



Understanding Health: Concepts, Relationships, and Dynamics

The schema of the epidemiologic transition, which Chapter 1 outlined, describes the general character of change in mortality patterns that richer countries have experienced over the course of two centuries and through which the developing world has been moving very rapidly in recent decades. It also provides some general help in thinking about how human health may change over the next half century. But as a broad framework, it leaves several critical, more specific questions fundamentally unaddressed.

Exactly what is human health? Mortality offers considerable biological clarity and methodological simplicity. All individuals will die, and we can measure the event in a “yes/no” fashion. In this chapter, and throughout the volume, we pursue the more ambitious goal of classifying the immediate cause of death and of providing at least some understanding of the character of ill-health as well as mortality.

Beyond the most immediate causes or character of death and ill-health, what are the

deeper drivers of population health? In particular, we have an interest in understanding those factors over which we may, as individuals or as societies, exercise some influence. Such exploration in this chapter will necessarily take us well beyond health and demographic systems and into consideration of economic, governance, and environmental systems.

To what does health, in turn, contribute? Amartya Sen’s human capabilities framework positions health—and freedom from the burdens, costs, and risks of poor health—as the fundamental element in a constellation of capabilities (including education and political freedom) that are essential to human flourishing (Sen 1985; 1987; 1998). Such a broad perspective on the importance of health leaves us with a wide set of justifications for health promotion: as a right in and of itself, as a marker of a just and well-governed society, as a raw material for productivity and growth, and as the foundation for broader human development.

■ This chapter provides a conceptual overview of health and begins consideration of health's position within the larger framework of human and environmental systems. ■

This chapter provides a conceptual overview of health and begins consideration of health's position within the larger framework of human and environmental systems. The rest of the volume will elaborate and explore the forces that affect health outcomes and also, notably in Chapter 7, the implications of health outcomes for broader human development. We seek to understand a complex web.

Measuring the Disease Burden

When analysts think about current and future levels of health (or as has been more typical, of ill-health), they usually refer to two related categories of outcomes: *mortality* (death) and *morbidity* (illness or disease). The collection of mortality and morbidity statistics by sex, age, and cause dates back to at least the 17th century in some parts of the world. Still, in spite of decades of investment and improvement in data quality, many measurement challenges remain.

Observing mortality

Families and societies have important reasons for the registration of deaths, including ensuring safe disposal of the body, cessation of state benefits provided to the living, and initiation of benefits for survivors. Indeed, almost all states have vital registration systems that nominally collect at least basic mortality information. However, the completeness of the data varies widely between countries.¹

Though high-quality data remain elusive, the United Nations Population Division (UNPD) uses an array of well-established techniques to estimate all-cause (as opposed to cause-differentiated), age-specific death rates since 1950 in its World Population Prospects database (UNPD 2009b). These data and estimates of mortality from all causes provide a critical anchor or base for estimating and reconciling societal measures of cause-specific mortality and morbidity. In turn, the World Health Organization (WHO) encourages national death registries to categorize deaths according to International Classifications of Diseases (ICD) criteria,² which provide highly detailed cause codes for clinicians and public health officials to assign to each death. Recording and reporting inconsistencies and gaps, however, are great.

WHO, in its efforts to systematize understanding of causes of death and their

patterns, developed a high-level classification system that organizes the ICD detailed causes of death into three major cause-groups (Murray and Lopez 1996a: 119). These groups, described briefly below, have become the standard high-level classification system of major causes of death and are used extensively throughout this volume.

Group I diseases are primarily *communicable diseases (CDs)* caused by infectious agents outside the body. Group I also includes all other causes of maternal and perinatal mortality as well as nutritional deficiencies. In combination, these are the diseases that tend to prevail in the early stages of the epidemiologic transition. The characteristics that unify them include their preventability (often at low cost), their outsize effect on vulnerable populations (especially the very young and women in childbirth), and the risks of infection to others. While most CDs are short in duration, resulting fairly quickly in either death or recovery, others (such as HIV/AIDS) are becoming increasingly chronic in nature. (As others often do, we use the term *communicable diseases* as shorthand to refer to all of Group I throughout this volume.)

Group II diseases are *noncommunicable diseases (NCDs)* resulting from genetic, cellular, or organic anomalies or degeneration occurring inside the body. While we can thus refer to NCDs as internal causes of death, external forces such as diet, tobacco use, and environmental factors influence many of them.³ NCDs tend to predominate in later stages of the epidemiologic transition, both because of the progression of underlying risks to health that accompany that transition and because of their tendency to accumulate as an individual reaches older ages. Many also share common characteristics with respect to the relative cost and difficulty of prevention and treatment.

Group III causes of death are *injuries*. Like CDs, injuries are external in cause, yet the external agent interacts with the body's skeletal or organ systems (not the immune system), and the sufferer poses no further risk to the health of others. Group III deaths are most likely among those with high exposure to risk (e.g., in the residence or workplace or through warfare or personal behavior), limited safety precautions, and limited access to treatment.

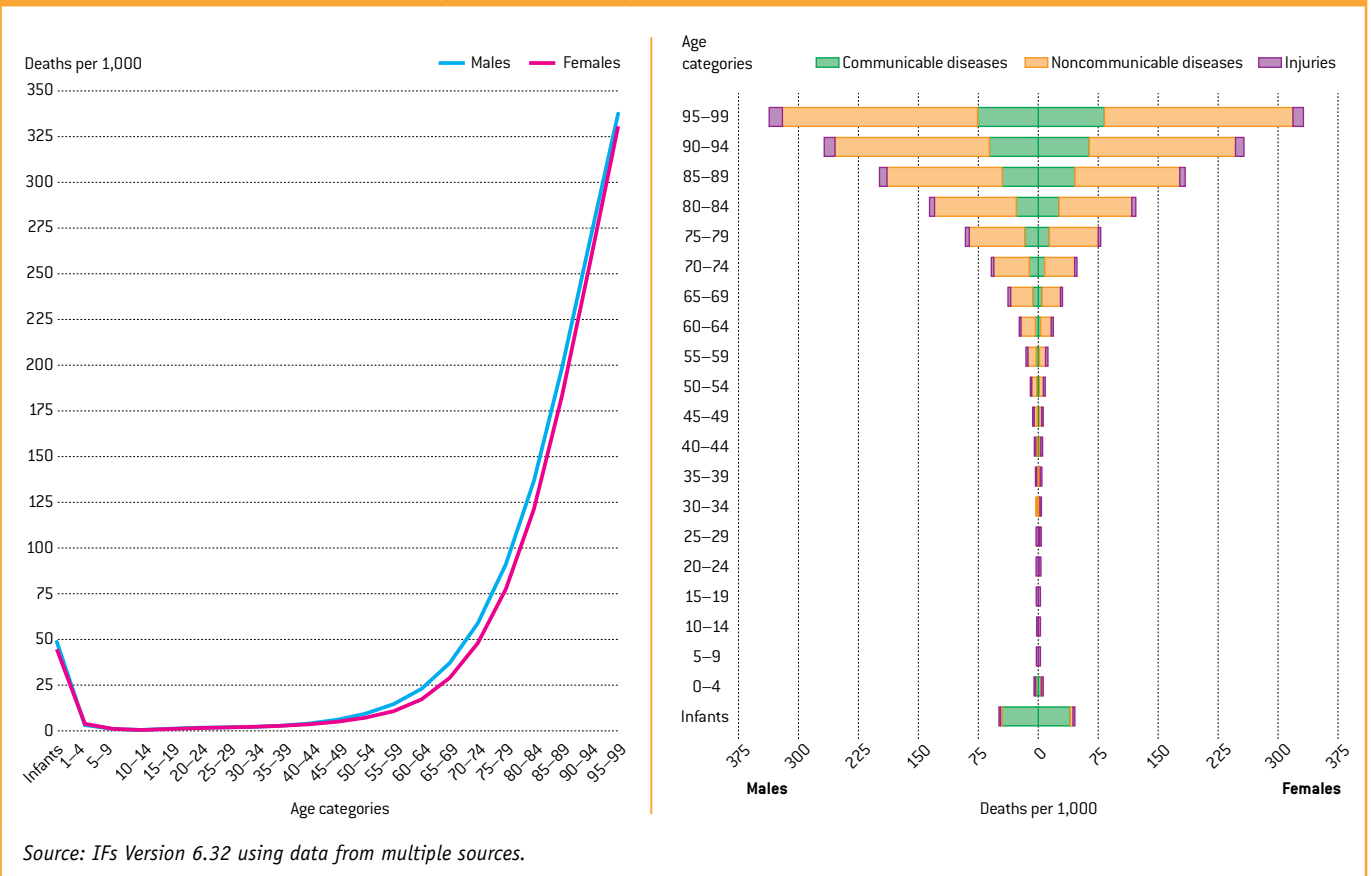
Limited data are available on cause-specific mortality and morbidity, even at the aggregated “major cause” level of Groups I, II, and III. In a 2005 study, WHO reported that only 23 of its 193 member countries provided “high-quality” cause-of-death data, and 75 member countries had provided no cause-of-death data since 1990 (Mathers et al. 2005). Largely in recognition of this data gap, WHO initiated the

Global Burden of Disease (GBD) study, which in 1990 began providing global estimates of age, sex, and cause-specific mortality and morbidity (Mathers and Loncar 2006; WHO 2008a). For most countries today, the development of cause-specific death rates begins with country cause-of-death registration systems but still relies largely on a series of estimates or imputations (Murray 2007).

Given mortality data or estimates, we can visualize mortality patterns using the J-curve (see Box 2.1 and the left side of Figure 2.1). A J-curve shows age- and sex-specific mortality rates for defined populations for an identified point in time. Plotting multiple J-curves on a single figure allows direct visual comparison across populations, causes of death, or time periods; for instance, the left side of Figure 2.1 compares mortality rates for males and females in Bangladesh from all causes in 2005. Mortality pyramids (see the right side of Figure 2.1) provide a different way to visualize the same data. It is not as easy to plot multiple pyramids

■ *Gaps in data, and data of varying quality, complicate efforts to measure global mortality—particularly mortality due to specific causes of death.* ■

Figure 2.1 J-curve and mortality pyramid (Bangladesh, 2005)



Box 2.1 The J-curve of mortality patterns

Demographers often refer to the mortality curve as the “J-curve,” reflecting the characteristic shape of the age-specific mortality pattern in humans and most other mammalian populations. The reproductive requirements of a skill- and resource-intensive species shape the human life cycle. Successful reproduction necessitates relatively predictable and low levels of mortality during the prime ages for childbearing and childrearing (Carey 2003; Kaplan 2006; Olshansky, Carnes, and Brody 2002). Durability in adulthood comes at a cost; humans acquire physical and mental skills over a relatively long period of development and dependency. Some deaths very early in life serve the evolutionary purpose of eliminating the least healthy so that resources can be focused on surviving children.

While the biological complexity necessary for a human’s high physical and mental functioning facilitates significant longevity—often well after

the childbearing years are concluded—the human body is nonetheless subject to a gradual process of senescence, or the gradual deterioration of cellular and organ function with age. Beginning at about age 25, mortality increases steadily due to the breakdown or obstruction of organs and cells (Carey 2003; Olshansky and Carnes 1994; 1997; Olshansky, Carnes, and Brody 2002). This pattern was first empirically substantiated and mathematically formalized by Benjamin Gompertz, an actuary working in Britain in the mid-19th century. Gompertz (1825) fit an exponential functional form to age patterns of British mortality data from the early 18th century, meaning that he observed that the hazard, or risk, of mortality increased in an accelerating fashion with advancing age from about age 25 through the end of life. This biologically determined age-pattern of mortality can suggest whether societies are reporting too few or too many deaths at certain ages.

on a single figure, but they do facilitate subgroup decomposition, as in the causes of death color-coded in Figure 2.1.

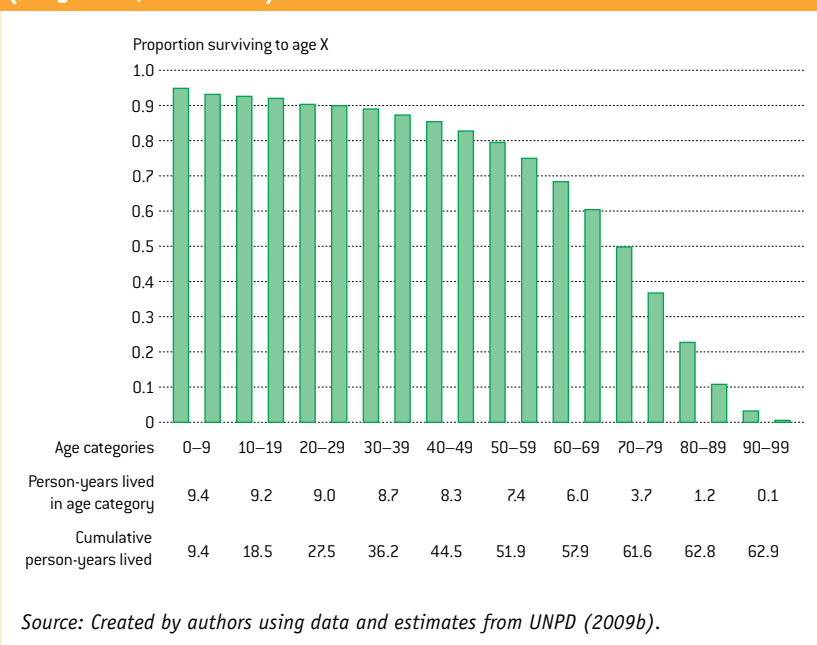
Summarizing mortality

A number of measures exist for summarizing mortality, each of which may paint a somewhat different picture. The most commonly reported measure at the societal level is the *crude death rate* (CDR), defined simply as the total number of deaths divided by the total population (often expressed as deaths per 1,000 people). While CDRs are simple to calculate, they conflate the effects of population distribution with the

effects of death rates at any given age. For example, in 2006, the reported crude death rate in Yemen was lower than in Sweden, which reported lower death rates at every age but had a substantially older population (UNPD 2009b).

An alternative approach to summarizing mortality rates is the calculation of life expectancy. Using age-specific mortality rates, a *life table* indicates the survival rate at each age of a hypothetical cohort of people who live their lives according to those rates; in turn, data from the life table can be used to construct a *survival curve* indicating the person-years lived in each of the age categories. Figure 2.2 shows a survival curve based on the estimated 2005–2010 death rates of the Bangladeshi population. We can visualize *life expectancy at birth*, or the average years a newborn could expect to live given current death rates, as the sum of the years of life lived by this population, as indicated at the bottom of Figure 2.2. Life expectancy at birth reflects the pace of mortality throughout the age distribution. For example, only 92 percent of Bangladeshis are expected to reach their 10th birthday, but the 8 percent who died before then nonetheless contribute some years of life; thus, a hypothetical average Bangladeshi can expect to live 9.4 out of 10 possible years through age 10. In older age groups, those dying in earlier periods make no contribution to life expectancy for those groups, and those dying during the period make only a partial contribution. The net result is a life expectancy at birth of 62.9 years for Bangladesh in 2005. As this example suggests, life expectancy is highly sensitive to infant and childhood mortality.

Figure 2.2 Survival curve and simplified life expectancy calculation (Bangladesh, 2005–2010)



The same approach can be applied to calculating life expectancy from any particular age. For example, the fact that a 65-year-old Bangladeshi has surpassed the population's life expectancy at birth does not mean that he will die imminently. Rather, given the death rates from that age on, he can expect to live an additional 12.8 years (life expectancy at age 65), or a total average age of 77.8 years.

In an earlier era, when the epidemiologic transition proceeded relatively smoothly from declining child mortality to declining adult mortality, life expectancy could fairly reliably capture a country's position on that path. Today, however, divergent patterns of adult mortality and a reduced correlation between life expectancy and child mortality have diminished the value of life expectancy as a single measure of health. For example, Yemen, the Russian Federation, and Namibia had very similar male life expectancies (60–61) in 2005. Yet the three populations had very different J-curves and survival curves. Yemen had a traditional pattern of high child mortality and high but “typical” adult mortality. On the other hand, the Russian Federation had child mortality levels comparable to a developed country coupled with extraordinarily high adult mortality rates related to noncommunicable diseases and injuries. And in a still different dynamic, Namibia's child mortality was significantly lower than Yemen's, yet high adult mortality due to HIV/AIDS (a Group I disease) brought down its life expectancy.

Other summary measures better capture the age-patterns of mortality. One pair of such measures is child and adult mortality rates. WHO defines the child mortality rate as “the probability of a child born in a specific year or period dying before reaching the age of five”;⁴ for this reason it is often referred to as the under-five mortality rate. Similarly, WHO defines adult mortality as the probability of surviving from one's 15th to 60th birthdays, a measure that ignores child mortality and old-age survival patterns (de Waal 2002). And studies of longevity sometimes use life expectancy starting from older ages to better isolate the effects of old-age mortality. For instance, the U.S. National Institute of Aging looks at populations 65–84 and 85 and older (sometimes 65–74, 75–84, and 85 and older; see Fries et al. 2000).

Measures such as life expectancy, child mortality, and adult mortality are very useful in conveying information about total mortality. To facilitate the attribution of mortality and changes in mortality to specific causes, however, it is useful to turn to a measure of life expectancy gap such as the *years of life lost* (YLL) measure used in the GBD study. Instead of describing the average years of life lived from birth, the YLL measures the number of years that a person loses upon dying compared to their life expectancy at the age of death. As noted above, one's life expectancy within one's own country can be calculated from any age. But GBD set the standard for how much longer one could have lived not based on one's own population, but based on that of Japan, which has the world's longest-lived population (Murray 1996: 16). Thus, an infant male dying in any country is estimated to have lost 80 years, the life expectancy of a male child born in Japan. The death of a 60-year-old man in any society would be associated with the 22-year life expectancy of a 60-year-old man in the Japanese standard population.

The YLL, which is generally aggregated across the population, offers two useful analytic benefits. First, by ascribing years of life lost at the time of death, we can easily disaggregate the contributions to YLL of specific causes of death just as we do for death rates. Second, the measure of total time lost to mortality is complementary to measures of disease prevalence and duration, to which we turn next.

Observing morbidity

The GBD project's ambitious goals included “generating the first comprehensive and consistent set of estimates of mortality and morbidity by age, sex and region for the world” (Mathers, Lopez, and Murray 2006: 45).

While morbidity (ill-health) and mortality (death) obviously are often related, in practice many individuals are never sick before death (e.g., accident victims) or are sick for very short periods (e.g., with avian influenza), while others experience long periods of morbidity from chronic diseases with widely varying levels of severity. Many of the sick also return to a state of health, though these individuals may differ in a variety of ways from those who were never sick.⁵ In all of these cases of sickness,

■ *We need to use measures beyond crude death rates and life expectancy at birth to understand the age-patterns of mortality.* ■

■ *Years of life lost measure an individual's premature death in comparison to a standard (long-lived) population.* ■

■ We are also interested in understanding morbidity—the incidence and prevalence of a disease and its impacts on the sick or injured person during life—but measurement is difficult. ■

the quality of one's existence is (presumably) neither as good as full health nor as bad as death. Put in terms of the earlier YLL example of a man dying at age 60, if he suffered from liver cancer for six years (from age 54), then the six years lost to cancer disability were qualitatively better than his 22 years of premature death, but worse than the preceding 54 years of health.

Clearly it would be desirable to account for the severity of illness in measures of health, yet there are serious conceptual and logistical impediments to doing so. Logistically, the science of collecting population-wide data on the diagnostic and functional dimensions of health remains in its infancy. The Demographic and Health Surveys (DHS) of the United States Agency for International Development and the Multiple Indicator Cluster Survey (MICS) conducted by the United Nations International Children's Fund are indispensable sources of population-wide morbidity data for children and women in most countries. More recently, WHO's World Health Survey (WHS) has begun the process of systematic estimation of reported health burdens and limitations in many countries, a process still largely based on a variety of indirect approaches described below.

Conceptually, severity may be indicated variously by the level of biochemical abnormality that defines a disease, by the extent of functional limitation, by the impact on quality of life, or by the risk of mortality. Each of these factors may vary independently over the course of a particular illness. Indeed, what counts as "healthy" may vary over time and space (Morrow and Bryant 1995). Finally, there is a difficult trade-off between the clarity offered by measuring the occurrence of a specific disease and the comprehensiveness of measuring aggregate morbidity irrespective of the disease. We discuss each of these issues in turn.

Measures of disease occurrence

We divide measures of disease occurrence into *incidence* (new cases of disease among a group at risk) and *prevalence* (presence of disease in a population). In a stock and flow model, incidence is a flow and prevalence is a stock. Prevalence reflects the combined effects of past disease incidence, recovery, and death; that is, it is the sum of those who contracted the disease

less those who recovered and those who died. It thus captures both the societal burden of an illness and the size of the pool of individuals experiencing a higher risk of mortality due to the disease. Although incidence measures are significant for disease surveillance and policy evaluation, collection is often costly and the actual onset of disease is often impossible to observe. The brevity of most acute illnesses poses a special challenge for collecting incidence or prevalence data in most poor countries.

Chronic diseases offer more opportunity to measure morbidity, both in terms of the marginal utility of doing so and their ease of measurement. Many societies, including some middle-income countries, conduct reasonably accurate surveillance of incidence (based on the timing of diagnosis) as well as prevalence. High-quality data are most common for the big three noncommunicable disease risks (cardiovascular diseases, cancers, and diabetes) as well as for HIV/AIDS, the communicable disease of greatest impact in many countries.

Even so, cost and lack of data-collection coverage severely limit development even of prevalence measures. Typically, prevalence data come from hospital reports (preferred) or surveys of individuals, which together systematically undercount prevalence if a significant number of patients have no access to a hospital or no recollection of a diagnosis. Recognizing this, researchers conducting household surveys in poor countries are increasingly assessing disease prevalence through methods such as physical examination and the collection of blood-drawn biomarkers for critical diseases such as HIV/AIDS, cardiovascular diseases, and diabetes. Limitations, however, persist: direct assessments are costly, often cover limited time spans and populations, and typically incorporate data on only a small number of diseases. Nonetheless, they represent a significant step forward, introducing at least the possibility of cross-national disease assessment on a wide scale.

In the end analysis, long-term forecasts are far more likely to employ measures of prevalence than incidence. In the absence of either prevalence or incidence data, analysts often rely on mortality data as either a predictor of, or proxy for, morbidity.

Measures of overall morbidity

Prevalence measures only bring us part of the way to the goal of creating a single morbidity rate that captures the overall burden of morbidity, much as the life expectancy or YLL measures do for mortality. The missing piece, noted earlier, is related to identification and measurement of the severity of each disease.

In order to address this deficiency, researchers with the GBD project developed the concept of a *disability weight* and used it to quantify each disease state along a 0–1 continuum (Lopez et al. 2006b).⁶ For example, the disability weight for liver cancer is 0.20 during the diagnostic and therapeutic stage, 0.75 in the metastasis stage, and 0.81 in the terminal stage. In the context of the earlier example, assume that five out of the six years of liver cancer were spent in the diagnostic and therapeutic stage, one year in the metastasis stage, and no significant time in the terminal stage. Thus, we could add $5 * 0.20 = 1$ for the first stage and $1 * 0.75 = 0.75$ for the final year of life, yielding a total of 1.75. These are referred to as the *years lived with disability* (YLD).

Combining morbidity and mortality into a single measure

The GBD researchers also developed a summary measure designed to combine the impact of both mortality and morbidity into a single statistic: *disability-adjusted life years* (DALYs), which combines years of life lost with years lived with disability.⁷ For instance, adding YLD and YLL in our ongoing example case of liver cancer, we find that the combined mortality and morbidity ($22 \text{ YLL} + 1.75 \text{ YLD}$) results in a total DALY estimate of 23.75 years.

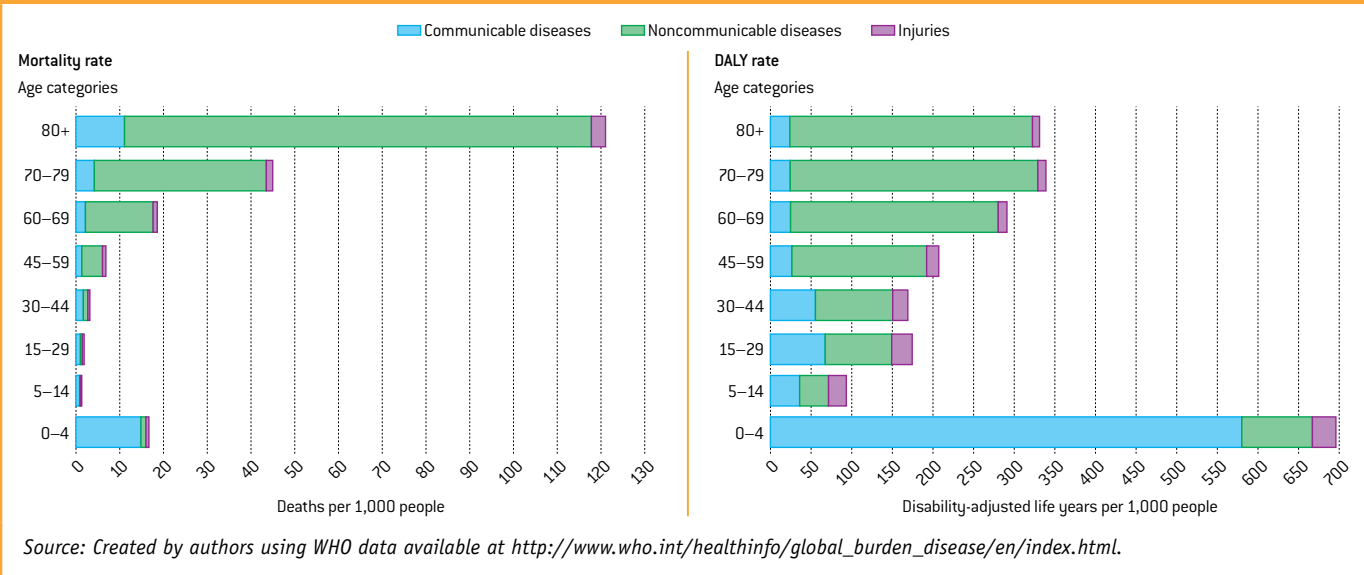
While the DALY represents a major step forward in disease burden reporting, it also presents analysts (and especially forecasters) with a number of problems related, first of all, to the disability weights. The base of data for estimating the disability weight for any specific disease remains quite thin. Initially, the disability weights were based on “tradeoff surveys” in which a panel of medical experts was asked to weigh the relative undesirability of particular disorders against one another (Arnesen and Kapiriri 2004); more recently, the estimation of disability weights has incorporated survey results linking key diseases to measures of physical limitation (Mathers,

Lopez, and Murray 2006: 50–51). Yet the disability weight concept does not relate to any specific dimension of disease severity, even though the varying effects of diseases on physical functioning, pain, depression, self-efficacy, or productivity could have widely divergent implications for the other dimensions of human progress included in a global health forecast (Anand and Hanson 1997; Arnesen and Nord 1999). Varying relationships between disability weights and the probability of death also complicate the estimation of disability weights; although the disability weight is closely concordant with mortality risk for a wide range of diseases, a number of important conditions (including chronic pain, severe psychiatric conditions, and vision and hearing impairment) carry high disability weights alongside relatively low probabilities of death. In summary, it is often difficult to tell whether the relative importance of certain conditions according to DALYs versus death rates derives from accounting for disability, from counting YLLs instead of deaths, or from two other adjustments applied to most DALY estimates to which we now turn—namely, age-weights and a discount factor.

Most published values of DALYs (and their YLL and YLD components) incorporate age-importance weights and a discount factor (typically 3 percent). The age-weights are primarily intended to reflect the potential impacts or forward linkages from death and disability to such societal functions as labor productivity and parenthood. These adjustments, however, are controversial, given variations in the relevance of different age groups for particular functions and in different societies, not to mention the ethical challenge of placing a value on a life. The choice of the discount rate also raises controversy in terms of what it implies with respect to the value of future life in relation to current life. Some researchers have pointed out that the combination of age-weights and discounting can create perverse cross-societal variations in the value of a life (Arnesen and Kapiriri 2004). While we follow the typical practice of incorporating discounting and age-weighting in our approach, we also make it possible with International Futures (IFs) to look at DALYs and their YLL and YLD components without discounting or age-weighting.

■ **Disability-adjusted life years combine the impacts of morbidity and premature death into a single summary measure of disease burden.** ■

Figure 2.3 Global female mortality and DALY rates by age and major disease group (2004)



Finally, there are issues particular to our analysis. One is that it is almost certainly unreasonable to assume that disability weights will not change over time, for instance in response to changing technology specific to the disease or to pain management. A second is that the use of current mortality and morbidity in our broader model already represents some of the forward linkages or impacts by age that are meant to be captured by the age-importance weights and the discount factor.

■ *To understand health outcomes, we need to understand distal and proximate causes of health within the context of biology, the natural environment, and human activity.* ■

Though not without problems and limitations, the DALY measure does usefully point our attention to what we miss if we ignore morbidity as a large part of the disease burden. In Figure 2.3, the left panel displays the age and causal structure of mortality for females globally in 2004 by the three broad disease categories reported by the GBD project, while the right panel illustrates the age and causal structure of DALYs in the same year. The two graphs together show that populations experience mortality and morbidity differently. Disease strikes across the age spectrum, with most deaths occurring among the elderly (from noncommunicable causes) or the very young (from communicable causes). The DALY panel not only shows a different age-profile, but also suggests a larger burden of disease from Group I causes than we might expect from the mortality data, especially among infants and children under the age of five.

Understanding Health Outcomes

Assuming that we can conceptualize and measure health outcomes in terms of both morbidity and mortality, the next task becomes understanding what causes differences and changes within them. Figure 2.4 provides a general overview of multiple categories of drivers of health outcomes. *Distal drivers*, or the deep drivers of health outcomes, include those that the GBD project uses in its forecasting, namely, income, education, and time as a proxy for other changes, including technological advance and broader social change. Distal drivers thus refer to societal conditions that do not have direct biological impacts but may enable biological interventions. There are also proximate drivers that more immediately and specifically relate to health outcomes (Lopez et al. 2006b: 2). These drivers often involve action that more directly addresses a human biological outcome, such as vaccination to prevent (or treatment to battle) a specific disease, behavioral changes to reduce obesity, or water-sanitation interventions to reduce exposure to biological agents.⁸ While proximate drivers are themselves strongly shaped by distal drivers over time, they suggest points of human leverage that do not require the fundamental restructuring of society (see Soper as edited by Kerr 1970), and many are amenable to intervention in even the poorest of countries.

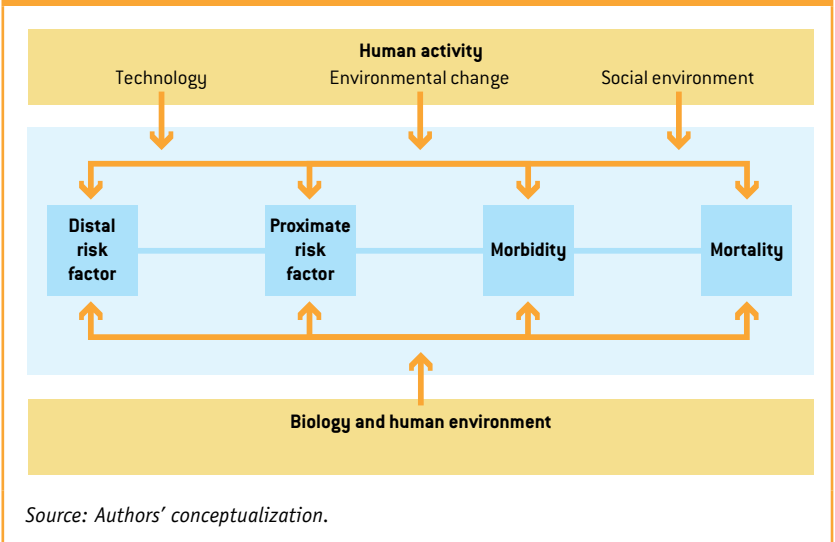
However, we can only properly understand distal and proximate drivers, and their strengths

and limitations, in the broader context of the biological and natural human environments. Health arises from a complex set of biological interactions occurring within our bodies, with other organisms (as in the case of infectious disease), with other humans (as in human-to-human disease transmission and many injuries), and with our broader ecosystem. Human morbidity and mortality would not exist in the absence of our biological vulnerability to external agents such as infectious diseases, physical trauma, and the gradual breakdown of bodily systems due to senescent decline.

A human biological framework also gives rise to the notion of *competing risks*, whereby no intervention can prevent mortality completely—it can only prevent mortality due to a specific cause and delay inevitable death due to some other cause (Cox 1959; Kalbfleisch and Prentice 2002). The deletion of a single cause of death can have a wide range of consequences for all-cause mortality, ranging from almost zero (if everyone dies of something else the next week) to a synergistic effect that saves lives due to multiple causes (if people become healthier in the process).⁹ As a result, the distal and proximate drivers of health not only have implications for current health and for health next week (when an epidemic may emerge), but also for longer-term future health.

In addition to the context provided by biology and the human environment and the interplay of distal drivers and proximate risk factors on health outcomes, some health literature makes a further distinction between distal drivers and even broader influencing factors. Smith and Ezzati (2005: 325) used the term *super-distal* to refer to factors that affect “essentially every disease, even if the pathways are not always well understood.” The analysis of this volume treats three variables as such super-distal drivers: technology, changes in the natural environment, and the social environment (examples include domestic social action such as health expenditures and global social initiatives such as poverty reduction). The super-distal drivers influence the course of all aspects of human development (e.g., the distal drivers of income and education as well as health outcomes with which income and education are associated). The distinction between distal and super-distal drivers is in

Figure 2.4 Drivers of health and disease



Source: Authors' conceptualization.

part a simple broadening and elaboration of the GBD project's distal category, but it is important because measures of income, education, and time alone cannot explain the profound global changes in health outcomes in recent decades.

These conceptual building blocks, including distal and proximate drivers in the context of biology and the super-distal factors linked to human action, provide some foundation for understanding the forces that drive change in human health. The remainder of this chapter, and then the broader volume, explores each in turn.

Distal Drivers of Health

Our current understanding of the distal drivers of health is affected by an ideologically charged debate between a growth-oriented perspective, in which living standards or a proxy such as income are seen as primary drivers of health improvements, and a support-led model in which health systems and interventions are viewed as the primary drivers (Pritchett and Summers 1996; Sen 1998). In the years since the onset of the HIV/AIDS epidemic, we have seen considerable convergence between these two models, driven by evidence of how improved living standards and health systems may interact to produce better outcomes. In fact, living standards, health systems, and health outcomes correlate highly across time, thereby obscuring meaningful understanding of the actual patterns of causation.

In this section, we introduce those distal drivers that the GBD project has thus far used—

■ WHO's *Global Burden of Disease* project uses income, education, and time (largely as a proxy for technology) as the distal drivers of health outcomes. ■

Income correlates highly with health outcomes at each point in time, but the cross-sectional relationship has shifted steadily upward across time. ■

GDP per capita, education, and time—and review the current evidence on other potential drivers. In subsequent sections we move to the proximate drivers that more directly mediate between distal drivers and health outcomes, as well as to the broader biological and human contexts that in turn shape these elements.

Income

For a number of years, the dominant paradigm among leading economists at the International Monetary Fund and at the World Bank suggested that living standards, measured typically by national GDP per capita, were the most important determinant of health. This growth-oriented approach first emerged with the work of Thomas McKeown, an English physician and epidemiologist who analyzed historical data from England and Wales for the period between 1837 and 1990. As a result of his analysis, McKeown attributed 50 percent of the reductions in mortality during this period (much of it preceding the advent of specific medical technologies associated with health improvements) to improved living standards

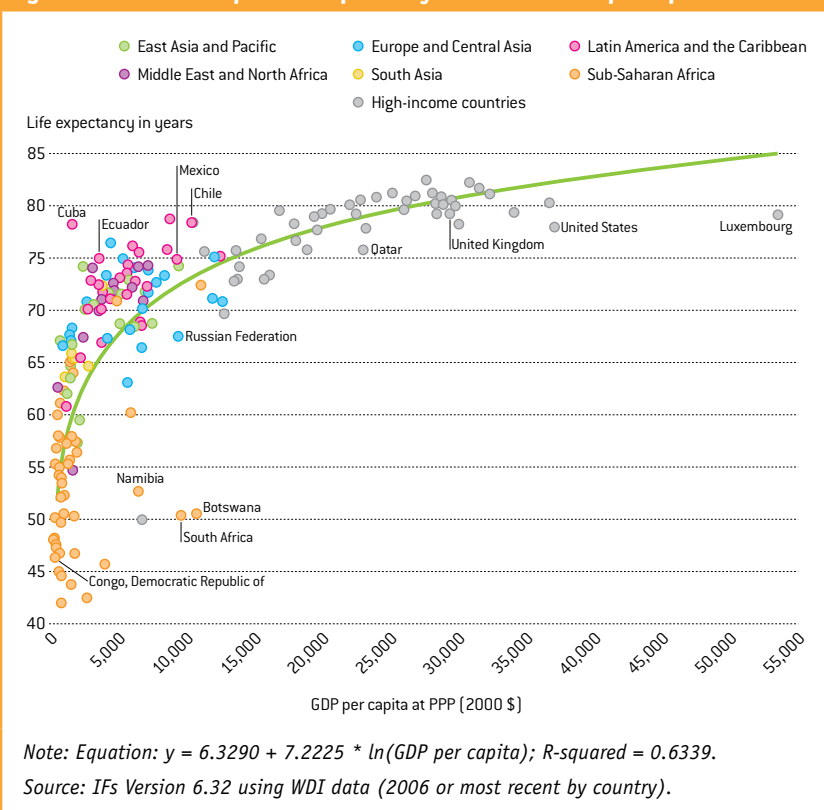
(McKeown 1976; McKeown and Record 1962). While subsequent research (Sen 1998) revised many of McKeown’s findings, his work pointed the way to a simple reality we now know to be consistent across historical periods—namely, that at any single point in time, between 65 and 90 percent of cross-national variation in human life expectancy at birth can be associated with a logged measure of GDP per capita (Filmer and Pritchett 1999; Pritchett and Summers 1996). Figure 2.5 shows this relationship in 2006.

The cross-sectional association is seductive in its simplicity, offering the possibility that one could model all future changes in health simply as a function of change in GDP per capita and that one could emphasize growth as the sole pathway of importance to improving human health outcomes. Yet the cross-sectional relationship is not all that it appears. First and most evidently, even a 65–90 percent correlation leaves considerable room for societies to outperform or underperform the level of health we would anticipate based on income alone. Second, there is uncertainty in the direction of causation and the influence of other factors in the relationship.

Nonetheless, considerable evidence does support a longitudinal, causal relationship between changes in income and subsequent improvements in health outcomes within a single country. The most-studied outcome is infant mortality, and the current “gold standard” estimate comes from Pritchett and Summers (1996). When they lagged infant mortality rates by five years with respect to changes in GDP per capita, Pritchett and Summers found a 24 percent decrease in infant mortality rates with a doubling of GDP per capita; controlling for level of education reduced this impact to 19 percent. Calculations based on these relationships lead them to conclude that “a country at the sample mean GDP would avert one death per 1,000 births if income were higher by 1%” (Pritchett and Summers 1996: 851).¹⁰

Yet some would argue that, as with the cross-sectional relationship, these within-country studies miss broader and more important trends emerging in the income/health relationship across multiple countries over multiple time periods. In a groundbreaking reconstruction of the relationship between income and health in 1905, 1935, and 1965, Samuel Preston (1975) observed that the

Figure 2.5 Relationship of life expectancy at birth and GDP per capita at PPP



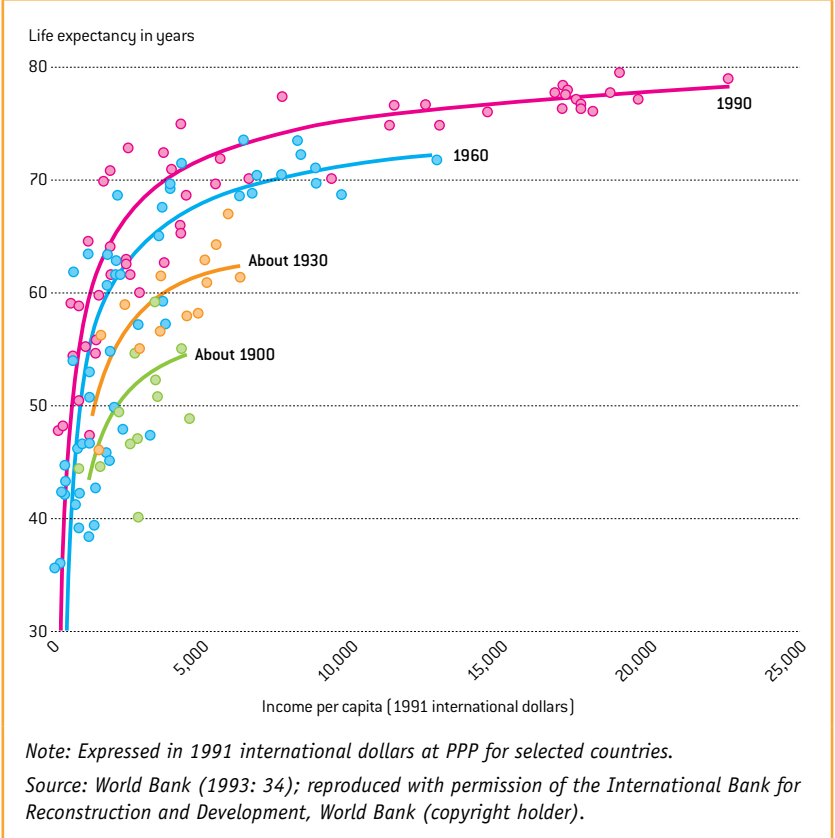
elasticity and shape of the cross-country income/health relationship was changing over time. This pattern, shown for four time points in Figure 2.6, clearly suggests that “the relationship between life expectancy and national income per head has shifted upwards during the 20th century” (Preston 1975: 236). Looking vertically, this means that over time the life expectancy associated with a particular GDP per capita has risen dramatically. Looking horizontally, we see that a nation “required an income level approximately 2.6 times higher in the 1930s than in the 1960s” to reach the same life expectancy.

Preston concluded that, over time, variation in income could explain only 10–25 percent of the variation in life expectancy (Preston 1975), and the World Bank’s influential 1993 *World Development Report* replicated this finding more recently (World Bank 1993). We thus face a challenging conundrum common to many forecasting efforts. No other single structural determinant of health will approach income’s magnitude or consistency across time period, cause, or country—yet it leaves 75–90 percent of the total variation over time unexplained.

Education

The roles of other structural factors such as education, health systems, and culture are potentially important, yet quite difficult to specify, in part because each of them tends itself to be correlated with income. Because living standards are such a powerful predictor of health at any single point in time, a useful entry point to the health systems literature has been to search for common patterns across the list of countries that perform better or worse on health outcomes than would be expected based on income alone. John Caldwell (1986) first used this approach by rank-ordering developing countries in 1983 in terms of income, life expectancy, and infant mortality rate and then qualitatively exploring the countries whose health outcome rankings were higher or lower than their income rankings. His analysis yielded a number of possible common factors relating to culture, gender, health spending, and health systems—factors to which we will return. One dominant theme emerged from the analysis. Better educated societies and societies with a greater tradition of widespread participation in education, particularly of women, had better-

Figure 2.6 Changes in the relationship between life expectancy and income per capita over time



than-expected health outcomes. Yet Caldwell’s analysis did not pinpoint the existence and direction of causality.

The relationship between education and health outcomes, even after controlling for income, is a strong one. In terms of causal effects, there is considerable evidence from the micro level that, in any society, individuals with higher levels of education attainment will be better able to take care of themselves and—arguably even more important—will be better able to take care of their children. At the macro level, structural regression models have isolated this relationship, with a uniquely identified effect of education on health outcomes (Boehmer and Williamson 1996; Frey and Field 2000; Lena and London 1993). Yet, none of these studies fully identifies the possible causation running from education to health or important confounding factors.

With regard to confounding factors, we note two of particular significance. First, a number of studies have identified that it is not education

● *While income and education explain much of the variation in health outcomes across countries, they also leave much unexplained (especially changes in health over time).* ●

● *Proximate risk factors, such as undernutrition and obesity, mediate between distal drivers and health outcomes and provide targets for focused health-related interventions.* ●

overall but rather women's education specifically that appears to be associated with better health outcomes (Shen and Williamson 1997; Summers 1994). While this might indicate simply that women's education is of greater causal significance as a determinant of health than men's, it might also indicate that women's education attainment, itself highly correlated with total education attainment, may be a proxy for an unobserved or untested societal factor. One such factor might be a greater degree of gender equity, which might result in narrower health gaps between men and women, or a greater level of women's empowerment, which might enable women to gain better access to prevention and treatment for children (Caldwell 1986; Frey and Field 2000; Nussbaum 2004).

Health and education are also subject to similar forces of demand and supply. Populations that demand greater educational opportunities may also demand greater health opportunities, and so the relationship between education and health may merely capture a societal taste for both outcomes. Similarly, there may be a strong correlation between a society's ability to provide quality educational and health services, whether via government, private, or nonprofit sectors. Thus, the association between education and health outcomes may merely capture the presence of a number of conditions that facilitate both health and educational achievement but do not indicate any kind of causal impact of education on health. We address many of these factors later, while noting that most of them at this time do not provide as consistent a body of evidence to justify inclusion as a distal driver as does education.

Time and its underlying elements

As the above discussion indicates, income and education correlate with a very large portion of the variation in health outcomes across countries. Yet they also leave a great deal unexplained, especially over time. Hence the GBD project has used time as a third variable in its distal-driver formulation. Although time is understood to be most significantly a proxy for technological advance (in close interaction with human biology and biological potential), it also potentially captures other changes related to human action. For instance, sanitation practices (including the

use of soap) involve cultural change even more than they do technology. We will return to this broader context later in the chapter, especially in discussion of super-distal drivers, but first we consider proximate risk factors and the manner in which they mediate between distal drivers and health outcomes.

More Proximate Determinants of Health

The previous section focused on the role of distal drivers—GDP per capita, education, and a time trend (thought to reflect technological change especially)—in determining health outcomes. In and of themselves, however, these factors do not cause health outcomes. Rather, as illustrated generally in Figure 2.4 and specifically in the example for diarrheal disease of Box 2.2, they do so through their effect on proximate risk factors relating to individual behavior and social and environmental conditions. There are very strong relationships between the distal and proximate drivers and health outcomes—so strong that the use of distal drivers alone does offer considerable forecasting power.

Yet simply using broad determinants to forecast health outcomes presents a series of problems. First, any assumption that the relationships across identified distal drivers, proximate risk factors, and ultimate health outcomes will remain stable is almost certainly wrong. Second, by obscuring the more direct or proximate relationships, reliance only on broad determinants offers limited potential for exploring the effects of specific health-related interventions.

It is not difficult to identify intervention points with respect to proximate drivers that suggest at least some degree of potential disconnect between the proximate risks and the distal causes. Clearly the correlations between distal and proximate factors are not perfect. Box 2.2 uses the example of diarrheal diseases to illustrate that although the distal drivers have much impact on the proximate risks, there are also ways in which incremental human effort (as organized, for instance, by national health systems or global initiatives, and targeting nutritional and environmental factors) can modify that relationship.

Over the past 200 years of progress in extending life expectancy, the “best-practice” societies have always been characterized not merely by advanced wealth but also

Box 2.2 Undernutrition and diarrheal disease in developing countries: An example of the interrelationship of distal and proximate drivers of health

Diarrheal infections offer a good example of interrelationships between distal and proximate drivers of health. Diarrheal disease accounts for about one-fifth of child mortality in the world's poorest countries, ranking among the top two or three causes of death in almost all such countries (Black, Morris, and Bryce 2003), yet deaths from diarrheal disease rarely occur in wealthy or middle-income countries. Thus, we can say that the distal driver of income is a major determinant of diarrheal disease, though the pathways linking rising income to reduced diarrheal disease mortality may not be readily apparent. Certainly many diarrheal diseases have treatments that would be easily procured and affordable in rich countries and prohibitively expensive in poor countries. Yet most rich countries drastically reduced diarrheal infections prior to the development of antibiotic drugs (Preston 1975), and few treatments exist for viral causes of diarrhea even today.

Many developing countries, including most "high achievers," have drastically reduced diarrheal disease mortality, both through reductions in incidence and, to a lesser extent, through effective treatment. These successes relate to health systems and to societal education and values, yet we cannot understand the impact of these societal inputs without understanding proximate drivers (Bryce et al. 2003; Jones et al. 2003; Keusch et al. 2006). Effective diarrheal disease interventions can be environmental (reducing the use of unsafe water and improving

sanitation), behavioral (improving hygiene and maternal education), and biomedical (diarrheal disease outreach, treatment, and oral rehydration). Most importantly, diarrheal diseases operate in a negative synergy with nutrition; poor nutrition creates greater vulnerability to and severity of infection, and diarrheal infections create nutrient loss and depletion (Mosley and Chen 1984). Most estimates suggest that more than half of all diarrheal disease mortality could be eliminated by ending undernutrition. Similarly, elimination (through vaccination or other means) of non-diarrheal infections such as measles may reduce vulnerability to diarrheal disease.

In summary, diarrheal diseases offer a textbook example of diseases that have a strong relationship to income but are also responsive to multiple proximate-driver-related solutions, such as reduction in undernutrition via provision of food in conjunction with school attendance. Many would argue that as long as global poverty exists, low-cost proximate interventions with respect to diarrheal and other childhood diseases offer a more immediate solution to health improvement than waiting for income to rise, and that such targeted interventions have the potential to produce healthier, better educated, and more productive societies going forward.

by positive environmental and behavioral circumstances and practices. Future mortality gains are likely to become increasingly hard won and possibly even more dependent on a combination of good treatments and good behaviors. Many modern behavioral and nutritional health risks predispose individuals to noncommunicable and some communicable diseases. The consumption of energy-dense foods and increasingly sedentary lifestyles that predispose societies to obesity and other noncommunicable diseases are of particular concern, and environmental stresses may compound these risks. Such concerns are best exemplified by the mortality crisis among men of Russian descent in Russia and former Soviet states, but there are also indications of a rising chronic disease burden and declining life expectancy in many parts of the United States (Ezzati et al. 2008; Rogers et al. 2007).

WHO's ongoing Comparative Risk Assessment (CRA) project provides our starting point for considering the burden of disease associated with proximate risk factors (Ezzati et al. 2004a). The CRA project has used two guiding criteria for including specific risks in its analysis: (1) selecting risks for which sufficient data and scientific understanding exist in order to assess the exposure and health effects associated with the risks; and (2) selecting risks "for which

intervention strategies are available or might be envisioned to modify their impact on disease burden" (Ezzati et al. 2004b: xx). Within this framework, the project has tried to provide conceptual and methodological consistency and comparability across the risk factors. Table 2.1 shows the 28 risk factors that the CRA project covered in its most recent report (WHO 2009a) and identifies the subset of those included in IFs forecasts in this volume.

The CRA project and other studies (Laxminarayan, Chow, and Shahid-Salles 2006; Prüss-Üstün and Corvalán 2006) have now provided guidance for identifying links between selected risk factors and specific health outcomes, making possible their inclusion in forecasts of future health. For example, we know that childhood undernutrition is associated with a range of communicable diseases, and obesity with certain chronic diseases (Gaziano et al. 2006; Narayan et al. 2006). However, a number of factors complicate quantitative analysis of proximate risk factors. First, they vary with respect to the size of their impact on health outcomes, their susceptibility to human intervention, and the degree to which they change independently of the distal drivers. Second, existing risk assessment analyses have not fully taken into account competing risks (the possibility that those

● WHO's Comparative Risk Assessment project has analyzed the relationship of 28 specific risk factors to health outcomes. ●

Table 2.1 Proximate health risk factors included in the World Health Organization's Comparative Risk Assessment project

Health category	Risk factor	Attributable mortality (%)	Attributable DALYs (%)
Childhood and maternal undernutrition	Underweight*	3.8	6.0
	Iron deficiency	0.5	1.3
	Vitamin A deficiency	1.1	1.5
	Zinc deficiency	0.1	1.0
	Suboptimal breast feeding	2.1	2.9
Other nutrition-related risk factors and physical activity	High blood pressure	12.8	3.8
	High cholesterol	4.5	2.0
	High blood glucose	5.8	2.7
	Overweight and obesity*	4.8	2.3
	Low fruit and vegetable consumption	2.8	1.0
Sexual and reproductive health	Physical inactivity	5.5	2.1
	Unsafe sex	4.0	4.6
Addictive substances	Unmet contraceptive need	0.3	0.8
	Tobacco use*	8.7	3.7
	Alcohol use	3.8	4.6
Environmental risks	Illicit drug use	0.4	0.9
	Unsafe water, sanitation, hygiene*	3.2	4.2
	Urban outdoor air pollution*	2.0	0.6
	Indoor smoke from solid fuels*	3.3	2.7
	Lead exposure	0.2	0.6
Occupational risks	Global climate change*	0.2	0.4
	Risk factors for injuries	0.6	0.8
	Carcinogens	0.3	0.1
	Airborne particulates	0.8	0.4
	Ergonomic stressors	0.0	0.1
Other selected risk factors	Noise	0.0	0.3
	Unsafe health-care injections	0.7	0.5
	Child sexual abuse	0.1	0.6

Note: Mortality and DALY values are for 2004. Risk factors marked with an asterisk (*) are included as proximate drivers in the IFs health model.

Source: Attributable mortality and DALY rates calculated by authors using data from WHO 2009a (Annex A, Tables A3 and A4, pages 50 and 52.)

saved from one cause of death will simply die from another) in their estimated relationships (Laxminarayan, Chow, and Shahid-Salles 2006). And finally, data for some of the factors are very limited. For these reasons, and because all modeling is time and other-resource limited, we currently incorporate a subset of the CRA proximate risk factors in our forecasts, as indicated by the asterisks in Table 2.1. Chapters 5 and 6 explore the implications of alternative

assumptions about these, independently from their expected evolution in the face of changes in distal drivers.

Super-Distal Drivers and the Broader Uncertainty Context of Health

Figure 2.4 portrayed a general model of changing health outcomes in which two major sets of contextual factors (human biology as part of the broader natural environment and human activity) shape the distal and proximate drivers, their relationships with each other, and their impact on health. We return to those general contextual factors now. As the above discussion indicated, income and education explain a large portion of the variation in health outcomes across countries. Yet they leave a great deal unexplained, especially over time.

Change over time, which almost certainly is in substantial part a proxy for technological change, was the most significant source of non-income variations in Preston's analysis of cross-national mortality differentials (Preston 1975). Clearly such change is associated with improved health outcomes in richer societies, and it also has an enormous impact on the pace at which poorer nations are able to achieve the health outcomes of richer ones. We can point to a number of key technological advances driving health improvements over the past two centuries, including the germ theory of disease. We must also consider the global role of innovations in disease management and technology transfer. In considering Preston's findings, Wilkinson (2007: 1) offered an interpretation that illustrates the challenges of modeling health:

What we have to explain is why, with the passage of time, the same amount of income buys progressively more health. It is as if the price of health goes down or, as I once put it (Wilkinson 1996), there is a change of gearing between income and health.

While the notion of a change in price or a change in gearing is conceptually helpful, our ability to predict technology as a function of anything other than time is severely limited. In 1945, it would have been difficult to forecast the dramatic global epidemiologic transition

that was about to take place. Looking to the future, it remains unclear whether this change in gearing is a permanent feature of human health, whether we have reached the highest gear possible, or whether we will, in fact, see a downshift to lower levels of health relative to living standards. Hence the GBD has used time as a third variable in its distal formulation (and we have adopted that model). Although time is understood to be most significantly a proxy for technological advance (in close interaction with biological context and potential), it also potentially captures change related to human action that more truly represents super-distal drivers, even including cultural evolution.

Overall, many implications of these super-distal drivers are captured via distal or proximate drivers, yet we are left with a great deal of uncertainty with respect to the overall trend of health improvement and the pattern and level of convergence or divergence we might expect between countries whose health outcomes are currently very different. Technology has a particularly complicated place in our schema, partly a distal driver and partly a function of broader human activity, but always also interacting closely with the constraints imposed by biology. We look to that interaction first, before turning to other elements of human activity.

Technology and biological limits

Human action drives advances in technology. Vaccines and antibiotics are prime examples, but the range is huge and the future progression of technological advance is highly uncertain. The most significant source of uncertainty concerning technology relates to the frontiers of human longevity. As discussed in Box 2.3, the arguably plausible range for life expectancy of the longest-lived societies on earth in the year 2100 ranges anywhere from 87 to 105 years (for low estimates see Carnes and Olshansky 2007; and Olshansky, Carnes, and Brody 2002; see Oeppen and Vaupel 2002 for high estimates).

The most plausible forecasts both account for trends in age-specific mortality and attempt to separate extrinsic causes of death that are more amenable to elimination from intrinsic causes of death that are more directly related to human senescence (Bongaarts 2006).¹¹ Yet age-specific, disease-specific, and even overall mortality

progress may lag during certain periods when they are not a societal investment priority or between technological revolutions, only to once again accelerate. One notable example from recent history involves cardiovascular disease (CVD) mortality rates. Gains in reduction of cardiovascular disease mortality rates stagnated between 1945 and 1968 in the United States as societal investments focused on infectious disease control and the lagged burden of tobacco consumption manifested itself. Between 1968 and 1995, however, cardiovascular disease mortality rates dropped precipitously (Goldman and Cook 1984; Hunink et al. 1997).

That said, a considerable gap between what humans could achieve and what humans actually will achieve will probably continue. The magnitude of mortality reduction necessary to maintain continued linear improvement in life expectancy is staggering. Many of the technology-dependent therapies underlying such improvements would carry considerable financial costs and would confront serious debate surrounding interventions that might violate bioethical or religious standards or could create unanticipated mutation risks. These debates have already begun with respect to stem cell research and treatments and the genetic modification of lower organisms, and they will almost certainly grow louder in the future. Also, there are clearly ethical issues around expenditures directed at the leading edge of longevity advance rather than toward closing large inequalities within and across societies. Thus, economic and ethical constraints might impact the pace of mortality reduction in a putative “best practice” society and in the relative progress of trailing societies, despite technological possibilities.

In addition to issues about technology and health arising from economic and ethical constraints and from uncertainty about the relationship between technology and the limits of human biology, considerable uncertainty prevails in understanding technological prospects in the battle against infectious diseases, particularly in light of lagging progress in a number of countries, the overwhelming HIV/AIDS epidemic, and the emergence or reemergence of a number of drug-resistant infectious diseases.¹²

In Preston’s analysis of changes in health over time, the greatest change in the positive

■ *Human activity affecting health has a broader scope than the proximate and distal drivers represent; we need also to consider “super-distal” drivers.* ■

■ *Biological limits restrict what technology can accomplish, but technology may or may not achieve all that is possible.* ■

■ *Human action drives advances in health-related technology and in the transfer of that technology through social systems.* ■

Box 2.3 Prospects for human longevity: A debate

There is heated debate between so-called longevity pessimists and optimists, each garnering a range of epidemiologic, demographic, and biological support. In general, the core pessimistic argument depends first on biological limits to the human life span, and second on the policy difficulties and high costs of preventive mortality, particularly in societies with high levels of behavioral or environmental risk. By this logic, chronic obstructive and metabolic disorders, such as CVDs and diabetes, are more endogenous to the human body than communicable diseases and thus more difficult or expensive to prevent. The pessimist school draws heavily on the “Hayflick limit”—the notion that reproductive and functional life spans of cells are time limited and that human evolution provides no mechanism for selecting for longer life spans (Hayflick 1996). According to this view, as exogenous causes of disease are being gradually eliminated, the remaining causes of death are more likely to result from not easily reversed genetic defects and processes of senescence. Pessimists point to demographic evidence showing the compression of mortality in advanced societies into increasingly narrow age ranges, such that continued mortality improvement would require substantial mortality reductions at ages that only a handful of humans have ever experienced, as well as the near-total elimination of deaths of those 65 to 85 years of age. They argue further that it would require the continued pushing back of causes of death that seemed inevitable until only recently (CVDs, many cancers) and of causes that have even yet to be imagined.

While the pessimist argument is biologically compelling, a considerable array of demographic evidence points to continued improvements in longevity. Over time, pessimistic projections have repeatedly been surpassed, often only shortly after the publication of the purported limit (Oeppen and Vaupel 2002; Wilmoth 1997). Optimists point to tremendous improvements in the survival of the oldest-old (those 85+), which have of late outpaced advance at younger ages, resulting in a dramatic increase in the population over age 100. Optimists supplement the demographic evidence by pointing to specific technological innovations such as cellular regeneration and replacement, nanotechnologies for treatment delivery, tools for reprogramming the human genotype for greater reliability, and genomic analysis for the better application of existing therapies (Carey 2003; Carey and Judge 2001; Oeppen and Vaupel 2002). Thus far, they argue, there is no evidence of a hard limit to life, little evidence of approaching limits to life, and ever more promising technological opportunities.

gearing between income and health resulted not from pushing outward the frontiers of modern technology but rather from the gradual transfer to poor countries of existing technologies, such as antibiotics, vaccination, and chemical spraying of infectious disease vectors (Preston 1975). Thus, improvements were the result not merely of technology itself, but also of efforts to transfer technology, provide technical support, and fund the delivery of technologies, carried out through a robust post-World War II intergovernmental framework embodied in WHO (Henderson 1998; Mosley 1984; Preston 1975; 1980; 1996). While future progress in the fight against communicable disease will certainly depend on the pace of new technologies, our ability as a planet to address the proximate drivers of disease via changes to the broader social environment will also determine much of the progress.

The social environment

The final report of the WHO Commission on the

Social Determinants of Health emphasized the indelible imprint that the social environment leaves on human health (CSDH 2008). Some of its effects occur at the national level. For instance, we can observe these relationships in the dramatic decline in male life expectancy in post-Soviet Russia, a crisis whose causes extend not merely into individual and collective risk behaviors (e.g., stress and alcohol consumption) but into the role of broader shocks to political continuity, macroeconomic stability, commodity prices, and national identity. We note also the important role of global factors—the global economy, trade, and health actions—in shaping the national context and in directly shaping human health. For example, systems of global aid, trade, travel, and information affect the availability of pharmaceutical treatments, the transmission of disease vectors across borders, and exposure to hazardous substances such as tobacco and air pollution. While understanding their close interactions, we look, in turn, at the domestic and global elements of the social environment.

Domestic health expenditures and other social influences on health

Inasmuch as a wealthy country completely bereft of health services or health spending would have notably poor health standards, a nation’s social and health systems must have an impact on health. Although researchers have found it difficult to pinpoint measurable societal factors (including health services and health spending) that drive health improvements, those countries that achieve better or worse health outcomes than their income and education would lead us to expect have offered a natural starting point in the search for further systemic determinants of health. This literature has tended to focus separately on developed and less-developed countries.

Among developed countries, a great proportion of the deviation from the expected income-health relationship is explained by one country, the United States, which has one of the world’s highest incomes per capita but ranks only 38th in infant mortality, 32nd in female life expectancy, and 18th in male life expectancy (UNPD 2009a). Other “Anglo-Saxon” societies such as Canada and the United Kingdom also perform worse than expected, albeit to a much

lesser extent. Possible explanations for this Anglo-Saxon deviation include inefficient or poorly managed health systems, high rates of poverty, weak welfare states, high levels of inequality, two-party political systems, and higher rates of behaviors that pose health risks and/or levels of individualism (Berkman et al. 2000; Starfield and Shi 2002; Subramanian, Belli, and Kawachi 2002).

Caldwell's study (1986) of health outcomes in less-developed countries offers a similarly dazzling range of potential sources of structural health variation, some of which are policy-relevant, many of which are not. In addition to education, his findings pointed to the association of the following with positive health outcomes: high levels of health spending or physicians per capita; equity of income or public service availability; a reputation for efficient public service provision; socialist or egalitarian government; a history of civilian, democratic rule; and high levels of gender equity and female autonomy. On the other hand, Caldwell found varying relationships between religious beliefs and practices and health outcomes. Each of his findings has received considerable and often controversial study over the past 25 years.¹³

Perhaps the most interesting and provocative proposition suggests that high levels of societal inequality lead to poor health outcomes, both through the greater prevalence of poverty as well as through higher levels of stress, distrust, violence, and problem behavior (Kawachi, Kennedy, and Lochner 1997; Navarro 2004; Navarro and Shi 2001; Wilkinson and Pickett 2006). However, the inequality hypothesis has been the subject of great debate, offering limited statistical support and few well-understood explanatory pathways (Deaton 2002; 2003; Mellor and Milyo 2001). More generally, Nathanson (1996) offered a framework for understanding the recurring and remarkably stable role over time of broader national norms of social organization and justice in determining health outcomes, identifying three important variables: degree of state centralization, the presence or absence of active grassroots organizations, and societal constructions of risk toward individual versus collective outcomes.

More recently, a number of studies have begun to explore the connection between health outcomes and the more readily forecastable

and policy-relevant measure of societal health expenditures, but even this relationship is complicated. We can imagine that, above and beyond the possibility that public health spending would lead to improved health, it would also demonstrate a broader societal commitment to human welfare. Then again, a society with high levels of public health spending might merely be a very sick society (HIV/AIDS has certainly led to increased public health expenditures) or a very inefficient society (as implied by the high levels of health expenditure without noteworthy health outcomes in the United States). Moreover, any analysis that controls for income, the single biggest determinant of health spending, is bound to offer only limited support for the impact of health spending itself.

Despite these complications, some recent studies point to the potential role of health spending and health systems in determining health outcomes, particularly for child health outcomes in poorer countries. Some estimates of the effect of health spending per capita, even after controlling for GDP per capita, have found coefficients in the range of -0.1 to -0.2—meaning that, above and beyond the effect of income, a 1 percent increase in health expenditure will lead to between a 0.1 percent and 0.2 percent decrease in child mortality (Anand and Ravallion 1993; Bidani and Ravallion 1997; Jamison et al. 1996; Nixon and Ulmann 2006; Wagstaff 2002).

Other studies that measured the independent effect of public sector health spending as a percentage of GDP, also after controlling for GDP per capita, tended to find smaller, often insignificant effects (Filmer and Pritchett 1999; Musgrove 1996). Filmer and Pritchett (1999) estimated that a doubling of average health expenditures per capita from 3 percent to 6 percent would reduce the child mortality rate by only 9–13 percent. They also noted that their estimates implied a cost of between \$47,000 and \$100,000 per child death averted, well above the typical cost associated with prevention of deaths due to specific causes in the societies they studied. After controlling for the effects of factors such as income, education, and ethno-linguistic factors, Filmer and Pritchett noted there was little difference in average health expenditures between the 10 best and

■ *The social environment of health includes domestic health expenditures and policies and—increasingly—global health initiatives, including health governance structures.* ■

the 10 worst achieving countries, as well as considerable variation within those groupings. Filmer, Hammer, and Pritchett (2000; 2002) pointed to a number of factors limiting the effectiveness of public health expenditures, including corruption; ineffectiveness of existing or complementary inputs; redundancy of new services to existing services; and the failure to target incremental spending toward the areas of highest impact—particularly toward health problems affecting the poor (who also tend to have the highest rates of preventable illness and death) and to problems that people would not otherwise address on their own (public goods such as vaccination or sanitation). In conclusion, as it stands, the literature suggests that societal, and possibly governmental, action can influence health, but that systematic effects across society have been relatively small. Many of the most important efforts, however, can be captured through the forecasting of specific proximate drivers of health.

Global health initiatives

Aspects of the global social environment pertaining to health increasingly involve global efforts aimed at both transferring technologies directed at specific proximate risk factors and creating global public goods and a global health governance structure. In particular, in addition to the efforts of the World Health Organization, development agencies, private donors, and the World Bank have increased their global health funding and programmatic efforts. To a great extent, each of these constituencies has placed an emphasis on disease-specific interventions targeting proximate drivers. The World Bank, tasked not with directly promoting health but rather with poverty alleviation and development, became increasingly involved with global health policy beginning in the early 1980s. The Bill and Melinda Gates Foundation, the largest private donor, has focused on all-encompassing disease-specific initiatives aimed at conditions such as polio and on development of new low-cost technologies for communicable disease detection and treatment. These foundations are joined by public-private partnerships or *global health initiatives* (GHIs), involving multiple partners working together to solve a single or small range of issues (e.g., HIV/AIDS, malaria, and tuberculosis). Approximately 100 GHIs currently

exist, and the largest of these—the Global Fund to Fight AIDS, Tuberculosis, and Malaria, the Global Alliance for Vaccines and Immunization, the President’s Emergency Plan for AIDS Relief, and the World Bank Multi-Country AIDS Program—contribute substantially to international funding for communicable disease control overall.

Many observers have questioned the long-term sustainability of such efforts to address proximate drivers without targeting health systems and the broader social environment (Garrett 2007), pointing in part to the failures of a previous generation of post-World War II international health actions. Specifically, in countries where those earlier disease-control efforts were successful, communicable disease risks gave way to a new set of life-course health risks (due to noncommunicable diseases and injuries) that required greater expenditures and depended on the development of health systems. In some of the world’s more disadvantaged countries, and in isolated or disenfranchised areas of emerging countries, persistent and emerging communicable disease risks posed a challenge that demanded social and political changes on top of technological interventions. Significantly, in 2007 the World Bank refined its health strategy to move away from specific disease control, focusing instead on strengthening national health systems and partnering with private donors to fund initiatives (Ruger 2007). The WHO CSDH report signaled another shift toward programs targeting the social environment. Yet past failures such as the Alma Ata Declaration (which promised “health for all by the year 2000”) suggest that the challenges of changing the social environment will be many.

The current generation of efforts to alter the social environment is embodied in a growing push, led by WHO, for global health governance, a term that “refers to the formal and informal institutions, norms and processes which govern or directly influence global health policy and outcomes” (Sridhar 2009: 1366). This broad understanding recognizes that health governance can and does occur at multiple levels (from local to global) and might include a variety of mechanisms (both formal and informal, private and public). In fact, many observers question whether national

governments even have the ability to protect and promote adequately the health of their citizens in a rapidly globalizing world (Fidler 2008/2009; Lee et al. 2007). Civil society organizations—including nongovernmental organizations and community organizations that either provide health services directly or lobby for health change—are also playing increasingly important roles in global health agenda setting, monitoring, and enforcement (Doyle and Patel 2008; Sridhar 2009). Combined with the recent march of global economic integration, these emerging global governance efforts¹⁴ constitute a fundamental shift toward a truly global health system. Yet their future impact on health outcomes depends on many things.

The natural environment

One final super-distal driver subject to a number of ongoing global governance efforts is our changing natural environment. Reflecting its ubiquitous nature, Smith and Ezzati (2005: 325) refer to the natural environment as a super-distal risk factor in that it “affects essentially every disease, even if the pathways are not always well understood.” Given the role that it plays in the evolutionary mechanisms of mutation and natural selection, a complete consideration of the environment would include not only the environment to which individuals are exposed, but also the environment to which their ancestors were exposed (Smith, Corvalán, and Kjellström 1999). However, for the purpose of this volume—forecasting changes in health outcomes for the next half-century—the focus of our concern is changes in the natural environment that humans induce. These include factors as disparate as the introduction of local water pollutants and climate change. Similarly, an interaction of human and natural systems is shaping the rapid evolution and drug-resistance of many threatening infectious agents, presenting increased risk of emerging and reemerging infectious disease (Fauci 2001; McMichael, Woodruff, and Hales 2006). We attempt to capture some of the uncertainty relating to natural environmental change in Chapter 6.

Conclusion

The preceding sections have offered a survey of the conceptualization and measurement of health and of the drivers of change in health

patterns over time and across societies. The chapter has given special attention to the distal and proximate drivers of health outcomes and has also drawn attention to the broader contextual determinants, including human biology and human activity expressed through super-distal elements such as technology, the social environment, and the natural environment.

With respect to distal drivers, national income—as a source of health-seeking resources and a proxy for health-seeking behaviors—offers the most reliable predictor of future changes in health. Yet income cannot explain most national variation in health outcomes over time. Studies also identify the effects of education attainment. The GBD project’s representation of distal drivers also includes a time term, often seen to be technology. Yet technological advance is uncertain, dependent on both the biological context for it and the extent of human activity supporting it.

Tremendous uncertainty surrounds the effort to understand alternative global health futures, making attention to only distal and proximate drivers inadequate. Mapping uncertainty is, of course, a critical aspect of forecasting, and this chapter has only begun the process of doing so. Although analysis of past health trends offers some insight into the pace and pattern of technological change, we cannot know whether these trends will continue, and our lack of certainty with respect to physiological constraints on longevity interacts strongly with that uncertainty. Nor can we truly anticipate the extent of continued support for much positive human action beyond the development of technology—for instance, for continued growth in health spending and for large global initiatives to fight communicable diseases.

In addition, there are sources of uncertainty with much more negative overtones. With respect to the biological context, there is, of course, great uncertainty surrounding the continued unfolding of the HIV/AIDS epidemic and the wild-card possibility of other and even more virulent pandemic diseases. With respect to human activity, we are only beginning to map the extent of damage that environmental change, especially global warming, may do generally and to human health specifically. Chapter 8 in particular will return to the issue of uncertainty, but it pervades all of our analysis.

■ *Human-induced changes in the natural environment also act as super-distal drivers of health outcomes.* ■

■ *Tremendous uncertainty surrounds efforts to understand alternative global health futures.* ■

For instance, our discussion of proximate drivers will identify considerable uncertainty around human choices and behavior.

Still another significant source of uncertainty, and one on which this chapter has scarcely touched, relates to the forward linkages from health to other dimensions of human well-being and the feedback loops that those linkages create. The current era of global health action itself stems from a new interest in those relationships. Sen's human capabilities framework (Sen 1998; 1999a; 1999b) outlined such connections and placed health at the foundation of human needs that must be met in order to achieve development and human security. On the negative side, scholars such as de Waal (2002) have pointed out how extremely high mortality due to HIV/AIDS could shift the positive feedback loop between

health and socioeconomic development into reverse, leading to a downward spiral instead of continuing progress. On the positive side, Fogel's theory of technophysio evolution placed human physiology not merely as the basis of increased production but also as the catalyst for subsequent stages of broad economic, technological, and social development (Fogel 1994; Fogel and Costa 1997).

While we wait to elaborate and explore these forward linkages in greatest detail in Chapters 7 and 8, interim chapters will build on the concepts and understanding of change introduced here and will also begin building our maps of that which we understand best about possible global health futures and that which we understand least. We turn first to the tools that we can use for thinking about alternative futures.

- 1 A 2005 WHO study found that mortality registration coverage ranged from 100 percent in the European region to less than 10 percent in the African region. As a result, even all-cause mortality data for some countries are still derived from sample surveys, indirect estimation techniques, and imputations based on "model life tables" for typical populations rather than through registration of deaths (Mahapatra et al. 2007; Setel et al. 2007).
- 2 The ICD is currently in its 10th revision. For information about the ICD and its history, see <http://www.who.int/classifications/icd/en>.
- 3 Some noncommunicable diseases are also driven, at least in part, by communicable diseases, as in the case of human papillomavirus as a principal determinant of cervical cancer.
- 4 For these and other definitions, see the World Health Organization Statistical Information System (WHOSIS) Indicator Compendium, available at http://www.who.int/whosis/indicators/WHOS09_IndicatorCompendium_20090701.pdf.
- 5 While most would argue that the formerly sick are probably not as healthy as those who were always healthy (referring to a cumulative burden of disease), proponents of the selection hypothesis have argued that those who survive disease, particularly severe life-threatening ones, may have proven their relative fitness and may actually be healthier than individuals who never experienced illness.
- 6 For specific disease weights and methodology, see Mathers et al. (2003) and Mathers and Loncar (2005).
- 7 Using the methods developed for determining YLDs and DALYs, WHO also reports healthy life expectancy, defined as the number of years an individual can expect to live in "full health."
- 8 In practice of course, many drivers instead lead, both intentionally and unintentionally, to increased morbidity and mortality. Among the proximate drivers, roads, cigarettes, wars, and failed surgical interventions may have negative impacts. Similarly, improved living standards could, though they have not yet been shown to, lead to a net deterioration of health. In all cases, the impacts ultimately operate through biological pathways.
- 9 A classic example of negative competing risks involves oral rehydration therapies (ORTs), which mitigate the loss of fluids associated with cholera and other watery diarrheal disorders but have no effect on blood dysenteries such as shigellosis. If ORT is the only antidiarrheal intervention in place, many of those not dying from cholera merely go on to die of shigellosis. By contrast, epidemiologists are constantly in search of the positive competing risks (or synergistic beneficial outcomes) that emerge when a single treatment not only eliminates all deaths due to that particular cause but also improves health in a way that reduces deaths due to other causes. Measles vaccinations, for instance, not only effectively eliminated measles deaths but also led to a reduction in long-term measles-related effects (e.g., blindness, micronutrient deficiency, and cognitive impairment), thereby further reducing mortality and morbidity rates.
- 10 Pritchett and Summers also cited a number of other country-level studies reporting statistically significant elasticities linking rising income to declining infant mortality and other health measures. For infant mortality these include: -0.19 from Flegg (1982); -0.161 (after controlling for education, safe water, and physicians per capita) from Hill and King (1992); -0.21 from Subbarao and Raney (1995); and -0.27 when Pampel and Pillai (1986) looked only at developed nations. Comparable estimates have been found for child mortality (Pritchett and Summers 1996; Wagstaff 2002) and maternal mortality (Bokhari, Gai, and Gottret 2007).
- 11 Bongaarts (2006) pursued one of the more cogent efforts to remove extrinsic causes of mortality and deaths related to smoking from a standard demographic forecast and concluded that the best case high-end life expectancy in 2050 would be about 97.
- 12 We should note that the reversal of the pre-1990s trend of mortality reduction is observable only in sub-Saharan Africa for communicable diseases and in Eastern Europe for noncommunicable diseases. Other regions remain on trend, though a number of specific countries affected by long-running conflicts have seen sustained reversals (e.g., Iraq and Afghanistan).
- 13 Governance variables have offered a particularly contentious, and inconclusive, arena for macro health research. A slew of studies have addressed relationships between governance and health outcomes. A number of studies have found small but significant effects of democracy on improved health (Besley and Kudamatsu 2006; Franco, Álvarez-Dardet, and Ruiz 2004; Navia and Zweifel 2003; Przeworski et al. 2000; Shandra et al. 2004; Szreter 1997), but an equal number have found no effects (Ross 2006) or even negative ones (Gauri and Khaleghian 2002; Khaleghian 2004). Most of the confusion surrounds how exactly one defines a democracy. A line of research with more consistent findings relates to the impact of governance efficacy on health (de la Croix and Delavallade 2009; Gupta, Verhoeven, and Tiongson 2002; Kaufmann, Kraay, and Mastruzzi 2004; Rajkumar and Swaroop 2008; Shen and Williamson 1997; 2001). In particular, measures of government accountability, stability, violence, effectiveness, corruption, and legal institutions have been found to have a strong impact on infant mortality (de la Croix and Delavallade 2009; Gupta, Verhoeven, and Tiongson 2002; Shen and Williamson 1997; 2001).
- 14 One example is the International Health Regulations (IHR) enacted in 2007. The IHR is a legally binding instrument that requires the 194 WHO member countries to report certain disease outbreaks and public health events to WHO.