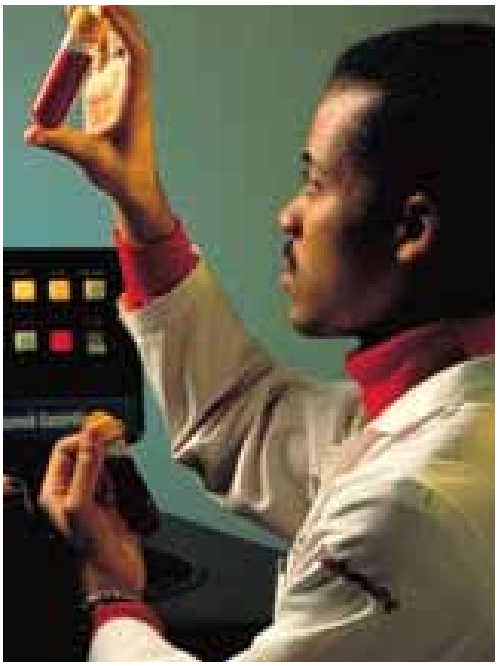
3



Forecasting Global Health

As a bridge between the largely conceptual discussion of Chapter 2 and the forecasting analyses to come, this chapter turns to the more technical topic of how we can best forecast global health futures. We can differentiate tools for looking at the future of global health on several dimensions of coverage and aggregation: whether they are country-specific or multi-country; whether they focus on morbidity, mortality, or both; whether they treat health in the aggregate or consider specific causes of morbidity and mortality; whether they forecast for 10 years or 25 years or even more.

We can also talk about them with respect to their concern with, and treatment of, related human development issue areas: whether they consider demographics, economics, and socio-political characteristics explicitly and dynamically in interaction with health; whether they consider primarily the impact of such other issues on health or also look to the implications of health for other aspects of human development. Further, we can distinguish tools and models in terms of their basic methodological characteristics—whether they focus on very select driving variables or more broadly and structurally portray multiple interacting determinants of human health; whether they tend primarily to be accounting systems that exogenously (externally) provide assumptions about change or whether they more dynamically and endogenously represent households, governments, and other potential agents in interaction.

As always in thinking about the future, the most important dimension on which to understand tools and their use is their purpose. We can identify at least three interrelated purposes of health forecasting systems, very much related to the purposes of this volume:

1. to understand better where patterns of human development appear to be taking us with respect to global health, giving attention to the distribution of disease burden and the patterns of change in it

2. to consider opportunities for intervention and achievement of alternative health futures, enhancing the foundation for decisions and actions that improve health
3. to prepare society for the demographic and other broad (for instance, economic and socio-political) impacts of changing health patterns

This chapter reviews many of the existing forecasting tools and identifies strengths and weaknesses relative to such purposes. It also explains the approach of this volume using the International Futures (IFs) modeling system. We begin with some information about that system because it provides the broader context of our health model and analysis with it.

The central purpose of IFs is to facilitate exploration of global futures through alternative scenarios.

Integrating Health with Broader Human Development: The Larger IFs System

IFs is a large-scale, long-term, integrated global modeling system. It represents demographic, economic, energy, agricultural, socio-political, and environmental subsystems for 183 interacting countries.[1] In support of this series on Patterns of Potential Human Progress, we have added models of education and health. The central purpose of IFs is to facilitate exploration of global futures through alternative scenarios.

The goals that motivated the design of IFs fall generally into three categories: human development, social fairness and security, and environmental sustainability. Across these domains, the project draws inspiration from seminal writers such as Sen (1999a) with his emphasis on freedom and individual development, Rawls (1971) with his emphasis on fairness within society, and Brundtland (World Commission on Environment and Development 1987) with her seminal definition of sustainability. In combination, these emphases provide a philosophical framework for the exploration of human beings as individuals, of human beings with one another, and of human beings with the environment.

IFs is a structure-based and agent-class driven integrated modeling system, producing forecasts for 183 countries through the year 2100.

Fundamentally, IFs is a thinking tool, allowing variable time horizons through 2100 for exploring human leverage in pursuit of key goals in the face of great uncertainty. IFs assists with:

- understanding the state of the world and the future that appears to be unfolding
  - identifying tensions and inconsistencies that suggest political, economic, or other risk in the near and middle term (a "watch list" functionality);
  - exploring longer-term trends and considering where they might be taking us;
  - working through the complex dynamics of global systems.
- thinking about the future we want to see
  - clarifying goals and priorities;
  - developing alternative scenarios ("if-then statements") about the future;
  - investigating the leverage we may have in shaping the future.

Human systems fundamentally involve agents (economists often represent them as individuals in households or firms; political scientists add governments) interacting with one another in various structures (economists focus on markets; political scientists look to action-reaction systems and international regimes; sociologists add societies and demographic structures; anthropologists focus on cultures; physical scientists extend the reach to ecosystems). In general, scientists seek to understand the co-creation and evolution of agent behavior and structural characteristics.

IFs attempts to capture some of that richness. It is a structure-based (with extensive representation of underlying accounting systems such as demographic structures and the exchanges of goods, services, and finance), agent-class driven (so as to provide a basis for representing change), dynamic modeling system. That is, IFs represents typical behavior patterns of major agent-classes (households, governments, firms) interacting in a variety of global structures (demographic, economic, social, and environmental). The system draws on standard approaches to modeling specific issue areas whenever possible, extending those as necessary and integrating them across issue areas. For instance, the demographic model uses the typical "cohort-component" representation, tracking country-specific populations over time by age and sex (extended by education). Within that structural or accounting framework, the model represents the fertility decisions of households

(influenced by income and education) as well as mortality and migration patterns. Similarly with respect to health, we have attempted to build on existing approaches to its forecasting—particularly those of the World Health Organization's Global Burden of Disease (GBD) project—extending those as possible and integrating them with the larger IFs system.

As well as being rooted in the theory of various disciplines and subspecializations, IFs is heavily data based. Data come from the various member organizations of the United Nations family and many other sources. The database underlying IFs, and integrated with the system for use by others, includes data for 183 countries over as much of the period since 1960 as possible. The model system itself runs in annual time-steps from its initial year (currently 2005).[2] The menu-driven interface of the IFs software system allows the display of historical data since 1960 in combination with results from a base case and from alternative scenarios over time-horizons from 2005 through 2100, facilitating user-interventions flexibly across time, issue area, and geography. It provides tables, standard graphical formats, and a basic Geographic Information System or mapping capability. It also provides specialized display formats, such as age-sex and age-sex-education cohort structures and social accounting matrices.

Figure 3.1 shows the major conceptual blocks of the IFs system. The elements of the technology block are, in fact, scattered throughout the model. The named linkages between blocks and the linkages themselves are a small illustrative subset, not an exhaustive listing.

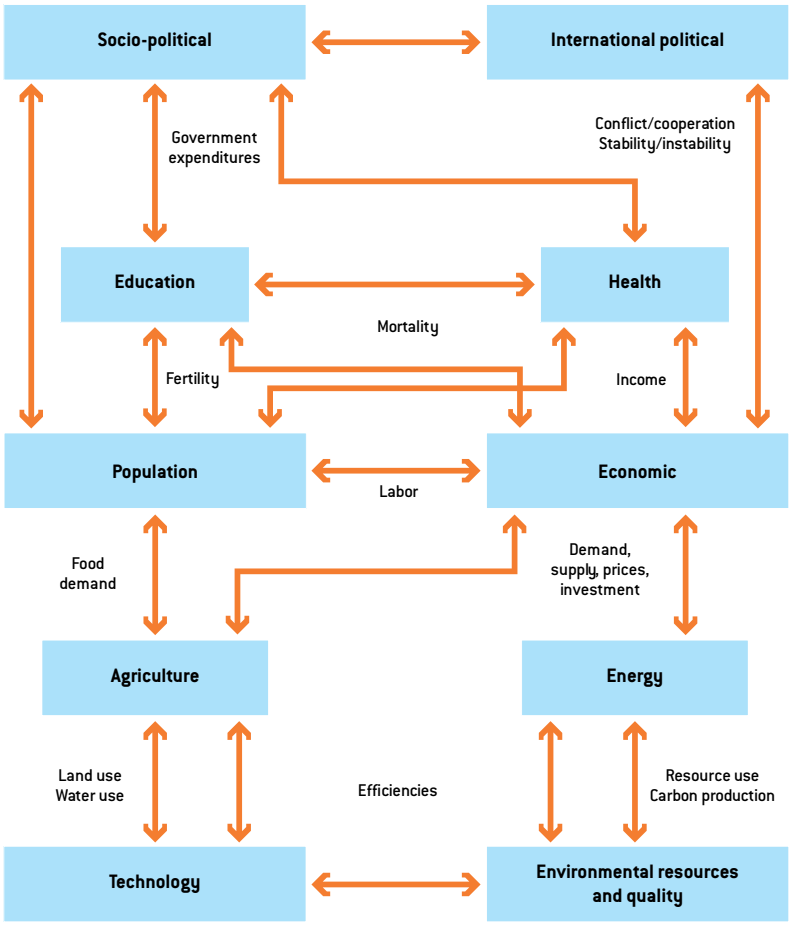
The two models within the IFs system that interact most closely with the health model are the population and economic models. Some of the key characteristics of the population model are that it:

- represents 22 age-sex cohorts to age 100+ in a standard cohort-component structure (but computationally spreads the 5-year cohorts initially to 1-year cohorts and calculates change in 1-year time-steps);
- calculates change in cohort-specific fertility of households in response to income, income distribution, infant mortality (from the health model), education levels, and contraception use;
- uses mortality calculations from the health model;
- separately represents the evolution of HIV infection rates and deaths from AIDS;
- computes average life expectancy at birth, literacy rate, and overall measures of human development;
- represents migration, which ties to flows of remittances.

Some of the most important characteristics of the economic model are that it:

- represents the economy in six sectors: agriculture, materials, energy, industry, services, and information/communications technology;
- computes and uses input-output matrices that change dynamically with development level;

Figure 3.1 Major models in the IFs modeling system and example connections



- is a general equilibrium-seeking model that does not assume exact equilibrium will exist in any given year; rather it uses inventories as buffer stocks and to provide price signals so that the model chases equilibrium over time;
- contains a Cobb-Douglas production function that (following insights of Solow and Romer) endogenously represents contributions to growth in multifactor productivity from human capital (education and health), social capital and governance, physical and natural capital (infrastructure and energy prices), and knowledge development and diffusion (research and development [R&D] and economic integration with the outside world);
- uses a linear expenditure system (LES) to represent changing consumption patterns;
- utilizes a "pooled" rather than a bilateral trade approach for international trade;
- has been imbedded in a social accounting matrix (SAM) envelope that ties economic production and consumption to representation of intra-actor financial flows.

The socio-political model also interacts with the health model as well as with the economic, demographic, and education models. Some of its relevant features are that it:

- represents fiscal policy through taxing and spending decisions;
- shows six categories of government spending: military, health, education, R&D, foreign aid, and a residual category;
- represents changes in social conditions of individuals (such as fertility rates, literacy levels, and poverty), attitudes of individuals (such as the level of materialism/post-materialism of a society from the World Values Survey), and the social organization of people (such as the status of women);
- represents the evolution of democracy;
- represents the prospects for state instability or failure.

The environmental model of IFs, important in many ways for our health analysis, is not as developed as that of many integrated assessment models, but among its capabilities it:

- forecasts exposure to indoor air pollution from the use of solid fuels for heating and cooking;
- computes outdoor particulate concentrations for urban areas;
- forecasts atmospheric accumulations of carbon dioxide from fossil fuel use and deforestation and replicates findings from more extensive general circulation models to compute associated changes in temperature and precipitation, which in turn affect crop yields.

Although initially developed as an educational tool, IFs increasingly supports research and policy analysis. It was a core component of a project exploring the New Economy sponsored by the European Commission (EC) in the TERRA project and a subsequent EC project on information and communication technology and sustainability. Forecasts from IFs supported Project 2020 (*Mapping the Global Future*) of the National Intelligence Council (USNIC 2004) and *Global Trends 2025* (USNIC 2008). IFs also provided driver forecasts and some integrating analysis for the *Global Environment Outlook-4* of the United Nations Environment Programme (UNEP 2007).

The system facilitates scenario development and policy analysis via a "scenario-tree" that simplifies changes in framing assumptions and agent-class interventions. Users can save scenarios for development and refinement over time. Standard framing scenarios (such as those from the United Nations Environment Programme's *Global Environmental Outlook-4*), are available with the model for users to explore and potentially develop further.

IFs is freely available to all users on-line at www.ifs.du.edu and in a somewhat richer downloadable version at the same address. The model's help system contains primary documentation, and the website provides extended reports and publications.

Before turning to the modeling of global health futures within IFs, we first review the foundations provided by other models, including aggregate mortality models and structural health models. We then discuss at some length the hybrid approach we have developed.

All-Cause Mortality Models

In a very real sense, health forecasting began as part of population forecasting, as the size and age structure of a population depend

on the relationship between fertility and mortality. As part of population or demographic forecasting, the emphasis has generally been on mortality as an event (e.g., mortality from all causes) rather than on specific causes of death, but there have been exceptions, such as special attention to HIV/AIDS.

The standard approach to forecasting future population size is the *cohort-component method*, which traces the movement of each *population cohort* through the life span of its members, subtracting out deaths at each age and adding new births to the bottom of the cohort structure. The key drivers of population change, beyond the simple mechanical process of aging (and setting aside the dynamics of migration), are age-specific mortality and fertility rates. Age-specific mortality rates determine life expectancy at any given year, including the typical specification of life expectancy at birth. Models that use cohort-component methods can also begin their analysis with specifications of change in life expectancy and then work backward to modify the age-specific mortality patterns according to mortality schedules standardized for typical populations (Coale and Demeny 1983). In either case, these methods tend to be primarily extrapolative (Bongaarts 2005), although expert judgment may also shape them substantially.

The United Nations Population Division (UNPD) produces the most widely used country-level population forecasts. With the exception of extended attention to HIV/AIDS, its forecasts do not deal with causes of death but instead focus on life expectancy as an aggregate measure. UNPD has summarized its approach as follows:

> Mortality is projected on the basis of models of change of life expectancy produced by the United Nations Population Division. These models produce smaller gains the higher the life expectancy already reached. The selection of a model for each country is based on recent trends in life expectancy by sex. For countries highly affected by the HIV/AIDS epidemic, the model incorporating a slow pace of mortality decline has generally been used so as to reflect a slowdown in the reduction of mortality risks not related to HIV/AIDS. (UNPD 2009a: 24)

The United States Census Bureau also produces basically extrapolative global population forecasts by country. Its methodology for forecasting changes in life expectancy at birth involves the fitting of a logistic or S-shaped curve to the most recent estimate for life expectancy; analysts fit mortality by age to the forecast through interpolation between past rates and rates representing especially low mortality.[3]

The International Institute of Applied Systems Analysis (IIASA) uses a somewhat different but still fundamentally aggregate method for forecasting mortality within a cohort-component model. IIASA describes its method as "expert argument-based probabilistic forecasting," that is, the use of Delphi-like processes[4] across multiple sources and expert surveys to map ranges of likely fertility and mortality (Lutz, Sanderson, and Scherbov 2004: 20). The efforts of the IIASA World Population Program are of special interest because of their purpose, namely, the linking of demographics to broader aspects of human development, such as education and health, and to policy-relevant aspects of demographics, such as the speed of population aging (Lutz, Sanderson, and Scherbov 2008; Lutz and Scherbov 2008), and population impacts on environmental sustainability.[5]

While these aggregate mortality models focus on population forecasting rather than on health forecasting, they alert us to some of the important characteristics that policy analysts and scientists increasingly want to see. For instance, the emphasis of the UNPD on HIV/AIDS as a critical uncertainty in population forecasting draws attention to the desirability of differentiating mortality by cause, especially when death rates from one or more specific causes may be rising and therefore behaving contrary to larger background patterns. And IIASA's emphasis on linking the analysis of population change to other human systems draws attention to both backward and forward linkages in the analysis of population and health.

Health forecasting began with mostly extrapolative forecasts of mortality and life expectancy as necessary components of population forecasting.

Forecasting focused specifically on health emerged during the 1990s through the combined efforts of the World Health Organization and the World Bank.

WHO's Global Burden of Disease project pioneered a structural approach to understanding and forecasting health.

The Emergence and Development of Structural Models

Samuel Preston, in his foreword to a major study published in 2006 (Lopez et al. 2006a: xv), noted that "before 1990, the global disease landscape was perceived 'through a glass darkly.'" Analysts had data on cause of death with relative accuracy for only a small number of countries, and "nowhere were estimates of disease incidence, prevalence, survival, and disabling sequelae consistently combined into population-level profiles of morbidity and mortality."

Circumstances began to change in the early part of the 1990s through the combined efforts of the World Health Organization (WHO) and the World Bank. At that time, WHO, through its Global Burden of Disease project, was building an emerging global database of health statistics, and the first major study of global health, *Disease Control Priorities in Developing Countries* (Jamison et al.), was published in 1993. The Disease Control Priorities project was sponsored by the World Bank, and served as the backdrop for the World Development Report of that year, *Investing in Health* (World Bank 1993). The World Bank reports were geared toward identifying priorities for interventions to achieve rapid health improvements in developing countries with constrained public resources, and they used the emerging WHO database in their analyses of the disease burden of developing countries and targeted interventions. Meanwhile, WHO was developing protocols for estimating and projecting disease-specific mortality and morbidity, and produced *The Global Burden of Disease* (Murray and Lopez 1996b) in 1996. This truly landmark study included 1990 data and provided global projections of mortality and morbidity for over 100 specific diseases through 2020 using new techniques, as discussed in the next section.

A stream of ongoing studies and reports from both WHO's GBD project and the World Bank's Disease Control Priorities project have appeared since those first reports,[6] and a major new GBD study updated with 2005 data is due for release late in 2010. As a result of these projects, a foundation for a structural approach to understanding current global health conditions and thinking about their dynamics in coming years has been established, consolidated, and extended.

Global burden of disease

The GBD project broke new ground not only by focusing specifically on global health, but also through its methodology and approach. First, rather than relying heavily on extrapolative techniques, it identified and used independent variables (income, education, and time) to understand and anticipate health outcomes and changes in them. Second, it disaggregated total mortality into multiple causes of death, important because the driver-outcome relationships vary with cause of death as well as with age and sex. Together these changes made possible a shift to a structural approach to understanding and forecasting health.

In the first GBD report (Murray and Lopez 1996b), the GBD researchers took a major step by building on data for 1990 to forecast the burden of disease in 2000, 2010, and 2020. As we discussed in Chapter 2, they also developed a measure of years lived with disability (YLD) and added it to years of life lost (YLL) to early mortality to create an aggregate measure of disability-adjusted life years (DALYs). Because Murray and Lopez used structural models of disease driven primarily by income and education, they were also able to develop three alternative scenarios of the future mortality and morbidity for over 100 diseases based on differing income and education assumptions for the eight global regions of their analysis.

Mathers and Loncar (2006) built on that path-breaking work in several ways. In addition to drawing on newer and far more extensive disease data and estimates from 2002, they updated driver-variable forecasts; separated diabetes from other noncommunicable diseases (reflecting expectations of increasing overweight and obesity); created regression models specifically for low- and lower-middle-income countries; and developed separate projection models for HIV/AIDS, tuberculosis, lung cancer, and chronic respiratory diseases. They also undertook analysis at the country level rather than at the regional level (although aggregating back to the regional level for presentation of results), and they extended the forecast horizon to 2030. A subsequent update with 2004 data and estimates was published in 2008 (WHO 2008a).

The major-cause typology of disease in the GBD approach builds from three broad cause-groups (Groups I, II, and III; see again Chapter 2) and major clusters within them. Beginning with the 2002 update (Mathers and Loncar 2006), all communicable diseases and maternal, perinatal, and nutritional conditions, with the exception of HIV/AIDS, constitute one cluster. Within Group II (noncommunicable diseases and conditions) the clusters are malignant neoplasms (excluding lung cancer), type 2 diabetes, cardiovascular diseases, digestive disorders, chronic respiratory conditions, and other noncommunicable diseases. And within Group III (injuries) the clusters are road traffic accidents, other unintentional injuries, and intentional injuries. In all, in the GBD projections accompanying the 2002 update, Mathers and Loncar (2006) developed models to forecast mortality and morbidity for 10 major-cause clusters and 132 specific causes within them, including HIV/AIDS. The same clusters and specific diseases were included in the 2004 GBD update (WHO 2008a).

As we discussed in Chapter 2, the conceptual foundation for GBD forecasting has been the use of broad distal drivers rather than directly causal independent variables; those drivers explain very high proportions of the variation in health outcomes. The specific distal drivers used for forecasting were GDP per capita (at purchasing power parity); years of education attainment of adults (extrapolated from the database of Barro and Lee 1996); and a time coefficient that in large part captures technological improvement. The GBD modelers also developed a measure of smoking impact. The GBD project's use of smoking impact in a selected subset of disease formulations reflected the delayed impact of smoking on the incidence of smoking-related diseases, as well as population-specific smoking patterns that the GBD researchers found were not well forecast by distal-driver formulations alone.

The GBD approach has enabled very significant progress with respect to the first major purpose for health forecasting, namely, the desire to understand better possible future changes in health. However, because the driving variables (with the exception of smoking) are not directly causal and therefore do not constitute points of immediately accessible leverage or intervention, the approach does not as directly as we might desire support the second purpose—providing a basis for understanding leverage and informing decision and action. To move in that direction, we now turn to discussion of more proximate drivers of change in the disease burden.

Comparative risk assessment and forecasting

Supplementing the work of the GBD forecasters and moving closer to the level of human choice and action, WHO's Comparative Risk Assessment (CRA) project has identified major disease risk factors and analyzed the burden of disease observed in a population with a given distribution of those risk factors, relative to that in a population with an alternative and theoretically minimal distribution of the risk factors, in order to quantify the impact of risk factors on diseases (Ezzati et al. 2004a; WHO 2009a). The project has identified 28 risk factors (see Table 2.1) grouped in seven categories: childhood and maternal undernutrition; other nutrition-related risk factors and physical activity; sexual and reproductive health; addictive substantives; environmental risks; occupational risks; and other selected risks.

Although the CRA project has not done so, theoretically, one could use the analysis that connects these risk factors to disease burden to forecast change in that disease burden. It would, of course require the development and use of models that represented risk factors. And the effort would struggle with the complex interactive effects of the risk factors (their effects are not simply additive) and with missing risks (it would never be possible to represent all of them). We return to these issues later in this chapter in the discussion of the IFs forecasting approach.

To date, the GBD project has not incorporated comparative risk assessments into its forecasting formulations except in the cases of (1) smoking impact on noncommunicable respiratory diseases, and (2) body mass index (BMI) on diabetes. Instead, efforts to explore choices and interventions tied to proximate risks have focused to a greater degree on detailed analysis of specific diseases and associated intervention options (including some attention to the role of larger health systems) without moving to the

The GBD approach has used broad distal drivers to analyze and forecast health outcomes for 132 specific diseases and conditions.

WHO's Comparative Risk Assessment project has identified 28 major disease risk factors and analyzed their relationship to disease burdens.

level of forecasts.[7] In short, there remains a gap between the GBD project's health forecasting approach and the attention of those interested in analyzing proximate action options. The gap exists with respect to the level of aggregation of disease types in the forecasting and, to an even greater extent, with respect to the drivers used in the forecasting formulations. Specialized models of specific diseases are now being developed and are partially closing that gap.

Hybrid health models could combine formulations based on distal drivers and risk analysis with more richly structural approaches.

Specialized disease-cause models and systems dynamics approaches

Even in the aggregate forecasting of the UNPD and the distal-driver-based work of the GBD, the projects relied in some instances on more specialized treatment of specific diseases and health risks, such as the separate modeling of HIV/AIDS in the otherwise aggregate mortality analysis of the UNPD and of smoking impact in the structural analysis of the GBD project.

The Spectrum system of the Joint United Nations Programme on HIV/AIDS (UNAIDS) is an important example of a specialized model.[8] It differentiates the prevalence of HIV (the stock of those afflicted) from the incidence of new infections, mirroring the common distinction between stocks and flows in the structural analysis of systems. It further represents the transition or flow rates from the prevalence of HIV to the manifestation of AIDS, as well as the rate of deaths of those with AIDS (in part as a function of the availability of antiretroviral therapy).

The GISMO model of the Netherlands Environmental Assessment Agency has begun to build a hybrid model that links environmental risks and health outcomes.

Although not explicitly using the terminology and computer software associated with systems dynamics, the UNPD, GBD, and UNAIDS modeling and forecasting of HIV/AIDS implicitly draws on that approach. Other efforts to examine specific diseases have drawn more explicitly on systems dynamics. For instance, Homer et al. (2004) described a diabetes model developed under the auspices of the United States Centers for Disease Control (CDC). Beginning with a generic model of chronic disease (with separate stocks representing the general population, the vulnerable population, those afflicted without complications, and those afflicted with complications), they proceeded in sessions with CDC staff to develop a specific model for diabetes. The model developers obtained data and parameters from a variety of sources and developed a base case that simulated well the historical growth of diagnosed adult diabetes in the United States after 1980. They also presented the model with the appropriate caveats that apply to systems dynamics models, including the difficulty of specifying the full initial condition and parameter sets and the resulting caution required in interpreting specific numerical output as opposed to more general system behavior.

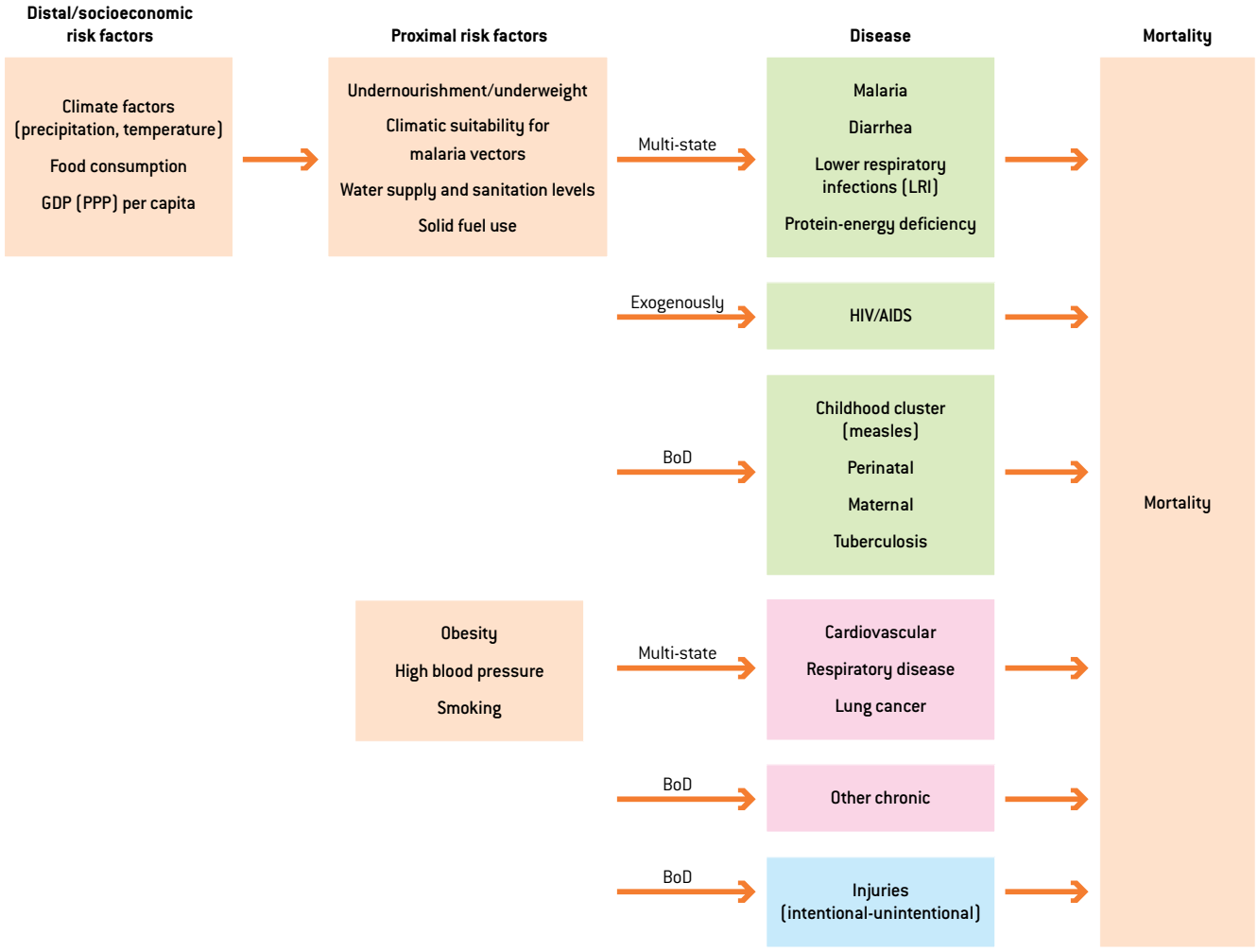
Regardless of the caveats and the difficulties that would face any attempt to generalize this diabetes model to countries around the world, it illustrates the potential for a deeper and, in a significant sense, more truly structural approach to health forecasting than that of the distal-driver models.[9] Such an approach can conceivably serve the purpose (number 2 in our earlier list) of those interested in choices and action, rather than simply forecasting patterns of change (our purpose 1). Such systems dynamics–like approaches also have the clear advantage of explicitly representing morbidity (the stocks of disease prevalence) as a stage of disease progression rather than an aggregate correlate of mortality.

One can begin to imagine a hybrid system of modules to forecast and study health, some of which might be relatively simple distal-driver formulations, some of which may tap knowledge about specific risk factors, and some of which may be more deeply and richly structural with respect to the progression of a disease. Such a system could help bridge the gap between the desire to fully understand patterns of morbidity and mortality and the desire to move the forecasting enterprise toward the goals of aiding choice and action. The development of such a system is a theme that this chapter will continue to develop.

GISMO: Integrating structural and dynamic representations

The Global Integrated Sustainability Model (GISMO) of the Netherlands Environmental Assessment Agency (Hilderink and Lucas 2008) is an emerging model with a number of the characteristics that allow bridging the forecasting of changes in human health with more detailed exploration of the determinants of such changes. The GISMO modeling system forecasts distal forces (such as GDP per capita)

Figure 3.2 Risk factors and health outcomes in the GISMO integrated model



Source: Hilderink, Lucas, Hoekstra, and Niessen. "The Future of Global Environmental Health: An Integrated Modelling Approach" (forthcoming); reproduced with permission of authors.

and uses those to drive change in a number of risk factors that then link to specific causes of death. Figure 3.2 illustrates the process.

Because of the environmental focus of GISMO's home institution, many of its pathways tend to emphasize driving variables, risk factors, and specific diseases related to the environment. Those risk factors and mortality outcomes are modeled using a multi-state approach, distinguishing proximal and distal determinants. Other health outcomes are modeled in GISMO using the GBD project's methodology (shown as BoD in Figure 3.2).

Although Figure 3.2 does not show it, the linkages in GISMO flow not just from driving modules to health outcomes but also from health outcomes back to other modules, notably the demographic one. This embedding of a health module in a broader system begins to help the system also serve the third purpose of health modeling identified earlier, namely, the exploration of how the future of health may affect broader demographic, economic, and even socio-political systems.

Returning to the general purposes of existing forecasting approaches

The beginning of this chapter identified three general and interrelated purposes of health modeling and forecasting:

- understanding better where patterns of human development appear to be taking us with respect to global health
- considering opportunities for intervention and achievement of alternative health futures
- preparing society for the broader (for instance, demographic, economic, and socio-political) impacts of changing health patterns

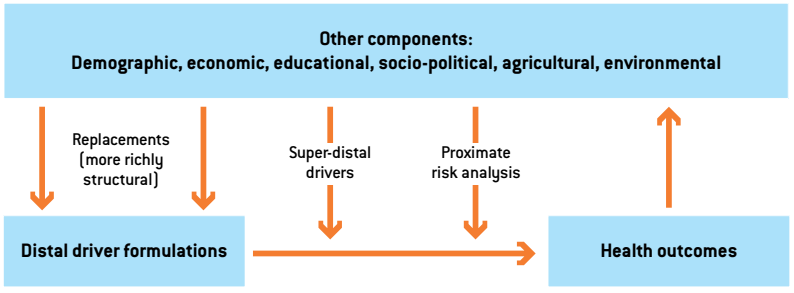
Different forecasting approaches serve different purposes, and hybrid models have the greatest potential to address multiple purposes.

No model is likely to serve all of these purposes well, and we have seen that the forecasting efforts have generally been quite limited in their intent. The GBD project's distal-driver models have opened the door for addressing the first purpose—mapping changing disease burdens. They also provide some foundation for thinking about decisions and actions to shape alternative futures because they quantify, by cause of death, the magnitude of current and (forecasted) future mortality and morbidity.

More specialized models potentially offer more targeted help with the allocation of resources and other interventions both across and within death-cause categories, because they can distinguish different stages of disease with potentially variable associated costs and benefits of intervention. Truly meeting the desires of those who wish to use a model to make cost-benefit decisions about alternative health interventions almost certainly requires a level of detail in representation that is at least at that of the CDC diabetes model referred to earlier. Moving to that level of detail generally means, however, that such modeling sacrifices any attempt to map the complete disease burden, as well as any effort to look at aggregate social implications.

The aggregate mortality models are perhaps currently best suited to helping with the third purpose, namely, the exploration of alternative mortality futures (with age-sex specificity) so as to help society paint, with quite a broad brush, the possible wide implications of different health futures. For example, those who think about financial requirements of pension systems regularly use such models. Analysis of forward linkages could potentially also further enrich the basis for action by providing information about the more indirect costs and benefits of alternative health futures. In reality, the level of aggregation in their treatment both of disease types and of social implications tends to limit analysis to large-scale demographic impacts.

Figure 3.3 Envisioning a hybrid and integrated health forecasting approach



Building a hybrid, integrated system

Although no model can do everything, the association of different approaches with different contributions suggests that a somewhat more hybrid and integrated model form could help with all three forecasting purposes and could provide a richer overall picture of alternative health futures. Figure 3.3 shows the general structure that such a system might take. Formulations based on distal drivers could remain at its core. Again, such a core structure is especially useful in accomplishing the first purpose, because the distal-driver formulations of the GBD offer an existing treatment of health outcomes that is comprehensive with respect to diseases and their related mortality and morbidity.

There is no inherent reason, however, that income, education, and time should be equally capable of helping us forecast disease in each of the major categories (let alone each of the specific diseases) that the GBD models examine. For example, distal-driver formulations tend to produce forecasts of constantly decreasing death rates. Yet we know that smoking, obesity, road traffic accidents, and their related toll on health tend to increase in developing societies among those who first obtain higher levels of income and education, and that only with further societal spread of income and education do smoking and road traffic deaths (and perhaps also obesity) typically decline.[10]

Richer structural models might help us capture such more complex patterns. Many death cause-specific distal formulations would

benefit from modifications and, in some cases, replacement. Deeper and richer structural formulations, like those for specific diseases such as HIV/AIDS or diabetes, are examples. So too, our exploration toward mid-century and beyond of forecasts of deaths from road traffic accidents generated by the distal-driver formulation of the GBD project suggested that a variety of floor and ceiling effects need consideration in the longer term, and that a more richly structural formulation could limit perverse forecasting behavior. (We should, of course, not expect GBD distal formulations that were built for forecasting 25–30 years to be fully capable of use over time-horizons of 50 years or more.)

A hybrid model of the form we wish to see also should contribute to the second purpose of health forecasting, as more specialized structural representations may help us identify opportunities for interventions to improve health futures. These interventions might occur either in the form of super-distal drivers (for example, policy-driven human action with respect to health systems) or the amelioration of proximate risk factors through changes in the behavior of individuals, or in the combination of super-distal drivers and proximate risk factors. As an example, the socio-political and environmental modules in IFs act, in part, as super-distal foundations for variables such as undernutrition and indoor air pollution, which in turn facilitate analyses of proximate risk factors and human action around them.

Finally, with respect to the third purpose (the connections of health with other human systems), representation of these connections would allow health outcomes to feed back to broader human development systems, closing the loop. Many linkages of all of these broader system elements with health should be bi-directional.

Modeling Health in IFs

The IFs health model system is a modularly hybrid and integrated system of the kind that Figure 3.3 sketches. Like any model, it has many limitations; it is an evolving and improving system. In the remainder of this chapter we describe its current form, and in Chapter 4 we will explore the behavior of the system in and of itself and in comparison with the health forecasts of others.

The IFs health model forecasts 15 individual and clustered causes of death and disability. We list them below, grouped by the GBD major cause categories:

Group I—diarrheal diseases; HIV/AIDS; malaria; respiratory infections; other communicable diseases

Group II—cardiovascular diseases; diabetes; digestive disorders; malignant neoplasms; mental health; respiratory conditions and diseases; other noncommunicable diseases

Group III—intentional injuries; road traffic accidents; other unintentional injuries

The GBD (mostly distal driver) model foundation in IFs

The IFs model begins with the driver outcome formulations developed by the GBD project for its analyses.[11] Mathers and Loncar (2006) built a general distal-driver formulation and applied it to selected major disease clusters (elaborating their forecasting to many other diseases with regressions linked to the major clusters). We applied their major-cause formulation to largely the same major clusters and elaborated a small subset of additional detailed cause-of-death categories (namely, malaria, respiratory infections, and mental health) in accordance with our needs for this volume (discussed below).[12] The core Mathers and Loncar formulation that we implemented is:

$$\ln(M_{a,k,i,r}) = C_{a,k,i} + \beta_1 * \ln(Y_r) + \beta_2 * \ln(HC_r) + \beta_3 * (\ln(Y_r))^2 + \beta_4 * T + \beta_5 * \ln(SI_{a,k,r})$$

where

M is mortality level in deaths per 100,000 for a given age group a, sex k, cause i, and country or region r; C is a constant; Y is GDP per capita at PPP; HC is total years of completed education for adults 25 and older; T is time (year = 1900); and SI is smoking impact.

Some important differences in our approach relative to that of the GBD required

Richer structural models might help us capture the complex patterns between health drivers, health outcomes, and forward linkages from health.

We are building and using a hybrid model that combines distal formulations, treatment of risk, extended structures, and integration with broader human development.

The IFs health model has at its core the distal-driver outcome formulations developed by the GBD project.

development of algorithms for computation of initial conditions and small multiplicative adjustments to formulations. Specifically, we begin forecasts in the base year of 2005, we maintain five-year age categories up through 100+, and we represent infants as a separate category. In contrast, the initial data we obtained from the GBD project provided country, sex, and cause-specific mortality, but from the year 2004 and in more aggregated categories at the youngest and oldest ages.[13] Moreover, we have our own sources of data for GDP per capita and education attainment level, which we forecast using our own models.

To reconcile our approach with data used by the GBD project and with GBD formulations we:

- computed a set of initial scaling parameters (by country, sex, and cause of death) that assure consistency of total deaths forecast using the GBD formulations and our 2005 values of driving variables with the cause-specific mortality data in the GBD's detailed death file;[14]
- calculated a second set of scaling or normalization parameters (by country, sex, and age category) that force the sum of all deaths to be the same as the UNPD mortality data for each five-year age and sex category for the year 2005. This process also spreads the more highly aggregated 2004 mortality data of the GBD project[15] into five-year age categories.[16] This process assures that we have initial conditions consistent with UNPD mortality data in our base year.[17]

We modified or replaced some GBD distal-driver functions to improve forecasting capability over our longer forecast horizon.

These adjustments mean that, except for total mortality by age and sex from the UNPD, our numbers in the 2005 base year will not match other data precisely, but that the overall pattern of deaths by cause should be quite close to the GBD data.[18] In the forecasts themselves, we keep the multiplicative scaling and normalization parameters constant over time because there is no clear reason for changing them.

Lumped within the major cause categories forecast with the formulation above are certain diseases we wish to deal with explicitly. These include three diseases in the category "communicable diseases other than HIV/AIDS"—diarrheal diseases, malaria, and respiratory infections—and two diseases under "other noncommunicable diseases"—chronic respiratory conditions (discussed later in conjunction with proximate-driver and relative-risk analysis) and mental health (for which we represent constant death rates). For the first three, the GBD project (Mathers and Loncar 2005 and 2006) provides a distal-driver formulation for detailed causes of mortality that we also use.

The regression equations for the detailed causes take the form:

$$\ln(M_{a,k,i,d,r}) = C_{a,k,i,d,r} + \beta_{a,k,i,d,r} * \ln(M_{a,k,i,r})$$

where

M is mortality rate in deaths per 100,000 for age group a, sex k, general cause i, and country or region r; calculated using the Mathers and Loncar formulation for the major cause category; d is the specific disease.

For the base year, the death rates for the specific diseases are calculated using the above equation.[19] For future years, given the form of the equation, a 1 percent change in the mortality rate for the general cause-group is associated with a ß percent change in the mortality rate for the detailed cause.[20] As an example, the ß for diarrheal diseases for males in the 0–4 age group is 1.493 (Mathers and Loncar 2005: 115). Thus, a 1 percent decline in the mortality rate for communicable diseases other than HIV/AIDS for males in the 0–4 age group in a specific country implies a 1.493 percent decline in the mortality rate for diarrheal diseases for the same group.

In an early phase of model development, we replicated the basic GBD distal-driver models (with the above selected breakouts of detailed causes) and then analyzed forecasts made with our drivers from the integrated IFs system, both in order to compare our results with those of the GBD and also to explore the behavior of the formulations beyond the time-horizon for which they were initially estimated and used. Although our primary focus is 2060 (a 50-year horizon), we pushed the horizon to 2100 in order to understand better the behavior of the equations. The extensions, modifications, and replacements of distal-driver-only functions that we made and describe below resulted from our desire to improve long-term forecasting capability.

Specialized structural model formulations and approaches in IFs

The distal-driver formulation serves well for many disease and death categories. The approach serves less well in other cases, particularly those in which mortality rates tend not to monotonically increase or decrease, often because a complex and perhaps sequential set of factors drive morbidity and mortality patterns. In such cases, still richer structural models can be helpful. One example is smoking, where the GBD approach uses an alternative smoking impact series, but a forecast of smoking rates itself as a driver of impact could be very useful. Another example, which drove the GBD project itself to look for an alternative approach, is HIV/AIDS. We needed either to do as the GBD did and rely on the forecasts of others (such as UNAIDS) or to develop our own approach, which might ultimately allow us to build more scenario "handles" into our own analysis; we chose the latter course. A third example is road traffic accident deaths, where our work with the formulation of the GBD suggested inadequate ceiling effects (upper limits beyond which a forecast should not reasonably go) and long-term forecasts of deaths that appeared unrealistic. A fourth example is health spending as it relates to communicable disease deaths of children. Although subject to significant debate, as Chapter 2 discussed, there is much reason to represent the possibility that health expenditures augment the distal drivers in affecting at least some health outcomes.

Smoking, smoking impact, and chronic diseases

In 1992 Peto et al. proposed a method for calculating the proportion of deaths caused by smoking that was not dependent on statistics on prevalence of tobacco consumption. This method involved developing an indicator for accumulated smoking risk, termed the *smoking impact ratio* (SIR). Ezzati and Lopez (2004: 888) defined the SIR as "population lung cancer mortality in excess of never-smokers, relative to excess lung cancer mortality for a known reference group of smokers." In other words, the ratio is derived by comparing actual population lung cancer mortality with the expected lung cancer mortality in a reference population of nonsmokers. Because the SIR is derived from age-sex lung cancer mortality it can also provide an indication of the "maturity" of the smoking epidemic (the extent to which the population had been exposed to tobacco in the past) (Ezzati and Lopez 2004: 888). Once the SIR has been determined, one can then use it to estimate the proportions of deaths from other diseases attributable to smoking (Peto et al. 1992).

For the GBD project, Mathers and Loncar developed country-level smoking impact (SI) projections to 2030 (Mathers and Loncar 2006; and Mathers and Loncar, Protocol S1 Technical Appendix, n.d.) and used them as part of their distal-driver formulation. The SI projections rely on expert judgment, and it was not possible for the IFs project to improve on them; thus, we used those projections without change. Forecasting beyond 2030 required, however, that the IFs project extend those series, taking into account a long lag between smoking rates and smoking impact. We therefore wanted smoking rates themselves to drive our approach. The development of a structural forecast system for those rates involved several main steps. First, we created a historical series of estimated smoking rates. This was necessary because historical smoking rate data are exceptionally sparse, and we needed to understand the patterns and trajectory of smoking behavior over time. We built the historical imputed smoking series on the most recent smoking rate data point of each country and the smoking impact forecasts of the GBD. Assuming a direct 25-year lag between smoking rate and smoking impact,[21] we used year-to-year percentage changes in the smoking impact series to change smoking rates before and after our smoking data point.[22] In spite of the simplicity of this approach, and the fact that smoking impact reflects more than smoking rates,[23] we found that the constructed series tended to match relatively well when more than one historical point for smoking rate existed.

Second, we constructed cross-sectional relationships that suggest expected rates of smoking based on GDP per capita at PPP for males and females separately:

$$ExpSmoking\_Rate_{Males} = 0.00224 * GDPPCP^2 - 0.3386 * GDPPCP + 38.3996$$

Our representations of smoking impact, HIV/AIDS, road traffic accidents, and health spending differ from or augment those of the GBD.

$$ExpSmoking\_Rate_{Females} = -0.00573 * GDPPCP^2 + 0.6893 * GDPPCP + 5.6634$$

Third, we initialized a moving average rate of change in smoking rate with the compound rate of change between 1995 and 2005. We advanced that moving average over time by slowly changing the moving average toward the expected values of the cross-sectional formulations (weighting the expected value 1/10 of the moving average value each year). We introduced a number of other algorithmic rules to produce what appeared to be reasonable forecasts of smoking rates given the general notion of a bell-shaped curve (or rise and then fall) of smoking with income and time. These included bounding the expected value formulations at $30,000 for females and $50,000 for males so as to avoid complete collapse of smoking rates at high income levels.

Finally, for forecasting we used the same process in reverse that we had earlier used to estimate the smoking series. With the year-to-year percentage change in smoking rate forecasts from 2005 forward, we changed the year-to-year values of the smoking impact series 25 years later.

HIV/AIDS

The ultimate objective of the calculations around HIV infections and AIDS is to forecast annual deaths from AIDS by age category and sex. We began, however, by forecasting country-specific values for the HIV prevalence rate (*HIVRATE*).[24] For the period from 1990–2007 we have reasonably good data and estimates from UNAIDS (2008) on prevalence rates and have used values from 2004 and 2006 to calculate an initial rate of increase (*hivincr*) in the prevalence rate across the population (which for most countries is now negative).[25]

There will be an ultimate peak to the epidemic in all countries, so we need to deal with multiple phases of changing prevalence: continued rise where rates are still growing steadily, slowing rise as rates peak, decline (accelerating) as rates pass the peak, and slowing rates of decline as prevalence approaches zero in the longer term. In general, we need to represent something of a bell-shaped pattern, but one with a long tail because prevalence will persist for the increasingly long lifetimes of those infected and if pockets of transmission linger in selected population subgroups.[26] As a first level of user-control over the pattern, we add scenario specification via an exogenous multiplier on the prevalence rate (*hivm*).

The movement up to the peak involves annual compounding of the initial growth rate in prevalence (*hivincr*), dampened as a country approaches the peak year. Thus, we can further control the growth pattern via specification of peak years (*hivpeakyr*) and prevalence rate in those peak years (*hivpeakr*), with an algorithmic logic that gradually dampens growth rate to the peak year:[27]

$$HIVRATE_r^t = HIVRATE_r^{t-1} * (1 + hivincr_r^t) * \mathbf{hivm_r}$$

where

$hivincr_r^t = F(\mathbf{hivincr_r^{t=1}}, \mathbf{hivpeakyr_r}, \mathbf{hivpeakr_r})$;
t is time (shown in this chapter only when equations reference earlier time points); and r is country (geographic region in IFs terminology). Here and elsewhere, names in bold are exogenously specified parameters.

As countries pass the peak, we posit that advances are being made against the epidemic, both in terms of social policy and technologies of control, at a speed that reduces the total prevalence rate by a certain percentage annually (*hivtadvr*). To do this, we apply to the prevalence rate an accumulation of the advances (or lack of them) in a technology/social control factor (*HIVTECCNTL*). In addition, if decline is already underway in the data for recent years, we add a term based on the initial rate of that decline (*hivincr*), in order to match the historical pattern; that initial rate of decline decays over time and shifts the dominance of the decline rate to the exogenously specified rate (*hivtadvr*). This algorithmic formulation generates the slowly accelerating decline and then slowing decline of a reverse S-shaped pattern with a long tail:

$$HIVRATE_r^t = HIVRATE_r^{t-1} * (1 - HIVTECCNTL_r^t)$$

where

$HIVTECCNTL_r^t = HIVTDCCNTL_r^{t-1} * (1 + \mathbf{hivtadvr} * \mathbf{t}/100) + F(\mathbf{hivincr_r^{t=1}})$

Finally, calculation of country- and region-specific numbers for HIV prevalence is simply a matter of applying the rates to the size of the population (POP) number.

$$HIVCASES_r^t = POP_r^t * HIVRATE_r^t$$

The rate of death of those with HIV would benefit from a complex model in itself, because it varies with the medical technology available, such as antiretroviral therapy (ART) and the age structure of prevalence. We have simplified such complexities because of data constraints, while maintaining basic representation of the various elements. Because both the manifestation of AIDS and deaths from it lag considerably behind the incidence of HIV, we link the death rate of AIDS (*HIVAIDSR*) to a 10-year moving average of the HIV prevalence (*HIVRateMAvg*). We also posit an exogenously specified technological advance factor (*aidsdrtadvr*) that gradually reduces the death rate of infected individuals (or inversely increases their life span), as ART is doing. And we allow the user to apply an exogenous multiplier (*aidsratem*) for further scenario analysis:

$$AIDSDRATE_r^t = HIVRateMAvg_r^t * HIVAIDSR_r^{t=1} * (1 - \boldsymbol{aidsdrtadvr}_r^t / 100) * \boldsymbol{aidsratem}_r^t$$

where

$HIVRateMAvg_r^t = F(HIVRATE_r^t, \text{last 10 years})$

We spread this death rate across sex and age categories. We apply a user-changeable table function to determine the male portion as a function of GDP per capita at PPP, estimating that the male portion rises to 0.9 with higher GDP per capita.[28] To specify the age structure of deaths, we examined data from large numbers of studies on infections by cohort in Brazil and Botswana (in a U.S. Census Bureau database) and extracted a rough cohort pattern from those data.

Road traffic accident deaths

Deciding that the distal-driver formulation alone was producing unrealistic estimates of deaths from road traffic accidents in the long-term, we replaced the distal formulation with a more deeply structural one tied to the growth of the vehicle fleet (occurring pretty much around the world with income growth but saturating at higher income levels) and the declining rate of accidents and deaths per vehicle (which occurs also at higher income levels). Thus, the overall forecast pattern is one of rather rapid growth in road traffic death rates when the vehicle fleet is growing most rapidly, followed by slowing growth of road traffic death rates and ultimately by their decline.

We based our forecast of vehicles per capita (*VEHICLFLPC*) on the formula of Dargay, Gately, and Sommer (2007):

$$VEHICLFLPC = (852 - RF) * e^{(-5.987 * e^{(-0.2 * GDPPCP(R))})}$$

where

GDPPCP is GDP per capita at PPP (thousand dollars) and RF is an adjustment factor (changing over time) to compensate for different land densities, taking the United States as the base. We computed the adjustment factor using the formula from Dargay, Gately, and Sommer:

$$RF = 38.8 * \left(\frac{POP(R)}{LANDAREA(R)} - \frac{POP(USA)}{LANDAREA(USA)} \right)$$

where

RF is the adjustment factor, POP is the population (millions) of country R, and LANDAREA is the total land area (10,000 square kilometers) of country R. We computed the adjustment factor only when country R had higher density than the United States.[29]

Deaths per vehicle tend to fall with income. R. J. Smeed originally proposed a quite widely accepted relationship, now labeled Smeed's Law,[30] which in his notation and without units relates deaths to vehicle ownership:

$$D = 0.0003(np^2)^{\frac{1}{3}}$$

where

D is annual road deaths, n is number of vehicles (which we compute from vehicles per capita above), and p is population. We spread deaths across age categories using information from the GBD project's detailed death tables.

Distal drivers influence health outcomes through their effects on health risk factors (proximate drivers) rather than direct.

Based on concepts from the CRA project, we extended the IFs health model to include selected health risks as proximate drivers of health outcomes.

Public spending on health

The GBD project's distal-driver formulation does not take public spending on health into account. However, we add a term to the basic GBD distal-driver formulation to incorporate the relatively consistent inverse relationship of public spending on health with child mortality rates in poor countries (Anand and Ravallion 1993; Bidani and Ravallion 1997; Nixon and Ulmann 2006; Wagstaff 2002). For countries having a GDP per capita (at PPP) of $15,000 or less, our model applies a simple elasticity for the effects of government health expenditure as a percentage of GDP on all-cause mortality for the 0–4 age group from the distal-driver formulation (the base calculation that health expenditures adjust):

$$\ln({}_5q_0^{adj}) = \ln({}_5q_0^{base}) - 0.06 * HealthExp\%$$

where

${}_5q_0$ is the mortality rate for age 0–4.

In IFs this formalized version becomes

$$MortAdj^t_{j=0\text{–}4,r,k=1} = Mort^t_{j=0\text{–}4,r,k=1} * (1 + HlExpFct^t_r)$$

where

$$HlExpFct^t_r = \textbf{\textit{elhlmortspn}} * (100 * GDS^t_{r,g=health} / GDP^t_r) - GDSHI^{t=1}_r)$$

where

$$GDSHI^{t=1}_r = GDS^{t=1}_{r,g=health} / GDP^{t=1}_r * 100$$

$$\textbf{\textit{elhlmortspn}} = -0.06$$

GDS is government expenditure; GDSHI is initial government expenditure; HlExpFct is health expenditure factor; elhlmortspn is the elasticity of mortality with health spending; j is age category; r is country/region; k is cause (1 is communicable); t is time-step

In this calculation we use health expenditure as a percentage of GDP, rather than health expenditure per capita, to avoid any confounding with the distal driver for GDP per capita. We established this coefficient for all-cause mortality in the 0–4 age category on the basis of multivariate regressions using the GBD distal-driver specifications as a base model and compared the coefficient with the results of existing studies (Anand and Ravallion 1993; Filmer and Pritchett 1999; Wagstaff 2002).[31]

Model extensions to include proximate drivers in IFs

As we have noted previously, the distal drivers do not, in and of themselves, cause health outcomes. Rather, they influence mortality and morbidity through their effects on a host of proximate risk factors. If these factors move in parallel with the distal drivers—that is, if changes in the distal drivers fully capture the risk factors and the efficacy of the health systems—then it would be reasonable to forecast solely on the basis of the distal drivers. To the extent that this is not the case, however, dealing with risk factors more explicitly may improve forecasts. Moreover, the proximate drivers provide some analytical leverage with respect to ways in which we might improve future health outcomes, the second forecasting purpose identified earlier. In summary, forecasting based on proximate drivers and risk factors brings us closer to the level of targeted human interventions.

In this section, we describe a method for modifying forecasts based solely on the distal drivers (and our specialized structural extensions to them) by addressing a number of the risk factors identified in WHO's CRA project discussed earlier (Ezzati et al. 2004a). We have not addressed all risk factors, or all health outcomes related to the selected risk factors, because of limitations of data, knowledge, and time. Still, the procedure we describe does allow us to deal with some of the more important risk factors and provides a foundation on which we and others can build further.

The basic proximate-driver approach in IFs

We build our approach on an understanding of two basic concepts used in the CRA project, specifically *relative risk* (RR) and *population attributable fraction* (PAF).

A relative risk is a "measure of the risk of a certain event happening in one group compared to the risk of the same event happening in another group."[32] We follow the approach taken by the CRA study, comparing our forecast population at risk to an "ideal" population with a "theoretical minimum" level of risk. For example, WHO estimates that children under five who are moderately or severely underweight are almost

nine times more likely to die from communicable causes than is a population of "normal-weight" children (Blössner and de Onis 2005).

As its name suggests, a PAF reflects the degree to which a specific risk factor is associated with the occurrence of a specific health outcome. Formally, it is the proportional reduction in disease or death rates for the total population (including those with and without the risk factor) that we would expect if we reduced a particular risk factor to a theoretically minimum level (Murray et al. 2004). The further the current situation is from the ideal, the closer the value of the PAF will be to 1.

A PAF is calculated as:

$$(\sum RR(x)P(x)-\sum RR(x)P'(x) / \sum RR(x)P(x)) = 1 - \sum RR(x)P'(x) / \sum RR(x)P(x)$$

where

RR(x) is relative risk at exposure level x;
P(x) is the population distribution in terms of exposure level, that is, the shares of the population exposed to each level of exposure;
P'(x) is the theoretical minimum population distribution in terms of exposure level. For certain risks this is defined as no exposure; where this is not realistic, WHO defines an international reference population

Following this definition, multiplying the mortality from a particular disease by the PAF yields an estimate of the number of people who would not have died had the risk factor been at its theoretical minimum level. If we assume that the values of RR(x) and P'(x) for particular risk factors and diseases do not differ across countries or change over time,[33] then changes in the PAF are solely a function of changes in P(x), the exposure of the population to the particular risk factor. Thus, it is necessary to be able to forecast the future levels of the risk factors. Later sections of this chapter describe how this is done for specific risk factors.

Since our forecast of health outcomes from distal drivers implicitly suggests certain proximate-driver levels, we are really interested in the effect of a difference in (1) estimates of the future levels of a risk factor based only on distal drivers, and (2) estimates based on a more complete set of drivers. Again, assuming this is possible, we can calculate two versions of the PAF, namely, PAF_{Full} and PAF_{Distal}. Defining $Mortality_{Distal}$ as the mortality calculated using only the distal drivers and $Mortality_{Final}$ as the mortality after accounting explicitly for the risk factor, we can state that:

- $Mortality_{Distal} * PAF_{Distal}$ represents the number of people who would not have died had the risk factor been at its theoretical minimum level using the distal formulations for mortality and the proximate risk factor; and
- $Mortality_{Final} * PAF_{Full}$ represents the number of people who would not have died had the risk factor been at its theoretical minimum level using a more complete formulation for mortality and the proximate risk factor.

If we assume that no other factors influence the difference in total mortality between the distal formulation and that using the full model, then:

$$Mortality_{Final} - Mortality_{Distal} = Mortality_{Final} * PAF_{Full} - Mortality_{Distal} * PAF_{Distal}$$

Yields:

$$Mortality_{Final} = Mortality_{Distal} * ((1-PAF_{Distal}) / (1 - PAF_{Full})) = Mortality_{Distal} * \sum RR(x)P_{Full}(x) / \sum RR(x)P_{Distal}(x)$$

The adjustment factor is the ratio of the weighted average relative risks based on the distributions using the distal-only versus the full formulations for estimating the value of the risk factor. A higher weighted average RR based on the full formulation implies that the distal drivers overestimate the improvement (or underestimate the deterioration) in the risk factor. Thus, the mortality forecast needs to be adjusted upward. Alternatively, if the weighted average RR is lower based on the full formulation than on the distal formulation, the mortality forecast will be adjusted downward. Note that this property of the calculation actually obviates the need to know the theoretical minimum population.

Mapping proximate drivers to diseases and age categories

We used this approach to modifying distal-driver forecasts by forecasts of proximate risks for

Our treatment of childhood undernutrition illustrates the IFs approach to modeling individual proximate drivers.

eight proximate risk factors (refer back to Table 2.1 for the broader list of factors included in the WHO CRA project and to Chapters 5 and 6 for the eight IFs analyses). Table 3.1 shows the particular diseases and age groupings that each of the risk factors in IFs affects.

An example of the proximate-driver approach in IFs: Undernutrition

We elaborate here the process for specification of the adjustment factor linking a proximate driver (in this case, undernutrition as measured by underweight) and disease (in this case, all communicable diseases other than HIV/AIDS) for children under five. For elaboration of our approach to the other proximate drivers included in this volume, please see the specific sections in Chapters 5 and 6 and the technical documentation of the health model at www.ifs.du.edu.

Fishman et al. (2004) discuss the many risks of death and disease that undernutrition, in the form of being underweight, poses to children under the age of five and to women of reproductive age. They point, in particular, to the potential consequences for children under the age of five from communicable diseases other than HIV/AIDS, one of the general cause-groups included in the GBD project. They break this category down into diarrhea, pneumonia (respiratory infections), malaria, measles, and a combined group of these and all other communicable diseases except HIV/AIDS, providing specific relative risks for each of the four specific disease groups, as well as for the combined group. As noted earlier, we also separate out diarrhea, respiratory infections, and malaria but define our fourth group to include measles as well as other communicable diseases except HIV/AIDS. Thus, we are able to estimate mortality rates, $Mortality_{Distal}$, for each of three separate and one combined cause-groups, as described earlier in this chapter.[34]

Fishman et al. (2004) specified the risk factor for undernutrition in terms of weight-for-age using an "average" population, with a given mean and standard deviation. For any particular country, children under five years of age are assigned to one of four categories: severely underweight (more than three standard deviations [SDs] below the mean weight for the "average" population), moderately underweight (3SDs to 2SDs below the mean weight for the "average" population), mildly underweight (2SDs to 1SD below the mean weight for the "average" population), and normal weight (no more than 1SD below the mean weight for the "average" population). This constitutes the population distribution in terms of exposure level, specified as P(x) in earlier discussion of the proximate-driver approach in IFs. Fishman et al. (2004) also described the theoretical minimum distribution, P'(x). Using their all-cause category as an example (Fishman et al. 2004: 64), children who are severely underweight are 8.72 times as likely to die from communicable diseases as those with a normal weight; those who are moderately underweight are 4.24 times as likely to die; and those who are mildly underweight are 2.06 times as likely to die.

In order to calculate the adjustment factor for the effect of undernutrition on children's mortality from communicable diseases, we need to know the population distributions (P) of undernutrition based on both the distal drivers, $P_{Distal}(x)$, and the full model,

Table 3.1 Risk factors and their disease impacts in IFs

Risk factor	Diseases impacted in IFs	Age group impacted in IFs
Childhood underweight	Diarrheal diseases Respiratory infections Malaria Other communicable diseases	<57
Body mass index	Cardiovascular diseases Diabetes	30+
Smoking	Malignant neoplasms Cardiovascular diseases Respiratory diseases	30+
Unsafe water, sanitation, and hygiene	Diarrheal diseases	All ages
Urban air pollution	Respiratory infections Respiratory diseases Cardiovascular diseases	30+
Indoor air pollution from household use of solid fuels	Respiratory infections Respiratory diseases	<5 (infections) 30+ (diseases)
Global climate change	Diarrheal diseases Respiratory infections Malaria Other communicable diseases	<5
Vehicle ownership and fatality rates	Road traffic accidents	All ages

Note: In IFs, global climate change affects the listed diseases for children under five years of age through its impact on childhood underweight.

Source: IFs project.

$P_{Full}(x)$. Using historical data, we developed formulations for calculating both of these. The latter draws on IFs representation of the food system. Most directly, it is a function of available calories per capita within a country, which reflects dynamics around income levels and food prices that, in turn, respond to land resources and use, crop yields, fish catch and aquaculture, energy prices, and more.[35]

The population distributions and relative risks provide all the information necessary to calculate the adjustment factor and the adjusted mortality, $Mortality_{Full}$, using the equation specified earlier in the section titled "The basic proximate-driver approach in IFs." Since we deal only with a single risk factor in this case, the formulation requires nothing further.

Dealing with multiple risk factors

Sometimes more than one risk factor will be linked to a particular disease. In theory, this requires estimating joint relative risks and exposure distributions. Under certain circumstances, however, a simple method can be used to calculate a combined PAF that involves multiple risk factors (Ezzati et al. 2004a):

$$PAF^{combined} = 1 - \prod(1\text{-}PAF^{i})$$

where

PAF^{i} is the PAF for risk factor i

The logic here is as follows: $1\text{-}PAF^{i}$ represents the proportion of the disease that is not attributable to risk factor i. Multiplying these $1\text{-}PAF^{i}$ terms yields the share of the disease that is not attributable to any of the risk factors, and subtracting this product from 1 leaves the share of the disease that is attributable to the set of risk factors considered.

Say that we have two risk factors:[36]

$$PAF^{combined} = 1 - (1\text{-}PAF^{1})(1\text{-}PAF^{2})$$

Following from the discussion above, the combined adjustment factor can be calculated as:

$$((1\text{-}PAF^{combined}{}_{Distal}) / (1\text{-}PAF^{combined}{}_{Full})) = [(1\text{-}PAF^{1}{}_{Distal})(1\text{-}PAF^{2}{}_{Distal})] / [(1\text{-}PAF^{1}{}_{Full})(1\text{-}PAF^{2}{}_{Full})]$$

$$= [(1\text{-}PAF^{1}{}_{Distal}) / (1\text{-}PAF^{1}{}_{Full})] * [(1\text{-}PAF^{2}{}_{Distal}) / (1\text{-}PAF^{2}{}_{Full})]$$

$$= [\sum RR^{1}(x)P^{1}{}_{Full}(x) / \sum RR^{1}(x)P^{1}{}_{Distal}(x)] * [\sum RR^{2}(x)P^{2}{}_{Full}(x) / \sum RR^{2}(x)P^{2}{}_{Distal}(x)]$$

In other words, the combined adjustment factor is a simple multiplication of the individual adjustment factors.

Other proximate-driver modifications of distal formulations

In limited cases, GBD researchers decided that model behavior necessitated proximate-driver modifications to the distal-driver approach. For example, while distal relationships suggest falling rates of noncommunicable disease over time, the popular assumption that BMI levels will continue to increase over the next decade(s) indicates that diabetes mortality might actually rise in the near to mid-future.[37] Similarly, in the GBD 2002 update, Mathers and Loncar (2006) introduced an adjustment factor related to smoking to re-estimate chronic respiratory–related mortality (a subset of other noncommunicable disease), citing concerns that distal-driver projections alone did not adequately reflect assumptions of decreasing smoking rates in high-income countries (Mathers and Loncar, Protocol S1 Technical Appendix, n.d.: 6). We generally followed the GBD approach to these two disease categories, with slight modifications for endogenizing our BMI forecasts into diabetes as described below.

Diabetes. For a population at a "theoretical minimum" level of BMI (mean 21 and one standard deviation), we assume that diabetes-related mortality will decrease over time at 75 percent of the rate of other noncommunicable disease–related mortality, following the logic employed by the GBD researchers (see again Mathers and Loncar, Protocol S1 Technical Appendix, n.d: 5). For a population with levels of BMI above the theoretical minimum, however, we compute a country-, age-, and sex-specific shift factor (labeled "diabetes relative risk" in the IFs model) that modifies this expected decrease in other noncommunicable disease–related mortality rates in order to

We used the approach of the CRA project to deal with multiple (interacting) risk factors.

determine the expected diabetes-related mortality rate.

For each unit of BMI increase, the relative risk of diabetes-related mortality (compared to a theoretical minimum population) ranges from approximately 1.4 for females and 1.2 for males, depending on age group.[38] Since we do not forecast age-specific BMIs in IFs (due largely to a lack of historical data), we initialize our diabetes shift factor to match those provided by the GBD project. However, in forecasts after 2005 we use our assumptions around future BMI to drive projections.

Respiratory diseases. The two subsets of chronic respiratory conditions—chronic obstructive pulmonary disease (COPD) and other chronic respiratory conditions—are computed separately from other noncommunicable diseases, and both follow the same formulation:

$$Mort = LN(SIR*RR + 1\text{-}SIR) *(Exp(ONCD\_Mort)^{0.75})$$

SIR is a "smoking impact ratio," calculated as smoking impact divided by an adjustment factor that is specific to age, gender, and to some regional differentiation.[39] Relative risks (RR in the above equation) are also specific to gender, age, and type (COPD or other respiratory disease) and were provided by the GBD authors. ONCD_Mort is other noncommunicable disease mortality.

IFs forecasts disability rates from disease-specific relationships between morbidity and mortality.

Disability and DALYs

To represent morbidity, we followed the path of Mathers and Loncar (2005 and 2006) and linked change in years of living with disability over-time to change in years of life lost over-time. In general, the GBD approach posits that disability declines at a rate that is some fraction of decline in mortality rates (from 0 or no decline in disability to 100 percent, or fully comparable decline). As Mathers and Loncar explained:

> YLD projections were generally derived from the YLL projections by applying the ratio of YLD to YLL for 2002. For ischaemic heart disease and stroke, future incidence rates were assumed to decline at 50% of their mortality rate declines reflecting declining case fatality rates as well as incidence rates. For causes where there is little or no mortality, age-sex-specific YLD rates per capita were generally assumed to remain constant into the future. For certain mental disorders, musculoskeletal conditions, and hearing loss, disability weights were assumed to decline somewhat with improvements in income per capita reflecting increasing treatment coverage. YLD rates for nonfatal communicable diseases and nutritional deficiencies were assumed to decline at between 50% and 100% of the mortality rate declines for Group I causes. (2006: 2016)

Table 3.2 shows the relative rates of decline that the IFs project adapted from the GBD discussion. In all cases, the rate of decline in morbidity rate is posited to be equal to or less than the decline in mortality rate, frequently only half as much. One of the strong implications of the approach and of the coefficients that are less than 100 percent is that the forecasts will generate an ongoing shift of total disease burden from mortality to disability.

In contrast to this expeditiously simple approach to forecasting disability in IFs, existing evidence provides a complicated picture of the relationship between declining mortality and morbidity. For chronic diseases such as cardiovascular diseases, reductions in cause-specific mortality result from treatment as well as from prevention, meaning that decreased mortality should be associated with relatively less decline in the incidence of the disease (Mathers and Loncar, Protocol S1 Technical Appendix, n.d.). As a greater proportion of incident cases survive and continue to be affected by the disease, prevalence rates should rise, an expectation confirmed by empirical data (CDC and The Merck Company Foundation 2007; Robine and Michel 2004). In other words, the decline in incidence or prevalence of a disease associated with a mortality decline should be determined by the relative prominence of prevention (reducing both incidence and prevalence) versus treatment (which should not affect incidence and should increase prevalence).

Moreover, the basic logic of the current approach in IFs does not address changes in disease severity as mortality declines. While the survival of those who would otherwise have been most likely to die might increase the average severity of disease among surviving cases, it is also quite possible that the very treatments that reduce mortality would also reduce disease severity across the entire distribution of illness. In fact, most recent evidence points to reductions in morbidity (as measured by self-rated health status and performance on the activities of daily living) that outstrip the pace of mortality reduction, meaning that even as populations grow older, they spend a greater proportion of those extra years in good health (Crimmins 2004; Payne et al. 2007; Robine and Michel 2004). In other words, even as prevalence increases due to greater survivorship, reductions in the average severity of disease may be so great as to reduce the overall burden of morbidity (Crimmins 2004; Mathers et al. 2004). See further discussion on this issue in Chapter 7.

Table 3.2 Percent changes in disability relative to declines in mortality by cause in IFs

	Percent changes in disability with changes in mortality
Communicable diseases	
Diarrheal diseases	75
HIV/AIDS	75
Malaria	100
Respiratory infections	100
Other communicable diseases	75
Noncommunicable diseases	
Cardiovascular diseases	50
Diabetes	100
Digestive disorders	100
Malignant neoplasms	100
Respiratory diseases	100
Other noncommunicable diseases	100
Injuries	
Intentional injuries	75
Road traffic accidents	75
Other unintentional injuries	75

Note: Mortality refers to years of life lost and disability to years of living with disability; IFs also represents mental health but does not model a mortality/morbidity relationship for it.

Source: Except for HIV/AIDs, IFs project estimates for communicable and noncommunicable diseases are based primarily on Table 6 (page 19) of Mathers and Loncar Protocol S1Technical Appendix [n.d.]; the estimates for HIV/AIDs and injuries are IFs project assumptions.

Conclusion

Considering the importance of health to us individually and as societies, modeling and forecasting health outcomes is a remarkably new activity. Movement beyond attention to life expectancy and age-specific mortality in the aggregate to the exploration of future multiple-cause mortality extends back only about two decades. The Global Burden of Disease project broke much important new ground in its analyses of causes of mortality and disability and in its two major sets of projections, each extending about 30 years.

We have been fortunate in being able to build significantly on the GBD project's distal-driver approach in our IFs work. There is, however, reason to believe that the future of forecasting will turn increasingly to a more hybrid, integrated analysis of systems, more regularly supplementing distal-driver analysis with attention to the kind of proximate-driver and elaborated structural representations that better allow modelers to connect forecasting with policy analysis. Moreover, almost inevitably there will be increasing efforts to integrate health modeling with the modeling of demographic and economic systems (minimally), and probably with some representation of environmental, socio-political, and other specialized systems, such as agriculture and energy.

We cannot pretend to feel highly confident in our efforts to construct and use such a hybrid, integrated health model in a broader modeling system. Nonetheless, the foundations do exist upon which to at least tentatively explore the possible futures of global health in larger context. The next chapter lays out a base case forecast of global health as a foundation, before it and subsequent chapters turn to exploration of possible alternative health futures.

1 For introduction to the character and use of the model, see Hughes and Hillebrand 2006.

2 More technically, the model structure is recursive (it computes equations sequentially in each time-step without simultaneous solution). It combines features of systems dynamics (notably the accounting structures with careful attention to both flows and stocks) and econometrics (using estimated equations for the dynamic behavior of the agent-classes).

3 The broader population forecasting methodology is available at www.census.gov/ population/www/ documentation/twps0038.pdf. The method uses fixed-point logistic models.

4 The full Delphi method involves multiple and systematic iterations across a group of experts to map (and generally narrow) disagreement and establish a central tendency (Gordon and Helmer-Hirschberg 1964).

5 Although IIASA historically forecast population by global region, it has moved to country and even intra-country analysis.

6 See, for example, Jamison et al. 1993; Jamison et al. 2006; Lopez et al. 2006a; Mathers and Loncar 2006; Murray and Lopez 1996b; WHO 2008a; 2009a; and World Bank 1993.

7 See the individual chapters in the recent Disease Control Priorities project report for examples of this approach (Jamison et al. 2006).

8 See http://data.unaids.org/pub/Presentation/2009/ 20090414_spectrum_2009_en.pdf and also Stover et al. 2008.

9 Similarly, Homer and Hirsch (2006) developed a systems dynamics model to explore the role of public health systems in prevention and care of chronic disease.

10 It is partly for this reason that the creators of the GBD models added exogenous specification of smoking impact to the otherwise mostly monotonically (one-direction only) changing specifications.

11 We are indebted to Dr. Colin Mathers, who generously shared with us his original database and regression models and provided responses to our many queries about them. We regret and accept full responsibility for any errors we may have made in our use of them.

12 Using the GBD historical data, we re-estimated the formulation for cardiovascular diseases in order to correct for a discrepancy in the direction of the coefficient for female smoking at older ages.

13 IFs represents populations in five-year intervals up through 100+, whereas the oldest age category in the GBD data combined all ages from 85+. In addition, we are able to represent infants separately (as well as within the 0–5 age category), and the GBD project only included them in the 0–5 group.

14 The GBD's detailed death file of mortality rates for 2004 was provided by Dr. Colin Mathers.

15 We began by spreading the same death rates for all five-year age categories within larger categories, but then used smoothing procedures for the initial spread so as to represent better the changing patterns of mortality by cause of death across five-year categories. We normalized the death rates across disease types so as to make the total death rates of countries consistent with UNPD data for each five-year category.

16 As we noted earlier, Mathers and Loncar (2006) did not separately estimate infants (those under one year of age); we used their coefficients for the under-five age category for infants also.

17 We used the UNPD's 2008 revision for initialization.

18 Complicating initialization further, the UNPD presents its data in five-year ranges, including 2000–2005 and 2005–2010. The age- and sex-specific survivor-table values in those ranges therefore do not correspond to specific years like our base of 2005. After correspondence with Kirill Andreev of the UNPD, which we acknowledge appreciatively, we decided to average the mortality values in the two five-year ranges ending and beginning with 2005.

19 Since Mathers and Loncar did not provide coefficients specifically for diarrheal diseases and malaria, we use those provided for the more general category of infectious and parasitic diseases. They provide separate coefficients for gender and seven age categories but not region. We calculate the adjustment factors in the same way as for other diseases, as described later in this chapter.

20 In theory, if ß > 1 (ß < 1) and the mortality rate for the general cause-group increased (decreased) sufficiently, the mortality rate for the detailed cause could exceed the mortality rate for the general cause-group. Furthermore, if ß > 1 and the mortality rate for the general cause-group decreased, the mortality rate for the detailed cause theoretically could fall below zero. As these would represent illogical results, we checked to make sure that these situations did not occur in the actual projections or, if they did, we—as Mathers explained to us the GBD researchers did—adjusted the sum of the specific causes to match the projected general cause-group rate.

21 Mathers and Loncar in their Protocol S1 Technical Appendix (n.d.: 8) said that their approach assumed 25-year time lags between tobacco consumption and smoking impact.

22 The IFs smoking impact forecast is age-cohort specific, while our smoking rate is not; thus, we needed a weighted average growth rate. The weighting used the population-cohort sizes in 2005.

23 There is not a one-to-one relationship, of course, between smoking rate and smoking impact based simply on a lag. Many other factors, including what is smoked and how (including frequency), will affect smoking impact. Treatment might also be an impact, although the GBD time/technology variable in the distal formulation could pick up some of that. We further understand that our historical smoking series is stylized.

24 The IFs approach does not use an incidence-based model, which would be an alternative. Such a model would also allow specification of mother-to-child transmission and of treatment coverage and success.

25 The IFs pre-processor calculates initial rates of HIV prevalence and annual changes in it using the middle estimates of the UNAIDS 2008 data. When middle estimates do not exist, as in the case of the Democratic Republic of Congo, it uses an average of high and low estimates. The system uses data for total population prevalence but also includes HIV prevalence for those 15–49.

26 A more satisfactory approach would use stocks and flows and would have a more strongly systems dynamics character. It would track infected individuals, presumably by age cohorts, but at least in the aggregate. It would compute new infections (incidence) annually, adding those to existing prevalence numbers, transitioning those already infected into some combination of those manifesting AIDS, those dying, and those advancing in age with HIV. But the data do not seem widely available to parameterize such transition rates, especially at age-category levels.

27 Table 17 of the Annex to "World Population Prospects: The 2002 Revision" (UNPD 2003: 77–78) provided such estimates for 38 African countries and selected others outside of Africa; the IFs project has revised and calibrated many of the estimates over time as more data have become available. By 2004–2006, however, quite a number of countries had begun to experience reductions, and this logic has become less important except in scenario analysis for countries where prevalence is still rising.

28 Early epidemic data from sub-Saharan Africa and the United States supported this assumption.

29 Dargay, Gately, and Sommer (2007) also describe an adjustment factor related to the percentage of a total population residing in urban areas; we did not implement that factor.

30 Http://en.wikipedia.org/wiki/Smeed%27s_law. See Adams (1987) and Smeed (1949). Others have disputed the law.

31 For each age-sex-cause-specific regression, HealthExp% was added and tested for significance. After considering Ordinary Least Squares (OLS), random-effects, and fixed-effects models, only the HealthExp% effect on all-cause mortality for the age 0–4 age group was considered sufficiently robust. Because HealthExp% effects are specified as linear, they could be quite large for countries with extraordinarily high levels of HealthExp%, particularly when combined with low GDP per capita. Few such cases exist within the existing distribution, however. For today's countries with GDP per capita below $15,000, HealthExp% has a mean of 6.3%, a standard deviation of 1.6%, and a range from 2.4% to 10.5%. HealthExp% also tends to be somewhat higher for wealthier countries in this group. Using the results implied by these regressions and sensitivity testing of the IFs base model, we find that the effect of a one standard deviation change in HealthExp% on ${}_5q_0$ (about 2.6% lower) is about one-fifth as large as the effect of a one standard deviation change in GDP per capita (about a 14% reduction).

32 "Dictionary of Cancer Terms," National Cancer Institute, http://www.cancer.gov/dictionary/ (accessed January 2010).

33 This is very reasonable for P'(x) by its definition. With respect to RR(x), we assume these to be the same for all countries unless otherwise specified in the CRA reports. Any change over time is likely to be picked up in other parts of our model dealing with changes in technology and the efficiency of health care systems.

34 The mortality rate for the residual category consisting of measles and other communicable diseases is calculated as the difference between the mortality rate for the general category "communicable diseases other than HIV/AIDS" as a whole and the sum of the separate mortality rates for diarrhea, respiratory infections, and malaria.

35 We present further details on these formulations in Chapter 5 in the section on undernutrition and in the technical documentation of the IFs health model at www.ifs.du.edu.

36 In the sequence of our calculations we decompose this equation in practice by finding the individual PAFs, computing their individual independent effects with (1-PAFDistal)/(1-PAFFull), and multiplying mortality independently and cumulatively.

37 In the CRA study on overweight and obesity, James et al. (2004: 498) reported that 58 percent of the global burden of type 2 diabetes was attributable to increases in BMI. Note that the study's assumption of rising BMI rates over time is not always replicated in IFs forecasts.

38 Relative risk estimates taken from Kelly et al. (2009).

39 Dr. Mathers provided us with the GBD project's adjustment factors, which remain constant over time in the forecast (although, since smoking impact changes over time, SIR does change with year). China and a subset of countries in Southeast Asia (Bangladesh, Bhutan, Democratic People's Republic of Korea, India, Maldives, Myanmar, and Nepal) were treated separately from one another and from the single "world" category in which all other countries were combined.