risk in riding a motorcycle?

In this report, the first two meanings are used. Risk is defined as a probability of an adverse health outcome, or a factor that raises this probability. Other important risk-related definitions are outlined below.

- ***Prevalence of risk*** – the proportion of the population who are exposed to a particular risk. For example, the prevalence of smoking might be 25% in a particular population.

- ***Relative risk*** – the likelihood of an adverse health outcome in people exposed to a particular risk, compared with people who are not exposed. For example, if people who smoke for a certain time are, on average, 15 times more likely to develop lung cancer than those who do not smoke, their relative risk is 15.
- ***Hazard*** – an inherent property, for example of a chemical, that provides the potential for harm.
- ***Population attributable risk*** – the proportion of disease in a population that results from a particular risk to health.
- ***Attributable burden*** – the proportion of current disease or injury burden that results from past exposure.
- ***Avoidable burden*** – the proportion of future disease or injury burden that is avoidable if current and future exposure levels are reduced to those specified by some alternative, or counterfactual, distribution.

Sources: *(1, 2).*

sented in this report builds on several similar estimates conducted in recent years. The first global estimates of disease and injury burden attributable to a set of different risk factors were reported in the initial round of the global burden of disease study *(4, 5)*. These estimates add to the many others made for selected risk factors in specific populations, for example, tobacco *(6)*, alcohol and other drugs *(7)*, environmental factors *(8)*, blood pressure *(9)*, and selected risk factors for certain regions *(10–12)*.

In the first round of the global burden of disease study, risk factors were assessed that were either exposures in the environment (for example, unsafe water), human behaviour (for example, tobacco smoking) or physiological states (for example, hypertension). However, in such early risk assessments, there was a lack of comparability between different risk factor assessments arising, in part, from a lack of standard comparison groups and different degrees of reliability in assessing risk factors. Also, the relevance of varying time lags between exposure and outcome – for example, short for alcohol and injuries and long for smoking and cancer – was not captured. A key aim of this analysis is therefore to increase comparability between the estimates of the impact of different risk factors and characterize the timing of these impacts.

Risk assessment estimates burden of disease resulting from different risk factors, each of which may be altered by many different strategies; it can provide an overall picture of the relative roles of different risks to human health. Specific strategies for identifying the appropriate sets of interventions, and the crucial roles of cost-effectiveness analyses in choosing from among them, are outlined in Chapter 5.

Key goals of global risk assessment

An effective risk assessment must have a well-defined scope, which in turn depends on the purpose of the analysis. For example, an evaluation of emissions from a particular industrial facility is likely to concentrate on their health effects on local populations. In contrast, a project to set national environmental priorities may be much broader in scope, covering such factors as emissions of greenhouse gases and ozone-depleting substances. Some trade-offs will inevitably be required. Governments and ministries of health oversee

overall population health and so, at the broadest level, need information from risk assessments that are comprehensive as well as being reliable, relevant and timely. Because the range of risks to health is almost limitless, it is essential for governments to have a quantitative approach to gauging their importance. Risks need to be defined and studied comprehensively irrespective of factors such as their place in a causal chain or the methods used (from the disciplines of the physical, natural, health, and social sciences) for their analysis. The following sections outline some of the different dimensions that should be considered.

STANDARDIZED COMPARISONS AND COMMON OUTCOME MEASURES

Ideally, the impact of each risk factor should be assessed in terms of a "common currency" that incorporates loss of quality of life as well as loss of life years. The principal metric used in this report is the DALY (disability-adjusted life year) – one DALY being equal to the loss of one healthy life year *(13)*.

A key initial question when assessing the impact of a risk to health is to ask "compared to what?" This report employs an explicit counterfactual approach, in which current distributions of risk factors are compared with some alternative, or counterfactual, distribution of exposure. Many different counterfactuals are potentially of interest. To enhance comparability across risk factors, the basis for the results in Chapter 4 is the theoretical minimum risk distribution, that is exposure levels that would yield the lowest population risk (for example, no tobacco use by any members of a population). For the analysis of the costs and effects of interventions to reduce risk in Chapter 5, a related counterfactual is used – based on the burden that would exist in the absence of relevant interventions. Risk factor distributions that are plausible, feasible and cost-effective will lie somewhere between the current risk factor levels and the related theoretical minimum. The envisaged shift from current to counterfactual scenarios has been termed the *distributional transition* (see Figure 2.1).

In many instances, the counterfactual of most relevance will involve small to moderate distributional transitions (for example, 10%, 20% or 30%), as these are most likely to be feasible and cost-effective. These estimates are also less susceptible to the influence of arbitrary choices of theoretical minima, and are likely to be the most reliable, as the dose–response is often least certain at low exposure levels.

Figure 2.1 Example of distributional transitions for blood pressure and for tobacco smoking
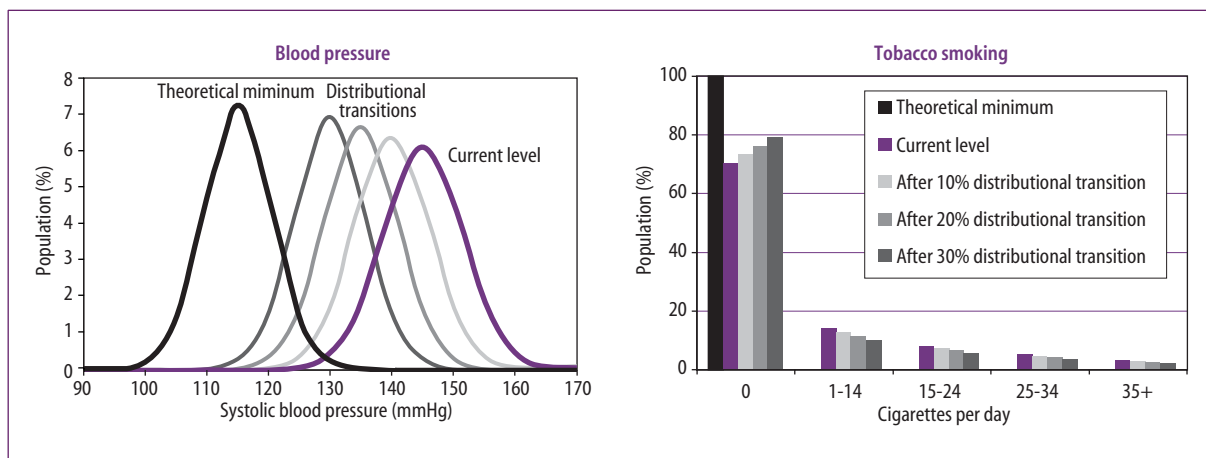
ASSESSING PROTECTIVE AS WELL AS HAZARDOUS FACTORS

Factors that affect risk of disease or injury are, of course, not all harmful. Risk factor does have a negative connotation, but ideally a risk assessment should include a range of protective as well as hazardous risk factors. For example, this report considers the protective benefits of fruit and vegetable intake and physical activity by assessing people with low levels of these factors. The important role of protective factors in adolescent health is outlined in Box 2.2.

INCLUDING PROXIMAL AND DISTAL CAUSES

Risks to health do not occur in isolation. The chain of events leading to an adverse health outcome includes both proximal and distal causes – proximal factors act directly or almost directly to cause disease, and distal causes are further back in the causal chain and act via a number of intermediary causes (see Figure 2.2). The factors that lead to someone developing disease on a particular day are likely to have their roots in a complex chain of environmental events that may have begun years previously, which in turn were shaped by broader socioeconomic determinants. For example, society and culture are linked to certain drinking patterns, which in turn influence outcomes such as coronary heart disease via physiological processes such as platelet aggregation. Clearly, there are risks over which an individual has at least some control (for example, inactivity) and risks that mostly or entirely rest at a population or group level (for example, ambient air pollution). It is essential that the whole of the causal chain is considered in the assessment of risks to health. Indeed, many risks cannot be disentangled in order to be considered in isolation, as they act at

---

Box 2.2 Protective factors

A growing body of cross-cultural evidence indicates that various psychological, social and behavioural factors are protective of health in adolescence and later life. Such protection facilitates resistance to disease, minimizes and delays the emergence of disabilities, and promotes more rapid recovery from illness.

Among the psychosocial factors that have been linked to protection in adults are: an optimistic outlook on life with a sense of purpose and direction, effective strategies for coping with challenge, perceived control over life outcomes, and expressions of positive emotion. Epidemiological studies have shown reduced morbidity and delayed mortality among people who are socially integrated. The quality of social relationships in the home (parent–child relations and spousal ties) and the workplace (employer–employee relations and coworker connections) are now recognized as key influences on physical and mental health. A growing literature underscores the protective health benefits associated with persistently positive and emotionally rewarding social relationships. Positive

health behaviours (e.g., proper diet and adequate exercise, and avoiding cigarettes, drugs, excessive alcohol and risky sexual practices) are also influenced by psychosocial factors.

The presence of psychosocial factors in understanding positive human health points to new directions for research and practice. The biological mechanisms through which psychosocial and behavioural factors influence health are a flourishing area of scientific inquiry: investigations in affective neuroscience are relating emotional experience to neural structures, function, dynamics and their health consequences. There is a need for greater emphasis in policy and practice on interventions built around the growing knowledge that psychosocial factors protect health.

Adolescence is a critical life stage when lifestyle choices are established, including health-related behaviours with impacts throughout life. Recent research has begun to focus on the role of protective factors in youth behaviour, complementing previous approaches concerned only with problems and risk taking.

Evidence from 25 developing countries, 25 European countries, Canada, Israel and the United States shows that adolescents who report having a positive connection to a trusted adult (parent or teacher) are committed to school, have a sense of spirituality and exhibit a significantly lower prevalence of risky behaviours. This is in addition to being more socially competent and showing higher self-esteem than adolescents without such a connection. Studies in the US have shown that these protective factors also predict positive outcomes (remaining connected to school, engaging in more exercise and having healthy diets) while diminishing negative behaviour (problem drinking, use of marijuana and other illicit drugs, and delinquent behaviour).
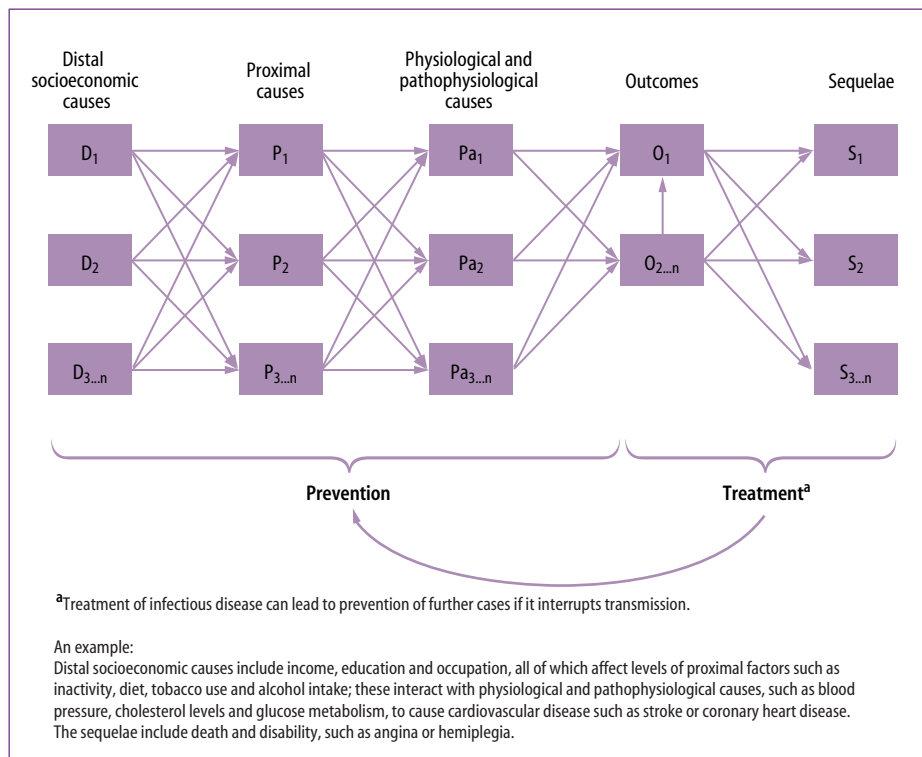
Protective factors promote positive behaviours and inhibit risk behaviours, hence mitigating the impacts of exposure to risk. Current efforts to reduce risks in the lives of adolescents should be broadened to include the strengthening of protective factors.

Sources: *(14–19).*

different levels, which vary over time. An appropriate range of policies can be generated only if a range of risks is assessed.

There are many trade-offs between assessments of proximal and distal causes. As one moves further from the direct, proximal causes of disease there can be a decrease in causal certainty and consistency, often accompanied by increasing complexity. Conversely, distal causes are likely to have amplifying effects – they can affect many different sets of proximal causes and so have the potential to make very large differences *(20)*. In addition, many distal risks to health, such as climate change or socioeconomic disparity, cannot appropriately be defined at the individual level. A population's health may also reflect more than a simple aggregation of the risk factor profile and health status of its individual members, being a collective characteristic and a public good that in turn affects the health status of its members *(21)*.

Research into the different levels of risks should be seen as complementary. There is considerable importance in knowing the population-level determinants of major proximal risks to health such as smoking. Similarly, there is value in knowing the mechanisms through which distal determinants operate. Understanding both proximal and distal risks requires contributions from different scientific traditions and different areas of health impact: environmental, communicable, noncommunicable, injury, and so on, and as a result different intellectual tools and methods, including those of health, physical and social sciences. This in turn requires consideration of the context of particular risks: some are likely always to have negative health effects (for example, tobacco use) while others may have a role that changes from setting to setting (for example, breastfeeding protects against diarrhoeal disease, to an extent that depends on the prevalent patterns of diarrhoea). Also, the same risk can be measured and quantified at various levels depending on measurement technology

Figure 2.2 Causal chains of exposure leading to disease

[a] Treatment of infectious disease can lead to prevention of further cases if it interrupts transmission.

An example:
Distal socioeconomic causes include income, education and occupation, all of which affect levels of proximal factors such as inactivity, diet, tobacco use and alcohol intake; these interact with physiological and pathophysiological causes, such as blood pressure, cholesterol levels and glucose metabolism, to cause cardiovascular disease such as stroke or coronary heart disease. The sequelae include death and disability, such as angina or hemiplegia.

and policy needs. For example, measuring iodine levels in food and in the environment requires different tools and the results have different implications.

When distal exposures operate through different levels of risk factors, their full impact may not be captured in traditional regression analysis methods in which both proximal and distal variables are included. More complex multilevel models and characterization of causal webs of interactions among risk factors may lead to more appropriate estimates, as well as facilitating estimation of the effect of simultaneous changes in two or more risk factor distributions. Some examples are shown later in the report.

Risk factors can also be separated from outcomes in time, sometimes by many decades. Box 2.3 shows how disadvantage can be accumulated across the life course.

ASSESSING POPULATION-WIDE RISKS AS WELL AS HIGH-RISK INDIVIDUALS

Many risks to health are widely distributed in the population, with individuals differing in the extent of their risk rather than whether they are at risk or not. Binary categorization into "exposed" and "unexposed" can substantially underestimate the importance of continuous risk factor–disease relationships. Consequently, much of this report estimates the effects of shifting distributions of exposures by applying a counterfactual approach, that is, by comparing the burden caused by the observed risk factor distribution with that expected from some alternative, or counterfactual, distribution. This approach allows assessment of population-wide interventions (see Box 2.4 and Figure 2.3).

INCLUDING RISKS THAT ACT TOGETHER TO CAUSE DISEASE

Many risks to health act jointly to cause disease or injury, and this has important implications for prevention opportunities, as outlined in Box 2.5. This report presents estimates of the individual effects of different selected risks to health, followed by analyses of the joint effect of selected clusters of risks.

Box 2.3 Risks to health across the life course

In recent years, a life-course approach to the study of health and illness – which suggests that exposure to disadvantageous experiences and environments accumulates throughout life and increases the risk of illness and premature death – has helped to explain the existence of wide socioeconomic differentials in adult morbidity and mortality rates.

Chronic illness in childhood, more common among children of manual workers, can have long-term consequences both for health and socioeconomic circumstances in later life. Slow growth in childhood (short stature for age and sex) is an indicator of early disadvantage. Early material and psychosocial disadvantage may also have an adverse impact on psychological and cognitive development, which in turn may affect health and labour-market success later in life. The impact of living and working environments – and lifestyle factors such as smoking – on health in-equalities has long been recognized. Cumulative differential lifetime exposure to health-damaging or health-promoting environments appears to be the main explanation for observed variations in health and life expectancy by socioeconomic status.

Disadvantage may begin even before birth: low birth weight is associated with increased rates of coronary heart disease, stroke, hypertension and non-insulin-dependent diabetes. These associations extend across the normal range of birth weight and depend on lower birth weights in re-lation to the duration of gestation rather than the effects of premature birth. The associations may be a consequence of "programming", whereby a stimulus or insult at a critical, sensitive period of early life has permanent effects on structure, physiology and metabolism. Programming of the fetus may result from adaptations invoked when the maternal–placental nutrient supply fails to match the fetal nutrient demand. Although the influences that impair fetal development and programme adult cardiovascular disease remain to be defined, there are strong pointers to the importance of maternal body composition and dietary balance during pregnancy.

Sources: *(22–24).*

USING BEST AVAILABLE EVIDENCE TO ASSESS CERTAIN AND PROBABLE RISKS TO HEALTH

It is important in any risk assessment to review quantitatively the best available evidence for both "definite" and "probable" risks. Estimation of the potential impact of a health hazard can never wait until perfect data are available, since that is unlikely to occur. Timeliness is essential. This area can be a source of tension between scientists and policy-makers. However, arguments are often clouded by the use of dichotomies – assertions of uncertainty or certainty when, in fact, there are different degrees of uncertainty and disagreement about tolerable thresholds. Similarly, it may be asserted that there are no data when some indirect data are available, or at least the range of levels in other parts of the world is known. For example, in estimating fruit and vegetable intake for countries with no known surveys on this topic, upper and lower ranges can be estimated from surveys undertaken elsewhere, and food sales and agricultural data can be used to produce indirect estimates that occupy a narrower range. Internal consistency can help put ranges on uncertainty: for example, mortality rates, population numbers and birth rates should be internally consistent, and reliable estimates for some of these components will put bounds on the uncertainty of the others. However, as outlined earlier, the sum of causes is unbounded and so internal consistency checks cannot be performed in assessments of different risks to health. Strategies to minimize this problem include full documentation of data sources, methods and assumptions, extensive peer review, explicit assessments of causality, and quantitative estimates of other uncertainty.

Box 2.4 Population-wide strategies for prevention

"It makes little sense to expect individuals to behave differently from their peers; it is more appropriate to seek a general change in behavioural norms and in the circumstances which facilitate their adoption." – Geoffrey Rose, 1992.

The distribution and determinants of risks in a population have major implications for strategies of prevention. Geoffrey Rose observed, like others before and since, that for the vast majority of diseases "nature presents us with a process or continuum, not a dichotomy". Risk typically increases across the spectrum of a risk factor. Use of dichotomous labels such as "hypertensive" and "normotensive" are therefore not a description of the natural order, but rather an operational convenience. Following this line of thought, it becomes obvious that 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. This leads to one of the most fundamental axioms in preventive medicine: "a large number of people exposed to a small risk may generate many more cases than a small number exposed to high risk". Rose pointed out that wherever this axiom applies, 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. For example,
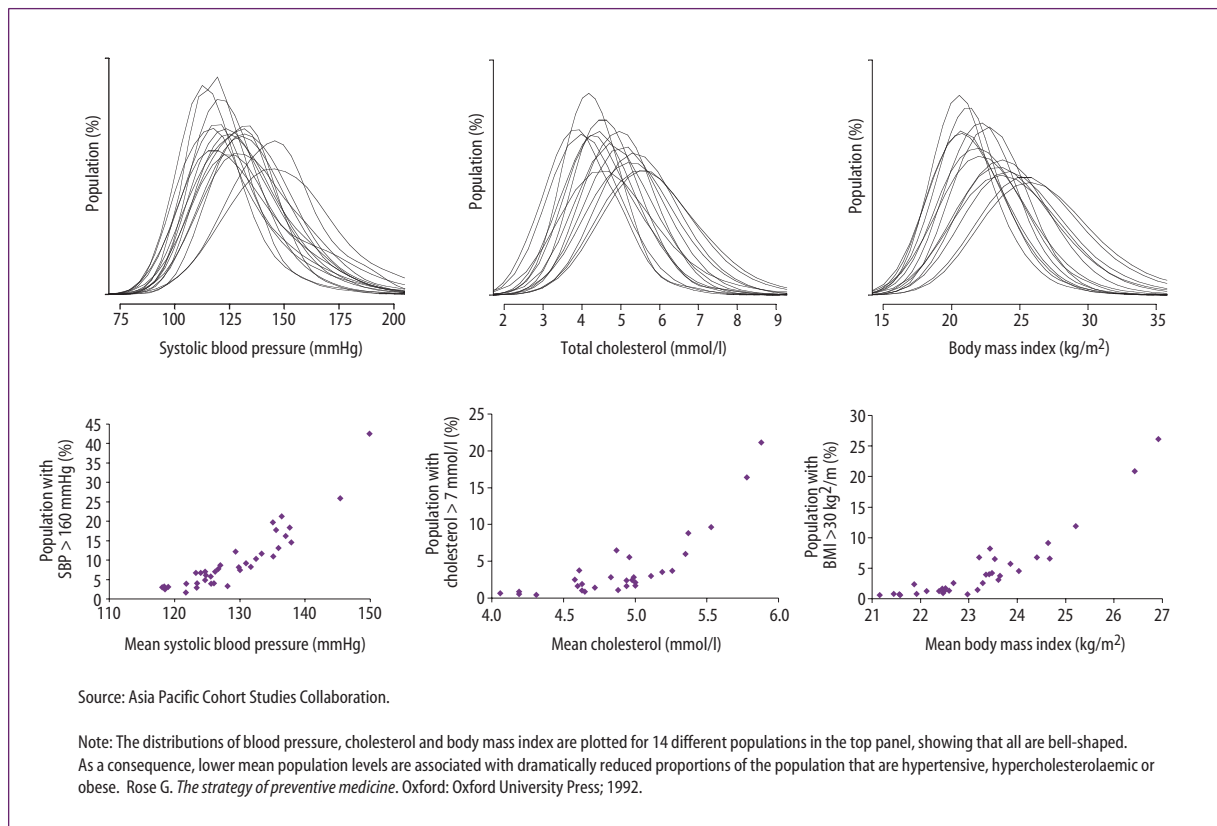
people with slightly raised blood pressure suffer more cardiovascular events than the hypertensive minority. While a high-risk approach may appear more appropriate to the individuals and their physicians, it can only have a limited effect at a population level. It does not alter the underlying causes of illness, relies on having adequate power 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 into social norms and thus lower the risk in the entire population. The potential gains are extensive, but the challenges are great as well – a preventive measure that brings large benefits to the community appears to offer little to each participating individual. This may adversely affect motivation of the population at large (known as the "prevention paradox").

Although most often applied to cardiovascular disease prevention, a population-wide approach is often relevant in other areas. For example,

a high-risk strategy for melanoma prevention might seek to identify and target individuals with three or more risk factors (such as a number of moles, blonde or auburn hair, previous sunburn, and a family history of skin cancer). However, only 24% of cases of melanoma occur in this 9% of the population, so a targeted approach would succeed in identifying those at high risk but would do little for population levels of melanoma – 75% of cases occur in the 58% of the population with at least one risk factor. A population-wide strategy would seek to make sun protection a social norm, so that the whole population is less exposed to risk.

These approaches are complementary: a population approach can work to improve and extend the coverage of a high-risk approach. A key challenge is finding the right balance between population-wide and high-risk approaches. Rose concluded that this will require a wider world view of ill-health, its causes and solutions, and will lead to acknowledgement that the primary determinants of disease are mainly economic and social, and therefore remedies must also be economic and social.

Sources: *(20, 25, 26)*.

Figure 2.3 The importance of population distributions of exposure

Source: Asia Pacific Cohort Studies Collaboration.

Note: The distributions of blood pressure, cholesterol and body mass index are plotted for 14 different populations in the top panel, showing that all are bell-shaped. As a consequence, lower mean population levels are associated with dramatically reduced proportions of the population that are hypertensive, hypercholesterolaemic or obese. Rose G. *The strategy of preventive medicine*. Oxford: Oxford University Press; 1992.

Extrapolations and indirect methods are often justified where there are implications in delaying estimates of health impacts and subsequent policy choices. If decisions await improved estimates, then not producing best current estimates (with appropriate indications of uncertainty) may mean inappropriate inaction. Alternatively, decisions may be made with other even more uncertain information, where the uncertainty will often be implicit. Nonetheless, there can be costs in making incorrect estimates and, ultimately, it is largely a matter of judgement to decide when data are adequate.

Whenever possible, the level of uncertainty should be reported explicitly in risk assessments. There is still considerable debate about how this is best done in a policy-relevant way, given the inevitable play of chance and uncertainties in both the likelihood of causality and the validity of the estimation methods. Major uncertainty should result in calls for more data. In particular, data are often absent or scanty in the developing countries, where many risks are highest and more information could produce the greatest gains in knowledge. The management of highly uncertain risks and the use of the precautionary principle are discussed in Chapter 6.

Assessing avoidable as well as attributable burden

Risk assessments to date have typically used only attributable risk estimates, basically addressing the question "what proportion of current burden is caused by the accumulated effects of all prior exposure?" However, often a more policy-relevant question is "what are the likely future effects of partial removal of current exposure?" Two key developments are therefore needed: an explicit focus on future effects and on less-than-complete risk factor

changes. This report presents estimates of attributable burden (current burden due to past exposure) and of avoidable burden (the proportion of future burden avoidable if current and future exposure levels are reduced to those specified by some alternative, or counterfactual, distribution). When the time between exposure and disease or death is short, the distinction between attributable and avoidable burden is not critical. However, for risk factors such as tobacco and some occupational exposures, a long time lag between exposure and health outcome may result in a major difference between attributable and avoidable burden. The distinction between attributable and avoidable burden is shown graphically in Figure 2.4.

OVERVIEW OF RISK ASSESSMENT METHODS

The overall aim of the analyses reported here was to obtain reliable and comparable estimates of attributable and avoidable burden of disease and injury, for selected risk factors. More specifically, the objectives were to estimate, by age, sex and region, for selected risk factors:
- attributable burden of disease and injury for 2000, compared to the theoretical minimum;
- avoidable burden of disease and injury in 2010, 2020 and 2030, for a standardized range of reductions in risk factors.

Box 2.5 Multiple causes of disease

The impact of a single risk factor on disease is often summarized as the proportion of disease caused by, or attributable to, that risk factor. The fact that diseases and injuries are caused by the joint action of two or more risk factors means that the sum of their separate contributions can easily be more than 100%. Consider a hypothetical situation of deaths from car crashes on a hazardous stretch of road. Studies may have shown that they could be reduced by 20% by using headlights in daytime, 40% by stricter speed limits, 50% by installing more traffic lights, and 90% by creating speed bumps.

As a further example consider a smoker, also a heavy drinker, who develops throat cancer. The cancer would not have developed on that particular day if the person had not smoked or drunk heavily: it was very likely caused by both tobacco and alcohol. There are three possible scenarios for throat cancer, each with a different set of causes that must be present for the disease to occur. In the first scenario, smoking and alcohol work together with other environmental and genetic causes to result in the disease ("environ-

mental" can be taken as all non-genetic causes). The second scenario is the same, except that throat cancer develops in a non-drinker. In the third, we do not know what caused the cancer, other than genetic and some unknown environmental causes. This simplified model illustrates the following important issues.
- Causes can add to more than 100%. If the scenarios were equally common, 66.6% of throat cancer would be attributable to smoking, 33.3% to alcohol, 100% to genetic causes, and 100% to unknown environmental causes, making a total of 300%. Causes can, and ideally should, total more than 100%; this is an inevitable result of different causes working together to produce disease, and reflects the extent of our knowledge of disease causation.
- Multicausality offers opportunities to tailor prevention. If these scenarios were numerically correct, throat cancer could be reduced by up to two-thirds with smoking cessation, by up to one-third with reduced alcohol intake, or by up to two-thirds with less marked decreases in both smoking and alcohol consumption. Fur-

ther reductions could also take place if research led to additional preventive strategies based on genetic or other environmental causes. The key message of multicausality is that different sets of interventions can produce the same goal, with the choice of intervention being determined by such considerations as cost, availability and preferences. Even the most apparently single-cause conditions are on closer inspection multicausal; the tubercle bacillus may seem to be the single cause of tuberculosis but, as improved housing has been shown to reduce the disease, living conditions must also be considered a cause.
- Prevention need not wait until further causes are elucidated. In the foreseeable future we will not know all the causes of disease, or how to avoid all the disease burden attributable to genetic causes. Nonetheless, multicausality means that in many cases considerable gains can be achieved by reducing the risks to health that are already known.

Sources: *(27, 28)*

Standard WHO age groups were chosen (0–4, 5–14, 15–24, 25–44, 45–59, 60–69, 70–79, and 80+ years) and epidemiological subregions were based on WHO regions, subdivided by mortality patterns (see the List of Member States by WHO Region and mortality stratum).
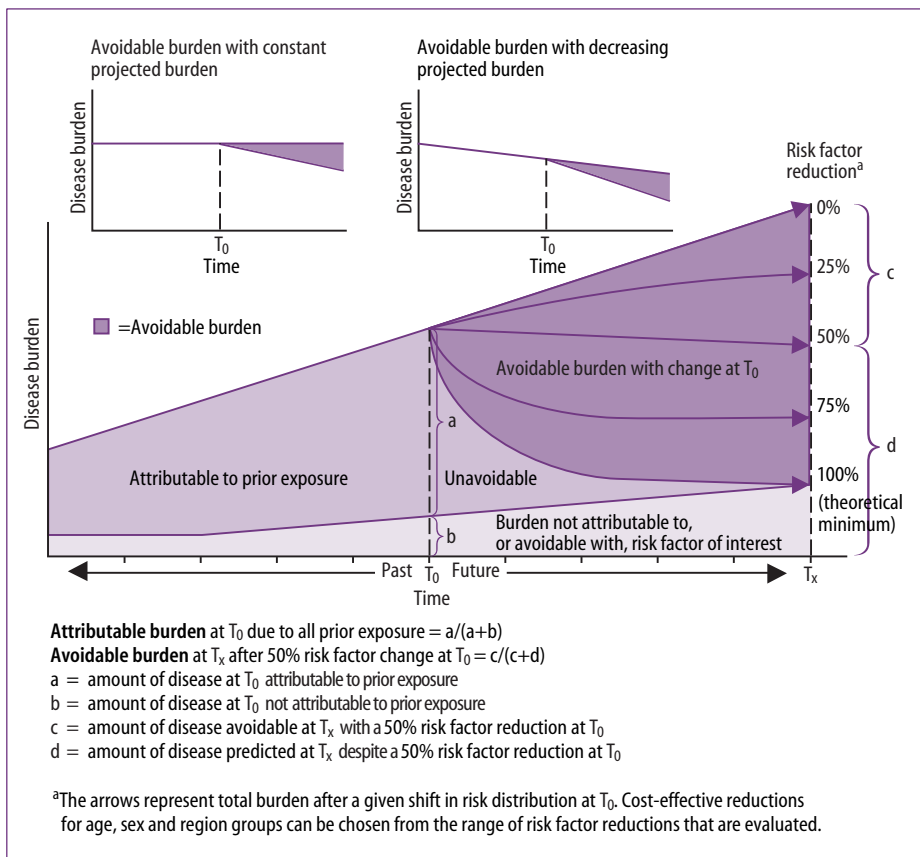
The methodology involved calculating population attributable risk, or where multi-level data were available, potential impact fractions. These measures estimate the proportional reduction in disease burden resulting from a specific change in the distribution of a risk factor. The potential impact fraction (PIF) is given by the following equation:

$$PIF = \frac{\sum\limits_{i=1}^{n} P_i\,(RR_i - 1)}{\sum\limits_{i=1}^{n} P_i\,(RR_i - 1) + 1}$$

where *RR* is the relative risk at a given exposure level, *P* is the population level or distribution of exposure, and *n* is the maximum exposure level.

Potential impact fractions require three main categories of data input, as summarized in Figure 2.5. The relationship between these key input variables and the basic methodology involved in calculating and applying population attributable fractions is summarized in Figure 2.6. It is clear from Figure 2.6 that risk factors that are more prevalent or that affect common diseases can be responsible for a greater attributable burden than other risk factors that have much higher relative risks.

Figure 2.4 Attributable and avoidable burdens

Attributable burden at T_0 due to all prior exposure = a/(a+b)
Avoidable burden at T_x after 50% risk factor change at T_0 = c/(c+d)
a = amount of disease at T_0 attributable to prior exposure
b = amount of disease at T_0 not attributable to prior exposure
c = amount of disease avoidable at T_x with a 50% risk factor reduction at T_0
d = amount of disease predicted at T_x despite a 50% risk factor reduction at T_0

[a] The arrows represent total burden after a given shift in risk distribution at T_0. Cost-effective reductions for age, sex and region groups can be chosen from the range of risk factor reductions that are evaluated.

Choosing and defining risks to health

The risk factors assessed in this report were chosen with the following considerations in mind.

- Potential global impact: likely to be among leading causes of disease burden as a result of high prevalence and/or large increases in risk for major types of death and disability.
- High likelihood of causality.
- Potential modifiability.
- Neither too specific nor too broad (for example, environmental hazards as a whole).
- Availability of reasonably complete data on risk factor distributions and risk factor–disease relationships.

There is unavoidably an arbitrary component to any choice of risk factors for assessment, as time and resource constraints will always operate and trade-offs will be required. For example, some factors like global warming where data are substantially incomplete may nonetheless be of such potential importance that they should be included and their impact estimated based on possible scenarios and theoretical models. These trade-offs should be made clear when the data sources, methods and results are reported in detail, including estimation of uncertainty.

Clearly, one risk factor can lead to many outcomes, and one outcome can be caused by many risk factors. For each possible risk factor–burden relationship, a systematic and documented assessment of causality was performed. Many approaches have been proposed for the assessment of causality. One that is widely known and reasonably well accepted is the set of "standards" proposed by Hill *(29)*. These are not indisputable rules for causation, and Hill emphasized that they should not be taken directly as a score. It is, however, widely agreed that a judgement of causality should be increasingly confident with the accumulation of satisfied standards including the following.
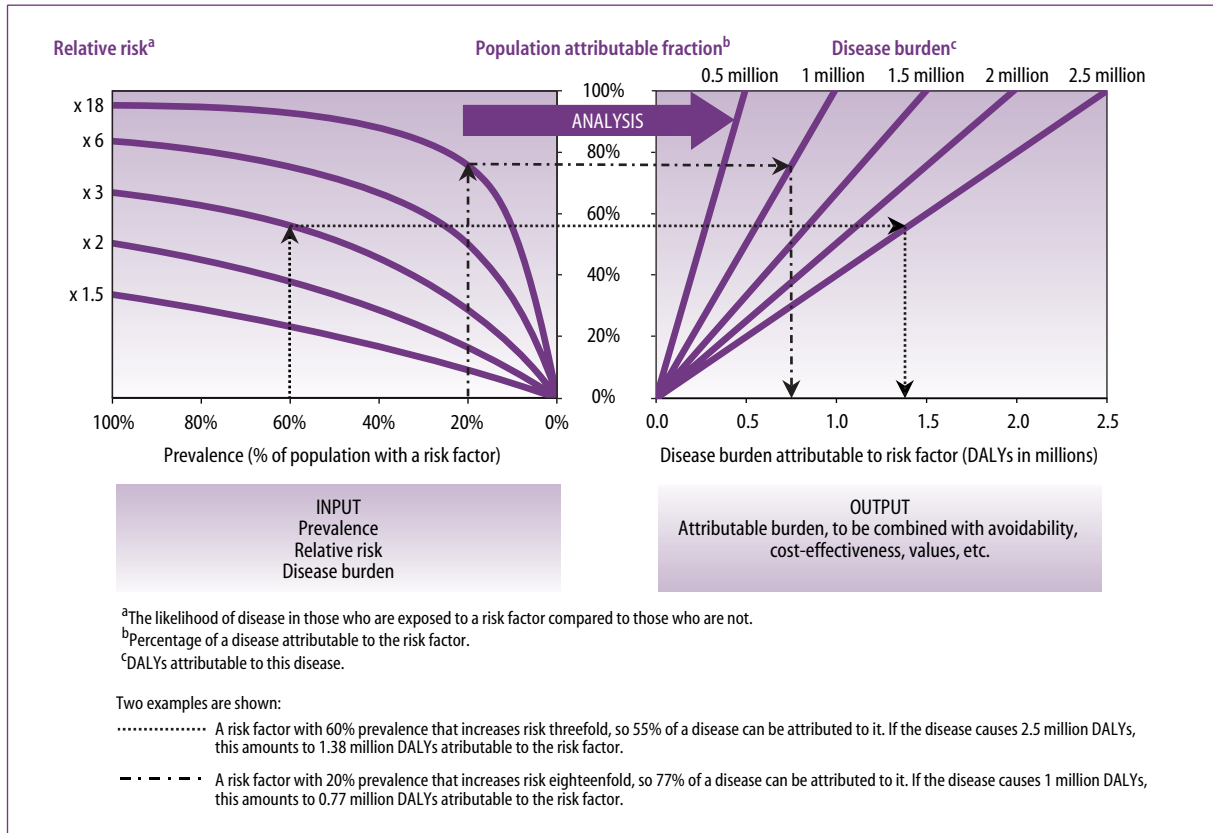
Figure 2.5 Key inputs for assessment of attributable and avoidable burdens
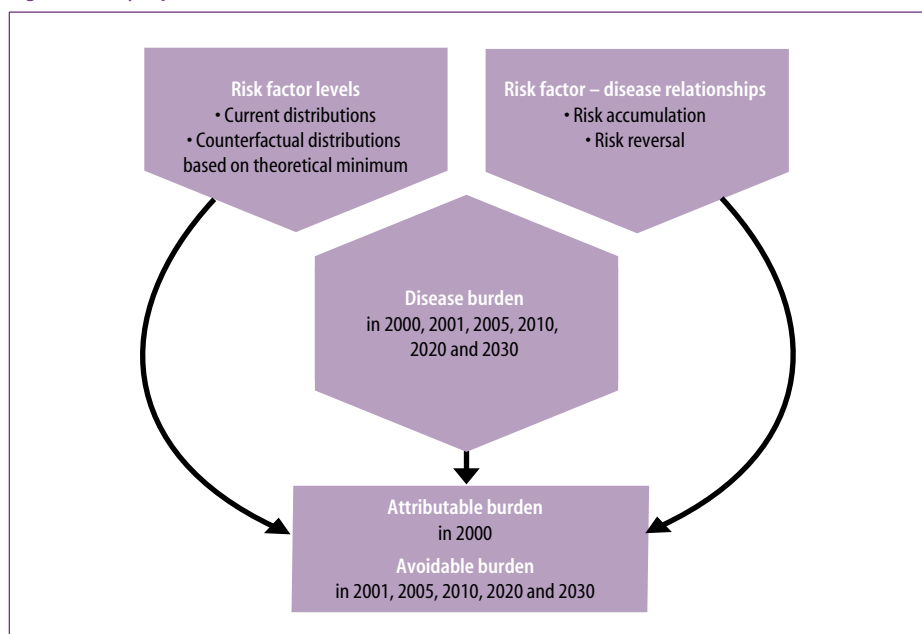
- **Temporality** – Cause must precede effect in time.
- **Strength** – Strong associations that are credible are more likely to be causal than weak associations, because if a strong association were wholly to result from some other factor, then it is more likely that other factor would be apparent. But a weak association does not rule out a causal connection.
- **Consistency** – Repeated observations of associations in different populations under different circumstances increase a belief that they are causal. But some effects are produced by their causes only under specific circumstances.
- **Biological gradient** – Presence of a dose–response curve suggests causality, although some causal associations do have a threshold, and for others the dose–response can arise from confounding factors.
- **Plausibility** – Biological plausibility is relevant, but can be subjective and is based on current level of knowledge and beliefs.
- **Experimental evidence** – Experimental evidence, in which some groups differ only with respect to the risk factor of interest, provides powerful evidence of causation. But evidence from human experiments is often not available.

Systematic assessments of causality, along with the other criteria listed above, led to the inclusion in this report of a number of risks to health and affected outcomes, which are discussed in Chapter 4.

Figure 2.6 Determination of attributable burden, taking account of prevalence and relative risk

Relative risk[a]   Population attributable fraction[b]   Disease burden[c]

Prevalence (% of population with a risk factor)

Disease burden attributable to risk factor (DALYs in millions)

INPUT
Prevalence
Relative risk
Disease burden

OUTPUT
Attributable burden, to be combined with avoidability, cost-effectiveness, values, etc.

[a]The likelihood of disease in those who are exposed to a risk factor compared to those who are not.
[b]Percentage of a disease attributable to the risk factor.
[c]DALYs attributable to this disease.

Two examples are shown:

⋯⋯⋯⋯⋯ A risk factor with 60% prevalence that increases risk threefold, so 55% of a disease can be attributed to it. If the disease causes 2.5 million DALYs, this amounts to 1.38 million DALYs atributable to the risk factor.

–·–·–·– A risk factor with 20% prevalence that increases risk eighteenfold, so 77% of a disease can be attributed to it. If the disease causes 1 million DALYs, this amounts to 0.77 million DALYs atributable to the risk factor.

ESTIMATING CURRENT RISK FACTOR LEVELS AND CHOOSING COUNTERFACTUALS

Risk factor levels in the population are the first main data input in estimating potential impact fractions. Extensive searches were required to estimate risk factor levels by the 224 age, sex and country groups used as the basis for analysis, particularly for data in economically developing countries. For all risk factors, there was a need to extrapolate data to some age, sex and country groups for which direct information was not available. Wherever possible, this extrapolation was based on generalizing from a particular subgroup that had similar health, demographic, socioeconomic or other relevant indicators.

The theoretical minimum was chosen as the counterfactual for all risk factors. For risk factors for which zero is not possible (for example, cholesterol), the theoretical minimum was the distribution associated with lowest overall risk. For some exposures (such as alcohol) there may be subgroups (by region, age or sex) for which zero exposure may not always be associated with the lowest risk. To maximize comparability, however, the theoretical minimum counterfactual was taken to be the same across population groups. This aided overall interpretation of the results, avoiding "shifting goal posts", yet still allowed for estimation of when minimum risks occurred at non-zero levels. Since policy-relevant reductions are likely to vary by, for example, age, sex or region, a range of estimates was made for counterfactual distributions at set intervals between the current situation and the theoretical minimum.

For the purposes of this report, risk factors were defined in light of data availability, the requirement for consistency, and a preference to assess multiple levels of exposure and hence the likely impact of shifting the risk factor distribution in the population.

ESTIMATING CURRENT AND FUTURE DISEASE AND INJURY BURDEN

The second data input into potential impact fractions is information on amounts of burden of disease and injury in the population, by age, sex and region. Current and future disease and injury burden was estimated as part of the ongoing global burden of disease project *(30)*.

ESTIMATING RISK FACTOR–BURDEN RELATIONSHIPS

The third data input into potential impact fractions comprised estimates of risk factor–burden relationships by age, sex and subregion. For most risks, direct information on such relationships came only from developed countries. This highlights the importance of assessing generalizability of data, in view of the need to extrapolate results to age, sex and region groups for which direct evidence is not available. For risk factor levels, there is often no particular reason to expect levels to be consistent between regions. Risk factor–disease relationships will, however, often be more generalizable, since they may, at least in part, be intrinsic biological relationships. Consistency between the results of reliable studies conducted in different settings is an indicator of causality and generalizability. While the representativeness of a study population is an essential component of extrapolating results for risk factor levels, study reliability and comparability will often be more important in assessing risk factor–disease relationships. Since relative risks tend to be the most generalizable entity, these were typically reported. When relative risk per unit exposure varied between populations, this was incorporated wherever possible. For example, the relative risk for current tobacco smoking and heart disease appears to be less in the People's Republic of China than in North America and Europe, principally because of a shorter history of smoking among the Chinese.

Estimates of avoidable burden

Current action to target risks to health can change the future but cannot alter the past. Future disease burden can be avoided but nothing can be done about attributable burden. For this analysis, avoidable burden was defined as the fraction of disease burden in a particular year that would be avoided with a specified alternative current and future exposure. Estimates of avoidable burden are particularly challenging, given that they involve all the uncertainty in the estimates of attributable burden plus those in a number of extra data inputs, described below.

- Projected global burden of disease.
- Risk factor levels under a "business as usual" scenario. Some projections were based on observed trends over the past few decades (for example, childhood malnutrition) and others based on models using exposure determinants and their expected trends (for example, physical inactivity, indoor smoke from solid fuels).
- Projected risk factor levels under a counterfactual scenario – for example, a 25% transition towards the theoretical minimum, starting from 2000 and remaining at 25% of the distance from business as usual and theoretical minimum exposure.
- Estimates of risk "reversibility". These may occur to different extents and over different time frames for various risk factor–burden relationships. After some time, the excess risk of a "previously exposed" group may reach that of the "never exposed" group, or may only be partially reversed. For all acute or almost-acute hazards, including injuries and childhood mortality risk factors, immediate reversibility was assumed. The impact of cessation of the use of alcohol and illicit drugs on neuropsychological diseases, while known to be delayed, was assumed to be fully reversed by 2010, the earliest reporting year. Thus ex-exposed in 2010 were assumed to have the same risk as never-exposed. For blood pressure *(31, 32)* and cholesterol *(33)*, most or all of the risks were assumed to be reversed within five years and all within 10 years. Since more distal risk factors such as obesity and physical inactivity operate in large part through these exposures, these data formed the basis of risk reversibility for other major causes of cardiovascular disease assessed here. For tobacco, data on risk reversibility after smoking cessation was obtained from the large American Cancer Society's Cancer Prevention Study *(34)*. This evidence shows that most excess risk for cancer, and almost all for vascular disease, is avoided within a decade of cessation. In the absence of similar studies for other risk factors, these data were also used to estimate the temporal relation between exposure reduction for other carcinogens and airborne particles and cause-specific disease outcomes. Lastly, a time-lag factor was used when appropriate, for example with childhood sexual abuse, reflecting the delay between cessation of abuse and the lower risks of adult mental health problems.

Estimating the joint effects of multiple risks

The main estimates presented in this report are for burden resulting from single risk factors, with the assumption that all others are held constant. Such estimates are valuable for comparative assessments, but there is also a need for estimates of the net effects of clusters of risk factors. When two risks affect different diseases, then clearly their net effects are simply the sum of their separate effects. However, when they affect the same disease or injury outcomes, then the net effects may be less or more than the sum of their separate effects. The size of these joint effects depends principally on the amount of prevalence overlap (for example, how much more likely people who smoke are to drink alcohol) and the biological effects of joint exposures (for example, whether the risks of alcohol are greater

among those who smoke) *(27)*. However, these have very little influence on net effects when the population attributable fractions are high for individual risk factors, as was often the case in these analyses – for example, more than 80% of diarrhoeal disease was attributed to unsafe water, sanitation and hygiene. The data requirements for ideal assessment of joint effects are substantial and assumptions were made of multiplicatively independent relative risks, except for empirical assessments of joint effects for two main clusters – risk factors that are major causes of cardiovascular disease and those that are major causes of childhood mortality. An alternative approach is outlined in Box 2.6. This simulation method based on individual participant data from a single cohort is compatible with the joint effects estimated from aggregate data as described above.

ESTIMATES OF UNCERTAINTY

Confidence intervals for the attributable burden were estimated by a simulation procedure *(37)* incorporating sources of uncertainty from domains of the exposure distribution and the exposure–response relationships. Briefly, the method involved simultaneously varying all input parameters within their respective distributions and reiterating the calculation of the population attributable fraction. An uncertainty distribution around each estimate of population attributable fraction was obtained after 500 iterations of the simulation and, from this, 95% confidence intervals were derived. Each risk factor group provided data characterizing the uncertainty in the estimates of exposure distribution and exposure–response relationships. To the extent possible, the uncertainty estimates accounted for statistical uncertainty in available data as well as uncertainty in the methods used to extrapolate parameters across regions or countries.

Still further refinements would improve the current estimates and are not reflected in the reported uncertainty indicators. These include uncertainty in the burden of disease estimates; lack of data on prevalence among those with disease, such data ideally being

---

Box 2.6 Estimating the combined effects of cardiovascular disease risk factors

There are several major risk factors for cardiovascular disease, and the actions of some are mediated through others. For example, overweight and obesity increase the risk of coronary disease in part through adverse effects on blood pressure, lipid profile and insulin sensitivity. The causal web model of disease causation reflects the fact that risk factors often increase not only the risk of disease, but also levels of other risk factors.

Separate estimation of the effects of individual risk factors does not typically take into account the effect of changes on the levels of other risk factors. One way of achieving this is to use measured relationships between the levels of the different risk factors to simulate what would happen in a 'counterfactual cohort', if levels of one or more risk factors were altered. The relationship between levels of risk factors

Sources: *(35, 36)*.

and disease can then be used to determine the rate of disease in the simulated cohort. The proportion of people in the population that would develop coronary heart disease (CHD) under each intervention is a counterfactual (unobserved) quantity. The g-formula (Robins, 1986) is a general nonparametric method that allows estimation of the counterfactual proportions under the assumption of no unmeasured confounders. This approach was taken using data from the Framingham Offspring Study on the risk factors body mass index, smoking, alcohol consumption, diabetes, cholesterol and systolic blood pressure.

A formula for predicting risk of CHD, given risk factor history, was estimated, and also the history of the other risk factors was used to predict future values of each risk factor following changes in some. A simulated cohort was generated from the study by sampling with replacement and various

scenarios were applied to the cohort to assess the impact on 12-year CHD risk, taking into account the joint effects of all the risk factors. A combination of complete cessation of smoking, setting all individuals' body mass index to no more than 22, and a simulated mean cholesterol level of 2.3 mmol/l and corresponding variance was estimated to halve the 12-year risk of CHD in both women and men. The estimated effect of all three interventions – a 50% relative risk reduction in coronary disease – was less than a crude sum of the separate effects (19%, 9% and 31%, respectively). This is because some people suffered CHD resulting from the joint actions of two or more of the risk factors, and this model estimates the size of these joint effects.

required in population attributable fraction estimates that incorporate adjusted relative risks (38); and the likelihood that reduction of exposure to risks such as unsafe medical injections in 2000 would lead to less infection in subsequent years and also a smaller pool of infected people from whom transmission could be propagated. Finally, competing risks – for example, someone saved from a stroke in 2001 is then "available" to die from other diseases in ensuing years – have not been estimated, which is likely to lead to an overestimate of the absolute amount of attributable and avoidable disease burden, although it may not substantially affect the ranking of risk factors. However, competing risks are accounted for in the dynamic models that assessed the joint effects of risks on healthy life expectancy. This topic, along with appropriate discount rates, is considered in Chapter 5.

REFERENCES

1. Last JM, editor. *A dictionary of epidemiology*. New York: Oxford University Press; 2001.

2. Slovic P. Informing and educating the public about risk. *Risk Analysis* 1986; 6:403-15.

3. Hope T. Rationing and life-saving treatments: should identifiable patients have higher priority? *Journal of Medical Ethics* 2001; 3:179-85.

4. Murray CJL, Lopez AD. Quantifying the burden of disease and injury attributable to ten major risk factors. In: Murray CJL, Lopez AD, editors. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge (MA): Harvard University Press; 1996. p. 295-324.

5. Murray CJL, Lopez AD. Global patterns of cause of death and burden of disease in 1990, with projections to 2020. In: *Investing in health research and development. Report of the Ad Hoc Committee on Health Research Relating to Future Intervention Options*. Geneva: World Health Organization; 1996.

6. Peto R, Lopez AD, Boreham J, Thun M, Heath CW. Mortality from tobacco in developed countries: indirect estimates from national vital statistics. *Lancet* 1992; 339:1268-78.

7. English DR, Holman CDJ, Milne E, Winter MG, Hulse GK, Codde JP, et al. *The quantification of drug-caused morbidity and mortality in Australia, 1995*. Canberra: Commonwealth Department of Human Services and Health; 1995.

8. Smith KR, Corvalan CF, Kjellstrom T. How much global ill health is attributable to environmental factors? *Epidemiology* 1999; 10:573-84.

9. Rodgers A, Lawes C, MacMahon S. 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* 2000; 18 (Suppl):S3-S5.

10. *Determinants of the burden of disease in the European Union. Sweden 1997*. Stockholm: National Institute of Public Health Sweden; 1997.

11. Mathers C, Vos T, Stevenson C. *The burden of disease and injury in Australia*. Canberra: Australian Institute of Health and Welfare; 1999.

12. *Our health, our future. The health of New Zealanders 1999*. Wellington: Ministry of Health; 1999.

13. Murray CJL, Lopez AD, editors. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge (MA): Harvard University Press; 1996.

14. Cacioppo JT, Berntson GG, Sheridan JF, McClintock MK. Multi-level integrative analyses of human behavior. Social neuroscience and the complementing nature of social and biological approaches. *Psychological Bulletin* 2000; 126: 829-43.

15. Ryff CD, Singer B. Biopsychosocial challenges of the new millennium. *Psychotherapy and Psychosomatics* 2000; 69(4):170-7.

16. Ryff CD, Singer B. The role of emotions on pathways to positive health. In: Davidson RJ, Goldsmith HH, Scherer K, editors. *Handbook of affective science*. New York: Oxford University Press; 2002.

17. Jessor R, Van Den Bos J, Vanderryn J, Costa FM, Turbin M S. Protective factors in adolescent problem behaviours: moderator effects and developmental change. *Developmental Psychology* 1995; 31:923-33.

18. *Broadening the horizon: balancing protection and risk for adolescents*. Geneva: World Health Organization; 2002. WHO document WHO/FCH/CAH/01.20 (Revised).

19. Jessor R, Turbin MS, Costa FM. Risk and protection in successful outcomes among disadvantaged adolescents. *Applied Developmental Science* 1998; 2:194-208.

20. Rose G. *The strategy of preventive medicine*. Oxford: Oxford University Press; 1992.

21. McMichael AJ, Beaglehole R. The changing global context of public health. *Lancet* 2000; 356:495-99.

22. Winslow CEA. *The cost of sickness and the price of health*. Geneva: World Health Organization; 1951. WHO Monograph Series, No.7.

23. Barker DJP. *Mothers, babies and disease in later life*. 2nd ed. London: BMJ Publishing Group; 1998.

24. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology* 2002; 31:285-93.

25. Rose G. Sick individuals and sick populations. *International Journal of Epidemiology* 1985; 4:32-8.

26. Rose G, Day S. The population mean predicts the number of deviant individuals. *British Medical Journal* 1990; 301:1031-4.

27. Rothman KJ, Greenland S. *Modern epidemiology.* 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1998.

28. Magnus P, Beaglehole R. The real contribution of the major risk factors to the coronary epidemics. Time to end the "only-50%" myth. *Archives of Internal Medicine* 2001; 161:2657-60.

29. Hill AB. The environment and disease: association or causation? P*roceedings of the Royal Society of Medicine* 1965;58:295-300.

30. Murray CJL, Lopez AD, Mathers CD, Stein C. *The Global Burden of Disease 2000 project: aims, methods and data sources.* Geneva: World Health Organization; 2001. Global Programme on Evidence for Health Policy Discussion Paper No. 36 (revised).

31. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, et al. Blood pressure, stroke, and coronary heart disease. Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; 335:765-74.

32. Collins R, Peto R, MacMahon S, Hebert P. Fiebach NH, Eberlein KA, et al. Blood pressure, stroke, and coronary heart disease. Part 2. Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990; 335:827-38.

33. Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *British Medical Journal* 1994; 308:367-73.

34. *Tobacco control country profiles*. Atlanta (GA): American Cancer Society; 2000. Also available at: URL: http://www1.worldbank.org/tobacco/countrybrief.asp

35. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: applications to control of the healthy workers survivor effect. *Mathematical Modeling* 1986; 7:1393-512.

36. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; 97:1837-47.

37. King G, Tomz M, Wittenberg J. Making the most of statistical analysis: improving interpretation and presentation. *American Journal of Political Science* 2000; 44:341-55.

38. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *American Journal of Public Health* 1998; 88:15-9.