



FREDERICK S. PARDEE CENTER
FOR INTERNATIONAL FUTURES
EXPLORE UNDERSTAND SHAPE

WORKING PAPER 2014.09.01

IFs HEALTH MODEL DOCUMENTATION

Authors: Barry B. Hughes, Cecilia M. Peterson, Dale S.
Rothman, and José R. Solórzano
September 2014

Note: If cited or quoted, please indicate working paper status.



UNIVERSITY *of*
DENVER

JOSEF KORBEL SCHOOL OF INTERNATIONAL STUDIES

PARDEE.DU.EDU

Health Model Documentation

Table of Contents

1.	Introduction.....	4
1.1	Overview.....	4
1.2	Dominant Relations	5
1.3	Structure and Agent Based System: Governance	6
2.	Health Flow Charts	7
2.1	Distal Drivers and Basic Indicators	7
2.2	Smoking and Smoking Impact.....	9
2.3	Proximate Drivers and Risk-Specific Population Attributable Fractions	10
2.3.1	Child Undernutrition.....	11
2.3.2	Body Mass Index and Obesity	12
2.3.3	Indoor Use of Solid Fuels	13
2.3.4	Outdoor Urban Air Pollution	15
2.3.5	Water and Sanitation.....	16
2.4	Specialized Models: Deaths from AIDS and Vehicle Accidents.....	19
2.5	Forward Linkages from Health.....	23
2.5.1	Forward Linkages of Health to Population and Labor Supply	24
2.5.2	Forward Linkages of Health to Capital Stock.....	24
2.5.3	Forward Linkages of Health to Economic Productivity	26
3.	Health Model Equations	28
3.1	Distal Driver Formulation.....	29
3.1.1	Forecasting income and education for the distal driver formulation	30
3.1.2	Forecasting technology for the distal driver formulation of mortality	31
3.2	Diabetes and Chronic Respiratory Diseases	33
3.2.1	Diabetes.....	33
3.2.2	Chronic Respiratory Disease.....	34
3.3	Diarrhea, Malaria, and Respiratory Infections	35
3.4	HIV/AIDS	36
3.5	Cardiovascular Disease.....	38
3.6	Road Traffic Accidents	38
3.7	Mental Health.....	40
4.	Modifications to the Basic Health Model	41
4.1	Mortality Transition for Low-Income Countries	41
4.2	Maintaining Increase with Age in Non-Communicable Death Rates.....	41
4.3	Elasticity of Child Mortality with Health Spending	45
5.	Data Initialization.....	47
5.1	Normalization and Scaling Factors	47
5.2	Cause-specific Mortality: Infants and the Elderly	48
6.	Incorporating Proximate Drivers	50
6.1	Adjusting mortality due to changes in a single risk factor.....	50
6.2	Multiple risk factors	52

6.3	Specific Risk Factors	53
6.3.1	Childhood Undernutrition.....	53
6.3.2	Adult Body Mass Index (BMI) and Obesity.....	55
6.3.3	Water and Sanitation.....	60
6.3.4	Indoor Air Pollution.....	63
6.3.5	Outdoor Air Pollution	65
6.3.6	Smoking Rate and Smoking Impact	66
6.3.6.1	Historical Smoking Rates	67
6.3.6.2	Smoking as a Function of GDP per Capita.....	68
6.3.6.3	Forecasting Changes in Smoking using Past Behavior and GDP per Capita Expectations	68
6.3.6.4	Computing Future Smoking Impact from Future Smoking Rates	70
6.3.6.5	Revised Approach: Forecasting Smoking Rates using a Stages Model ..	71
7.	Outcomes	74
7.1	Years of Life Lost (YLL).....	74
7.2	Years of Life Lost to Disability (YLD)	75
7.3	Disability adjusted life years (DALY)	77
7.4	Morbidity	78
7.5	Mortality probabilities	78
8.	Forward Linkages	80
8.1	Population and Labor Supply.....	80
8.2	Capital stocks	80
8.3	Productivity.....	82
8.4	Additional potential forward linkages	83
9.	Conclusion	85
	References.....	86
	Appendix.....	90

1. Introduction

1.1 Overview

The IFs health model allows users to forecast age, sex, and country specific health outcomes related to 15 cause categories (see table) out to the year 2100. Based on previous work done by the World Health Organization's (WHO) Global Burden of Disease (GBD) project¹, formulations based on three distal drivers – income, education, and technology – comprise the core of the IFs health model. However, the IFs model goes beyond the distal drivers, including both richer structural formulations and proximate health drivers (e.g. nutrition and environmental variables). Integration into the IFs system also allows us to incorporate forward linkages from health to other systems, such as the economic and population modules. Importantly, IFs provides the user the ability to vary model assumptions and create customized scenarios; as such, IFs is a tool exploring how policy choices might result in alternative health futures.

This documentation supplements the third volume of the PPHP series, “Improving Global Health,” (Hughes et al, 2011) by providing technical details of health model integration into the IFs system. It includes the specific equations used to forecast outcomes and drivers, relative risk values for proximate drivers, and data manipulations related to model initialization and projection. We intend the IFs model to be fully transparent to all users, and invite comments and questions at <http://www.ifs.du.edu/contact/index.aspx>.

Cause groups in IFs

Group I – Communicable, Maternal, Perinatal, and Nutritional Conditions

- Diarrheal diseases
- Malaria
- Respiratory infections
- HIV/AIDS
- Other Group I causes

Group II – Noncommunicable Diseases

- Malignant neoplasms
- Cardiovascular diseases
- Digestive diseases
- Chronic respiratory diseases
- Diabetes
- Mental health
- Other Group II causes

Group III – Injuries

- Road traffic accidents
- Other unintentional injuries
- Intentional injuries

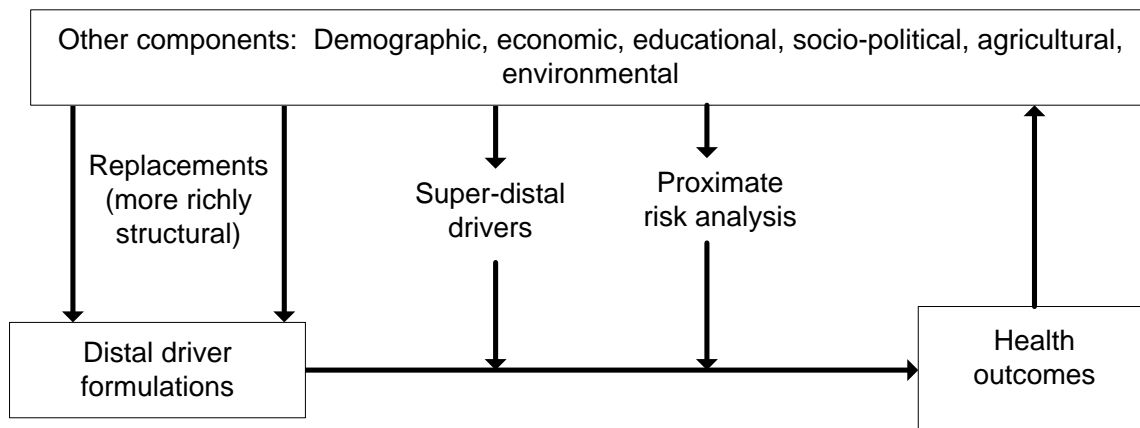
¹ See Mathers and Loncar (2006) for details on GBD projections of cause-specific mortality out to 2030.

1.2 Dominant Relations

Health forecasting systems typically can help us either (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; or (2) to consider opportunities for intervention and achievement of alternative health futures, enhancing the foundation for decisions and actions that improve health.

Broad structural models (e.g., that of the Global Burden of Disease or GBD) assist in the first purpose by relating deep or distal development drivers to outcomes. More specialized structural formulations and the inclusion of proximate drivers open the door to the second, allowing for consideration of interventions in the pursuit of alternate health futures. A more hybrid and integrated model form like that of IFs can help with both purposes and provide a richer overall picture of alternative health futures.

The figure shows the general structure. Formulations based on distal drivers (the GBD methodology) sit at its core. There is no inherent reason, however, that income, education and time (the distal drivers of the GBD approach) 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, for instance, 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; with further societal spread of income and education, at least smoking and road traffic deaths (and perhaps also obesity) typically decline.²



Envisioning a hybrid and integrated health forecasting approach

A hybrid model can therefore help us identify opportunities for interventions to improve health futures. These interventions might also occur in the form of super-distal drivers

² 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.

(for example, policy-driven human action with respect to health systems). The sociopolitical 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.

The integrated nature of the IFs modeling system further allows us to think about feedback loops between health outcomes and larger development variables such as economic progress and population structure.

1.3 Structure and Agent Based System: Governance

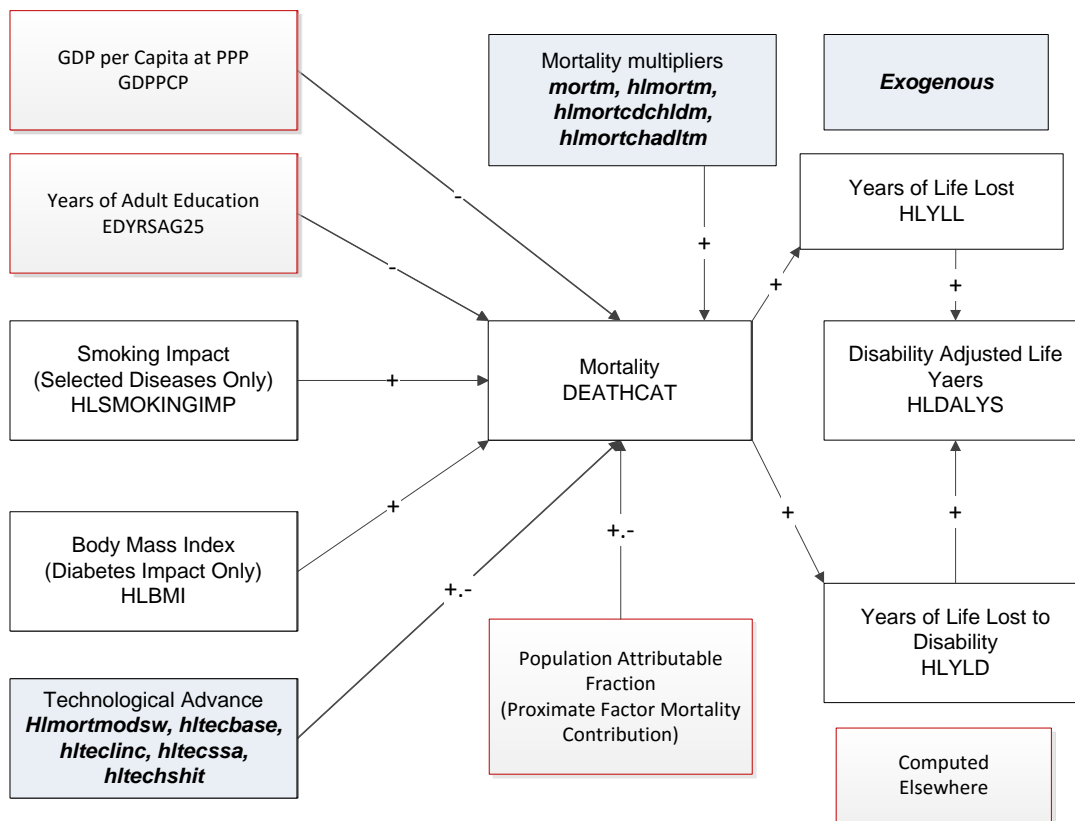
Structure and Agent System: Governance	
System/Sub System	Health
Organizing Structure	Hybrid structure using distal driver formulations supplemented by proximate drivers; integrated with larger IFs systems such as population and governance.
Stocks	Population by age-sex; stunted population; HIV prevalence.
Flows	Births, mortality and morbidity.
Key Aggregate Relationships	Distal driver formulations driven by income, education, and time as a proxy for technological advance. Proximate driver formulations driven by various social patterns and behaviors.
Key Agent-Class Behavior Relationships	Behavior related to proximate drivers such as smoking, indoor solid fuel use, obesity.

2. Health Flow Charts

Mortality from most causes of death is a function of a small number of distal or deep drivers and a larger number of proximate or more immediate drivers. For two specific mortality types, however, specifically deaths from AIDS and vehicle accidents, there are more specialized representations that rely on a number of more cause-related drivers.

2.1 Distal Drivers and Basic Indicators

To forecast mortality related to most of the major cause clusters we use the regression models and associated beta coefficients prepared for the GBD project (Mathers and Loncar 2006). Age, sex, cause, and country-specific mortality rate is a function of income (using GDP per capita as a proxy), adult education, technological progress. For specific death causes, smoking impact (for malignant neoplasms, cardiovascular disease, and respiratory disease) or body mass index (for diabetes only) add to the causality; see the discussion of flow charts and equations for information on the determination within IFs of smoking and smoking impact and of body mass index and obesity.



A number of parameters control technology in the distal functions. In the default mode ($hlmortmodsw = 1$), IFs modifies the technology (time) coefficient in recognition of slower than expected historical progress in many countries, an approach developed in the Global Burden of Disease (GBD project). Those country differences are controlled by

hltechbase, *hltechlinc*, and *hltechssa*. Setting the switch value to 0 activates an alternative IFs project approach to the impact of those parameters.

The user can also affect the mortality patterns directly with several parameters, including *mortm*, which allows simultaneous manipulation of all causes of death and *hlmortm*, which facilitates manipulation of each cause of death separately. *Hlmortcdchldm* changes the rates of all communicable diseases for children aged 5 and younger, while *hlmortcdadltm* affects rates of death from communicable diseases for adults aged 15-49.

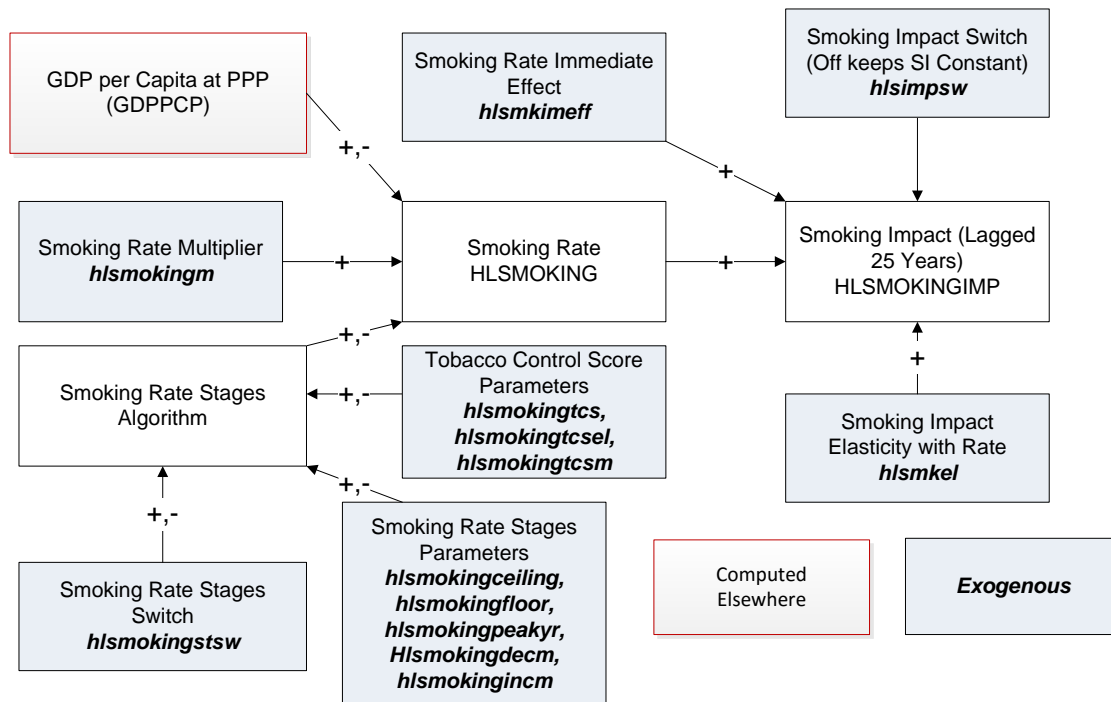
Based on the mortality level, it is possible to compute the years of life lost to each cause of death (HLYLL). Using WHO-based estimates, IFs links mortality also to years of living with disability (HLYLD). The sum of the two is disability-adjusted life years lost (HLDALYS)

The forecast of mortality in this figure, dependent almost entirely on distal factors, is not actually the final calculation in the model. See the discussion of the entry of proximate drivers into the discussion of population attributable mortality fractions (PAFs), in interaction with distal-based mortality, for the rest of the story.

Because of the importance of smoking impact in the distal driver formulation, it is important that we elaborate that term. Body mass index is, at this point, only linked to diabetes and we discuss that in the context of the PAFs.

2.2 Smoking and Smoking Impact

Of the various specific health risks that the model treats, smoking has a special place because its impact is in the distal driver formulation of the IFs health model. The figure shows that the impact is driven by the rate of smoking (differentiated by males and females) 25 years earlier, with the relationship controlled by an impact elasticity (*hlsmkel*). The user can also posit as a nearer term (in the model immediate) impact by setting a switch for that (*hlsmkimeff*) at some fractional value of the full delayed impact—the value in the base case is 0.1 or 10 percent. For analysis purposes, another switch (*hlsmimpsw*) can turn off the endogenous computation of smoking impact and leave it constant at the initial year value.



Smoking rate itself is computed in two different ways. The basic formulation uses only the initial condition and a function linked to the simple and squared values of GDP per capita at PPP. The more extended formulation is an algorithmic one based on the same general concept of a pattern that initially rises with GDP per capita, peaks, and then falls, but with a series of parameters that allow much more control over the stages.³ This staged algorithmic approach (see Lopez et al. 1994; Shibuya et al. 2005; Ploeg et al. 2009) is turned on with a switch (*hlsmokingstsw*).

Because control of tobacco is a major policy objective in many countries, there is also a representation of a tobacco control score on a 100-point scale (*hlsmokingtcs*) with an associated parameter to control the elasticity of smoking with that score (*hlsmokingtcsel*), as well as a multiplier on the score (*hlsmokingtcsm*).

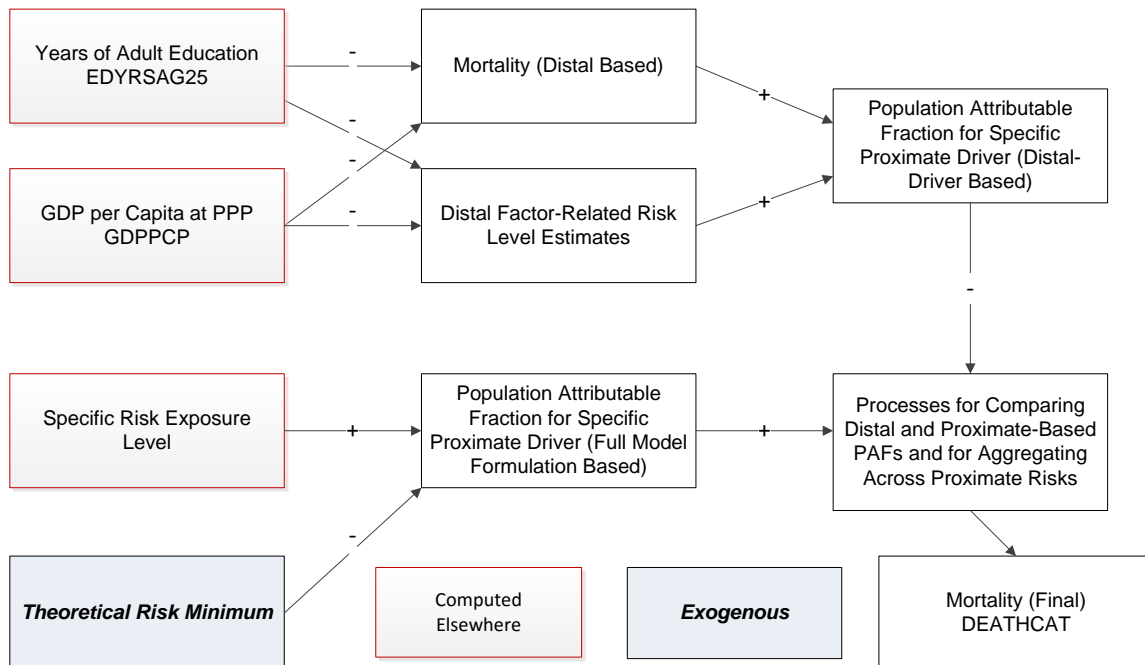
³ Cecilia Peterson developed this approach for IFs.

Finally, there is a multiplier that allows direct manipulation of the smoking rate, again by sex (*hlsmokingm*).

2.3 Proximate Drivers and Risk-Specific Population Attributable Fractions

Although mortality can be calculated solely from distal drivers such as income and education, it is better to calculate it from proximate or more immediate factors, such as undernutrition or exposure to pollutants. But IFs, and perhaps any model, will never be able to represent all such proximate drivers. Hence there is value in having an approach that combines the use of distal and proximate drivers, supplementing and adjusting the distal-driver based approach whenever possible.

The figure below shows such a combination. Each proximate driver (and there are many different ones in IFs, in spite of the generalized representation in the figure) can be associated with a fraction of the mortality of a society. That population attributable fraction or PAF (derivative from the risk exposure level relative to a theoretical risk minimum) can be used to adjust the mortality associated with any cause that would have an implicit risk-related mortality built into the distal driver formulation. IFs makes those implicit distal-driver associated risk levels explicit by using the distal drivers to identify a risk level that would be expected based on cross-sectional analysis using the distal drivers. That allows the computation of a distal-driver based PAF. In similar fashion a PAF can be calculated that relates an exposure level to the risk, calculated mostly elsewhere in IFs (such as in the food and agriculture model for undernutrition of children) to a PAF.



The complication mathematically lies in the interaction of (1) the distal-driver and proximate risk-based PAFs and (2) the multiple specific-risk PAFs, because avoidance of

death from one will generally increase the risk of death from others. See the equations associated with PAFs for details.

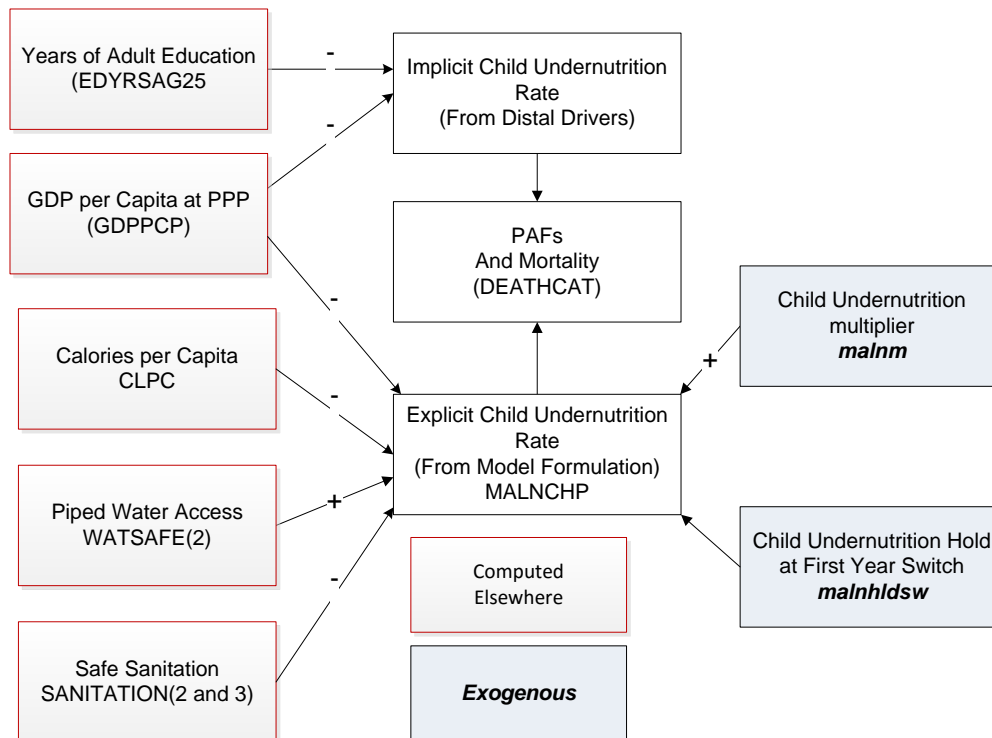
Among the specific risk factors treated in IFs are overly high body mass indices and associated obesity, undernutrition of children, access to unsafe water and sanitation, indoor use of solid fuels, and levels of urban air particulates.

2.3.1 Child Undernutrition

Although obesity is a growing problem and killer around the world, the most important risk factor for children in particular has traditionally been undernutrition (often simply referred to as malnutrition). The percentage of children undernourished (MALNCHP) affects mortality rates from communicable diseases in particular via the mechanism that the model uses to modify cause-specific mortality from the distal driver formulation by actual risk level in a country. The core of that approach is to compare the risk-specific population attributable fraction (PAF) of total mortality as calculated from the distal drivers with the PAF calculated from the actual level of the risk in the country.

The figure below shows the approach for childhood undernutrition. The two key variables in the distal driver formulation at any point in time (ignoring the technology factor that adds dynamics over time) are GDP per capita at purchasing power parity and years of adult education. They are used in a cross-sectionally estimated function to calculate an implicit body mass index that then produces the associated implicit PAF. IFs uses an alternative and more risk-factor specific formulation to forecast values of child undernutrition over time. The PAF associated with this explicit representation of MALNCHP is compared with the PAF from the implicit calculation and the comparison alters the actual mortality pattern.

To calculate MALNCHP the explicit formulation also uses GDP per capita, as in the distal formulation, but augments it with calories per capita and with access to safe water and sanitation (unsafe water can cause diarrheal disease and undernutrition even with caloric intake would be adequate). A multiplicative parameter (*malnchpm*) can be used to change child undernutrition in scenario analysis. Another parameter (*malnchpsw*) can be used to hold the level of undernutrition at the level of the first year, an approach useful for counterfactual scenario analysis.



Although not used in the health model, IFs contains two other measure of undernutrition. The first is an alternative measure of child undernutrition developed by Smith and Haddad (2000); MALNCHPSH is computed as a function of the ratio of female and male life expectancy, of female secondary school gross enrolment rate, and of access to safe water. The second is a measure of rate of undernutrition for the entire population (MALNPOPP), computed as a function only of calories per capita.

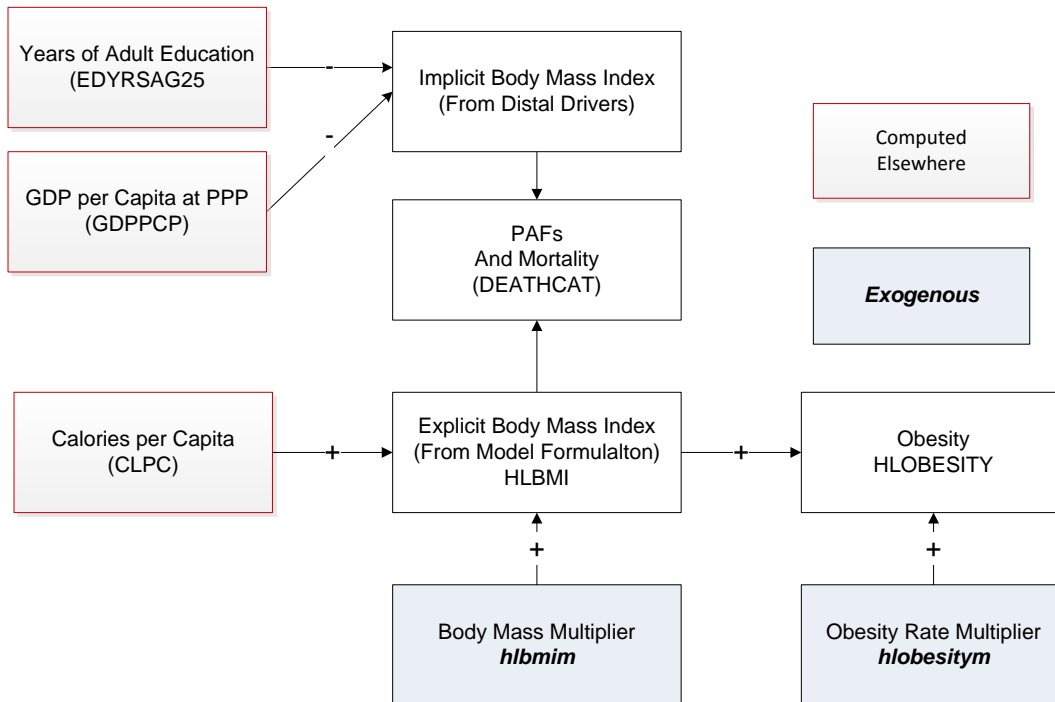
2.3.2 Body Mass Index and Obesity

The distal driver formulation used for forecasting mortality in IFs contains a country's average body mass index (HLBMI) for diabetes. HLBMI also affects mortality from cardiovascular disease in IFs via the mechanism that the model uses to modify cause-specific mortality from the distal driver formulation by actual risk level in a country. The core of that approach is to compare the risk-specific population attributable fraction (PAF) of total mortality as calculated from the distal drivers with the PAF calculated from the actual level of the risk in the country.

The figure below shows the approach for body mass index. The two key variables in the distal driver formulation at any point in time (ignoring the technology factor that adds dynamics over time) are GDP per capita at purchasing power parity and years of adult education. They are used in a cross-sectionally estimated function to calculate an implicit body mass index that then produces the associated implicit PAF. IFs uses an alternative and more risk-factor specific formulation to forecast values of body mass index over time. The PAF associated with this explicit representation of HLBMI is compared with

the PAF from the implicit calculation and the comparison alters the actual mortality pattern.

To calculate HLBMI the explicit formulation uses calories per capita as the sole driving variable. A multiplicative parameter (*hlbmim*) can be used to change HLBMI in scenario analysis. A forecast of the obese population as a percent of the total population (HLOBESITY) is driven by the body mass index. A separate multiplicative parameter can modify it (*hlobesity*).



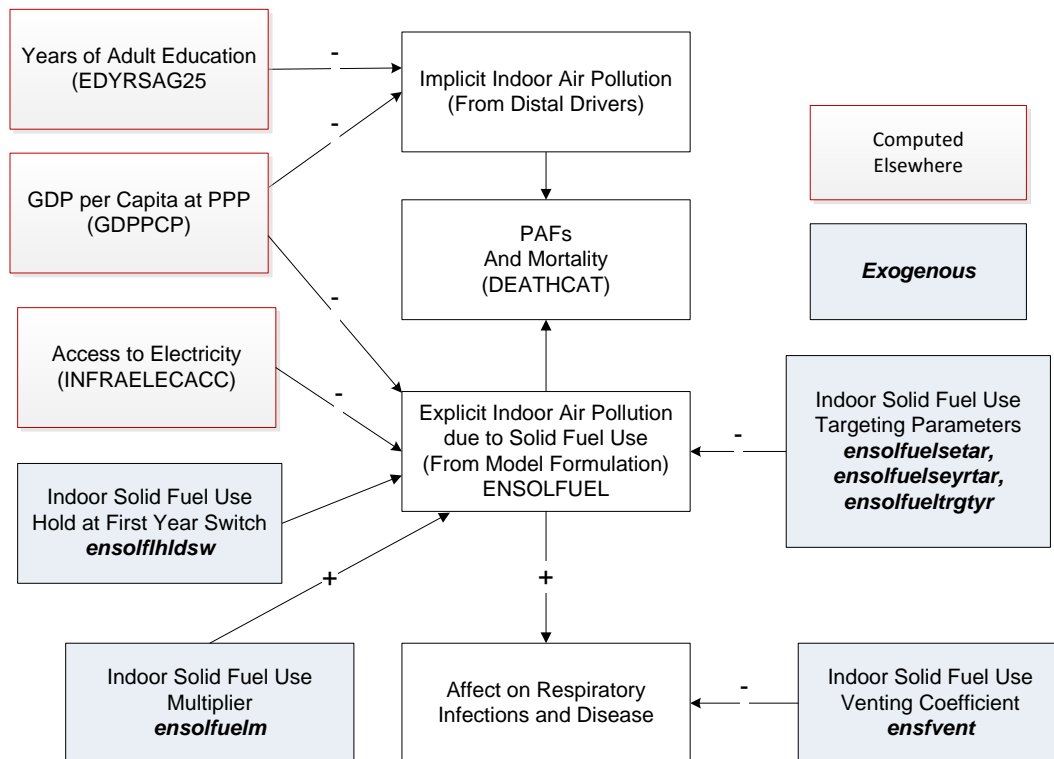
2.3.3 Indoor Use of Solid Fuels

One of the most important health risk factors in the developing world, especially for women and children under 5 is the use of solid fuels for cooking (and heating) indoors (ENSOLFUEL). It is a major cause of respiratory diseases. In IFs it affects mortality rates via the mechanism that the model uses to modify cause-specific mortality from the distal driver formulation by using information concerning actual risk level in a country. The core of that approach is to compare the risk-specific population attributable fraction (PAF) of total mortality as calculated from the distal drivers with the PAF calculated from the actual level of the risk in the country.

The figure below shows the approach for indoor air pollution from the use of solid fuels. The two key variables in the distal driver formulation at any point in time (ignoring the technology factor that adds dynamics over time) are GDP per capita at purchasing power parity and years of adult education. They are used in a cross-sectionally estimated function to calculate indoor air pollution (linked to solid fuel use) that then produces the

associated implicit PAF. IFs uses an alternative and more risk-factor specific formulation to forecast values of solid fuel use over time. The PAF associated with this explicit representation of ENSOLFUEL is compared with the PAF from the implicit calculation and the comparison alters the actual mortality pattern.

To calculate ENSOLFUEL the explicit formulation also uses GDP per capita, as in the distal formulation, but augments it with access to electricity. For the actual equation, see the topic on equations for solid fuel use in the infrastructure documentation. A multiplicative parameter (*ensolfuelm*) can be used to change solid fuel use in scenario analysis. Another parameter (*ensolhldsw*) can be used to hold the rate of solid fuel use at the level of the first year, an approach useful for counterfactual scenario analysis.



Major factors affecting the health impact of indoor solid fuel use are the efficiency and ventilation of the stoves. The model provides a coefficient (*ensfvent*) for scenario analysis concerning those factors.

Much analysis on this health issue will want to use control of solid fuel use, partly through the use of a multiplier (*ensolfuelm*). There is also targeting of solid fuel use and the model provides two different kinds of targeting parameters, absolute and relative. The absolute (or universal) targeting allows the setting of a year (*ensolfueltrgyr*) by which solid fuel use would be eliminated; it is available country by country. The relative targeting approach, available only globally across all countries, allows the setting of a value based on the typical rate of solid fuel use at different levels of GDP per capita (estimated cross-sectionally). A target rate (*ensolfuelsetar*) would normally be no higher than the typical rate at the country's level of GDP per capita and could be, for instance,

one standard error lower than the typical rate. An associated parameter (*ensolfuelseyrtar*) identifies the number of years over which a country would move to the target level. If a country already meets or exceeds a relative target, it will not move (adversely) toward it. Moreover, only the absolute or relative target should be used in analysis, not both together—an attempt to use both together will result in neither being used.

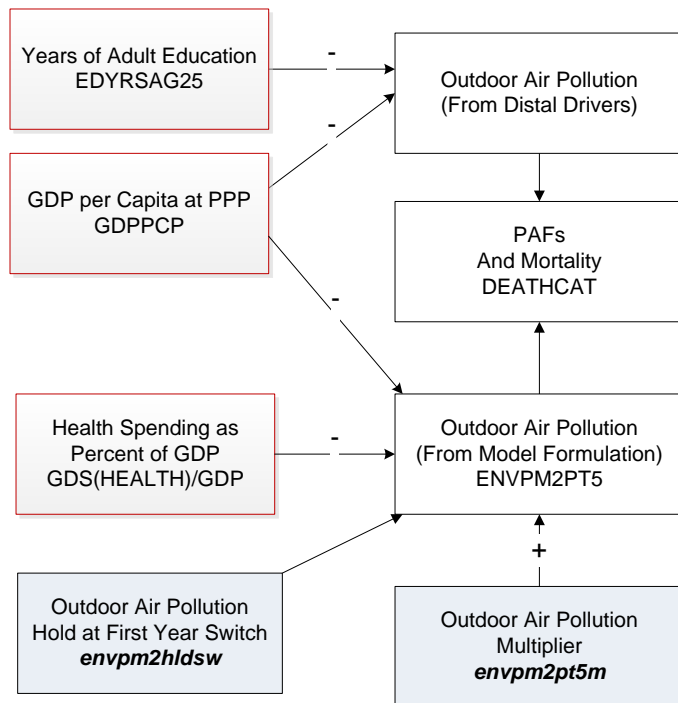
2.3.4 Outdoor Urban Air Pollution

One of the more important health risk factors in the developing and developed world alike, especially for middle-income industrializing countries, is the concentration of particulate matter of diameter 2.5 micrometers or less per cubic centimeter in urban air (ENVPM2PT5).⁴ It is a major cause of respiratory infections, respiratory diseases, and cardiovascular disease in adults 30 and older. In IFs it affects mortality rates via the mechanism that the model uses to modify cause-specific mortality from the distal driver formulation by using information concerning actual risk level in a country. The core of that approach is to compare the risk-specific population attributable fraction (PAF) of total mortality as calculated from the distal drivers with the PAF calculated from the actual level of the risk in the country.

The figure below shows the approach for outdoor urban air pollution, focusing on the measure of ENVPM2PT5. The two key variables in the distal driver formulation at any point in time (ignoring the technology factor that adds dynamics over time) are GDP per capita at purchasing power parity and years of adult education. They are used in a cross-sectionally estimated function to calculate outdoor air pollution that then produces the associated implicit PAF. IFs uses an alternative and more risk-factor specific formulation to forecast values of outdoor urban air pollution use over time. The PAF associated with this explicit representation of ENVPM2PT5 is compared with the PAF from the implicit calculation and the comparison alters the actual mortality pattern.

To calculate ENVPM2PT5 the explicit formulation also uses GDP per capita, as in the distal formulation, but augments the spending of a country on health as a portion of GDP (which appears to serve reasonably well as a proxy for more general attention to the environment). For the actual equation, see the topic on outdoor urban air pollution equations in the health documentation. A multiplicative parameter (*envpm2pt5m*) can be used to change urban air pollution in scenario analysis. Another parameter (*envpm2hldsw*) can be used to hold the level of urban air pollution at the level of the first year, an approach useful for counterfactual scenario analysis.

⁴ Initialized in IFs by converting World Bank data on PM10 concentrations.



2.3.5 Water and Sanitation

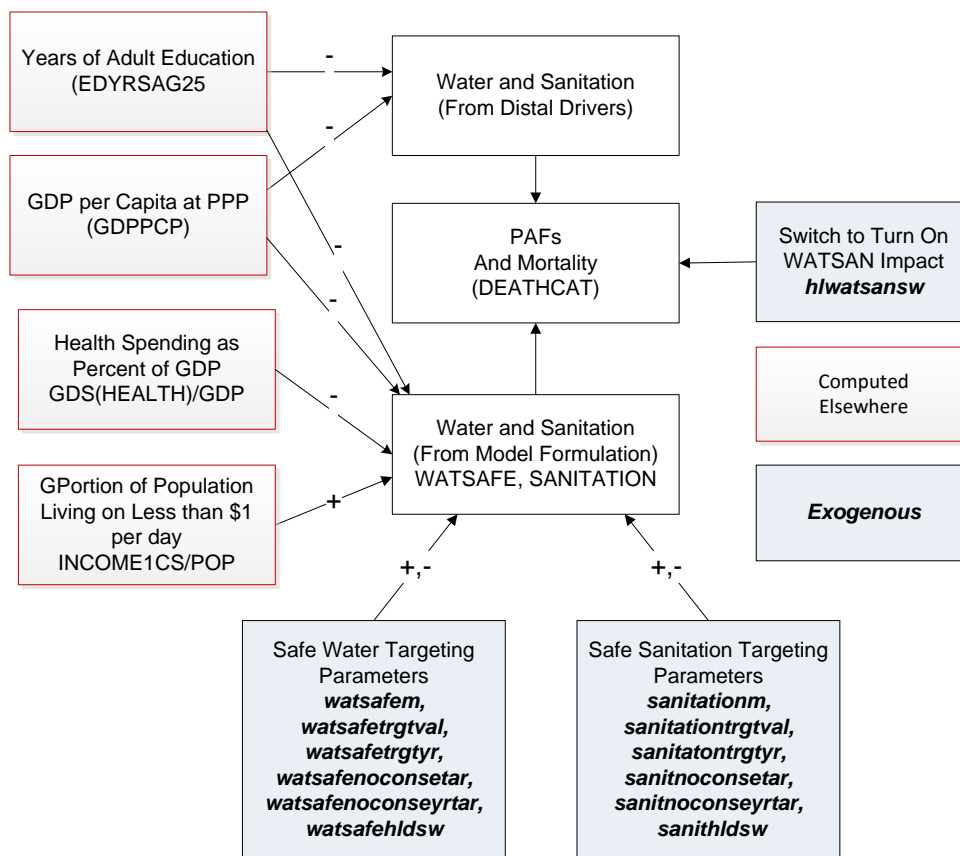
Although unsafe water and sanitation is a killer via its contribution to undernutrition of children in particular, it creates its own mortality risk especially via diarrheal disease. The variables of importance in IFs are access to safe water (WATSAFE) and safe sanitation (SANITATION). In IFs they affect mortality rates via the mechanism that the model uses to modify cause-specific mortality from the distal driver formulation by using information concerning actual risk level in a country. The core of that approach is to compare the risk-specific population attributable fraction (PAF) of total mortality as calculated from the distal drivers with the PAF calculated from the actual level of the risk in the country.

The figure below shows the approach for safe water and sanitation. The two key variables in the distal driver formulation at any point in time (ignoring the technology factor that adds dynamics over time) are GDP per capita at purchasing power parity and years of adult education. They are used in a cross-sectionally estimated function to calculate unsafe water and sanitation that then produces the associated implicit PAF. IFs uses alternative and more risk-factor specific formulations to forecast values of safe access to water and sanitation over time. The PAF associated with this explicit representation of WATSAFE and SANITATION (in combination) is compared with the PAF from the implicit calculation and the comparison alters the actual mortality pattern.

To calculate WATSAFE and SANITATION (separately) the explicit formulation also uses both average years of adult education and GDP per capita, as in the distal formulation, but augments those with the spending of a country on health as a portion of GDP (which appears to serve reasonably well as a proxy for more general attention to the

environment) and portion of the citizenry living on less than \$1.25 per day. For the actual equations, see the topic on outdoor urban air pollution equations in the infrastructure documentation.

Both access to safe water and to safe sanitation have ladders of access quality ranging from none to household connections. Parameters affecting them must thus take into account those ladders and the specific level(s) the parameter affects. Multiplicative parameters (*watsafem* and *sanitationm*) can be used to change access at any level on the two ladders (the model normalizes access across levels to assure summation to 100 percent). Another parameter pair (*watsafehldsw* and *sanithldsw*) can be used to hold the level of access at that of the first year, an approach useful for counterfactual scenario analysis.



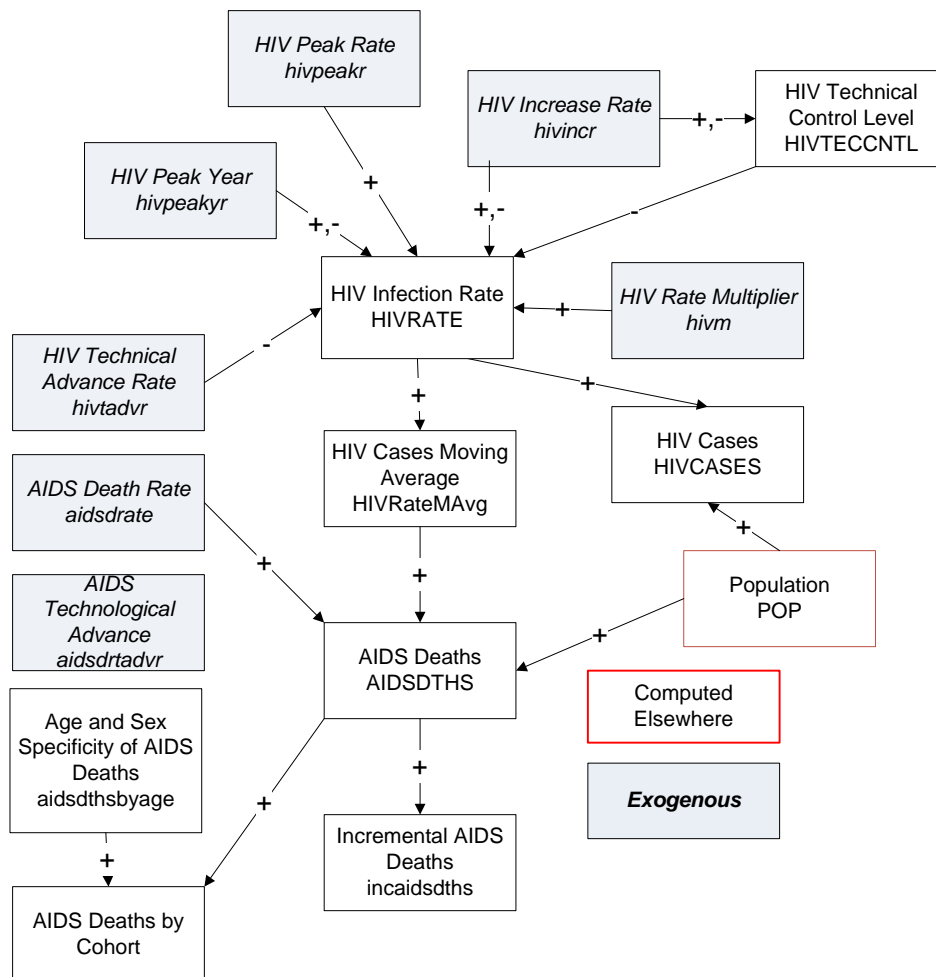
Other parameters control targeting, both universal and relative. With respect to absolute targeting, *watsafetrgtval* and *watersafetrgtyr* control those with no access to safe water (the proportion and the number of years to reach the target, respectively). Similarly, *sanitationtrgtval* and *sanitationtrgtyr* control those with access to household connections. The relative targeting approach, available only globally across all countries, allows the setting of a value based on the typical rate of access at different levels of GDP per capita (estimated cross-sectionally). A target level (*watsafenoconsetar*, *sanithhconsetar*, *sanitnoconsetar*) would normally be no better (which could mean no higher or no lower) than the typical level at the country's level of GDP per capita and

could be, for instance, one standard error better (higher or lower depending on the variable being targeted) than the typical level. An associated parameter (*watsafenoconseyrtar*, *sanithhconseyrtar*, *sanitnoconseyrtar*) identifies the number of years over which a country would move to the target level. If a country already meets or exceeds a relative target, it will not move (adversely) toward it. Only the absolute or relative target should be used in analysis, not both together—an attempt to use both together will result in neither being used.

2.4. Specialized Models:

2.4.1 Deaths from AIDS and Vehicle Accidents

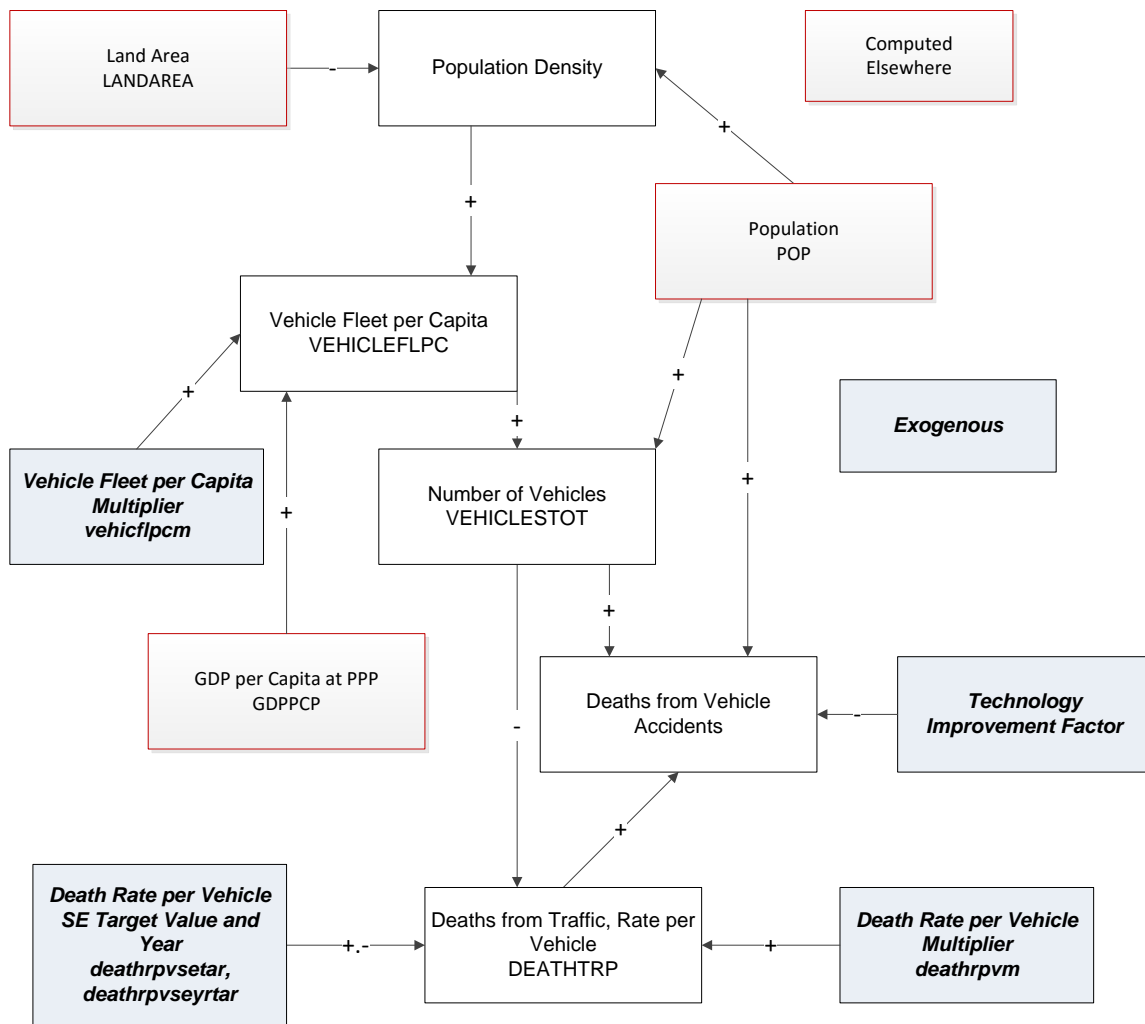
AIDS deaths depend very directly on the number (or stock) of HIV-infected individuals, which depends in turn on the HIV-infection rate (HIVRATE). Data on HIV infection rates were used to compute a basic, country/region-specific rate of HIV infection increase (hivincr), which the user can alter, as they can exogenous assumptions about the peak year of the epidemic (hivpeakyr) and the infection rate in that year (hivpeakr). If a country is beyond the peak year of the epidemic, control will be bringing the rate down over time (HIVTECCNTL). The user may also rely upon a country/region-specific multiplier to move rates up or down (hivm).



There is both a policy and medical effort underway to reduce the growth in infections. An HIV technical advance rate (hivtadvr) represents the success of that in rate of reduction in annual infection growth, and a variable (HIVTECCNTL) shows the cumulative impact of changes past a peak rate and year. Although highly speculative, the user will recognize the long-term importance of such assumptions.

Turning from the infection rate to the death rate, the user can make changes in the initial AIDS death rate (aidsrate) to reflect possible progress or lack of it in reducing the deaths from HIV (using aidsdrtadvr). When the deaths from AIDS are computed, they are used to compute an incremental number of deaths (since some are already in the mortality of the base year); and an exogenous vector spreads them by age and sex.

The computation of deaths from vehicle accidents starts with computing the number of vehicles per capita (VEHICFLPC) as a function of population density and GDP per capita. That allows computation of the total number of vehicles (VEHICLESTOT). The function used by IFs to compute the total number of road deaths uses the number of vehicles and the population.



2.4.2 Specialized Models: Severe Acute Malnutrition

Severe Acute Malnutrition (SAM) (also known as severe wasting) is a deficiency of protein or other minerals and vitamins leading to a loss of body fats and muscle tissues. SAM can be defined by a Weight-for-Height Z-score <-3 compared to the median WHO

growth standards; by visible severe wasting; or by the presence of nutritional oedema (Prudhon, Briend, Prinzo, Daelmans, & Mason, 2006). It has been identified as a significant cause of death and disability globally (Black et al., 2008, 2013). Children with SAM face a risk of morbidity and mortality nine-to-eleven times greater than their healthy counterparts (Lenters, Wazny, & Zulfiqar, 2016; UNICEF, Nutrition Section, United Nations System Standing Committee on Nutrition, World Food Programme, & World Health Organization, 2007). SAM may result in cognitive impairment, delayed school enrollment, vulnerability to chronic diseases, and reduced productivity.

A wide-range of interventions including policies to improve food security, social protection and supplementation exist to treat moderate forms of malnutrition. The 2015 Global Burden of Disease Study estimates that the global prevalence of childhood stunting declined by 33% between 1990 and 2015 (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016). Prevalence data used to estimate the global burden of SAM have fluctuated across time: 19 million (Black et al., 2008), 17.3 million (UNICEF, WHO, World Bank, 2012), and 19 million (Black et al., 2013). Despite the fluctuation, these data suggest SAM remains a persistent health challenge.

There is a consensus within the literature that SAM is the result of a combination of interconnected proximate (basic) and distal (underlying) factors. Proximate drivers of SAM include lack of access to food (Haddad & Smith, 2000), poor infant-feeding practices (Acharya, Choudhary, & Shekhawat, 2014), maternal health and education (Dereje, 2014; Haddad & Smith, 2000), and the general burden of infectious disease. Distal drivers of SAM include general level of economic development (Alderman, Appleton, Haddad, Song, & Yohannes, 2001), poverty levels (Black et al., 2013; Haddad & Smith, 2000), gender inequality (Bhutta et al., 2008; Mucha, 2012), conflict (Hall, Blankson, & Shoham, 2011), corruption (Halleröd, Rothstein, Daoud, & Nandy, 2013; Ngozi Uchendu & Olatunsbosun Abolarin, 2015), and poor governance (Haddad & Smith, 2000; Halleröd et al., 2013), and transportation infrastructure (Ulimwengu, Funes, Headey, & You, 2009).

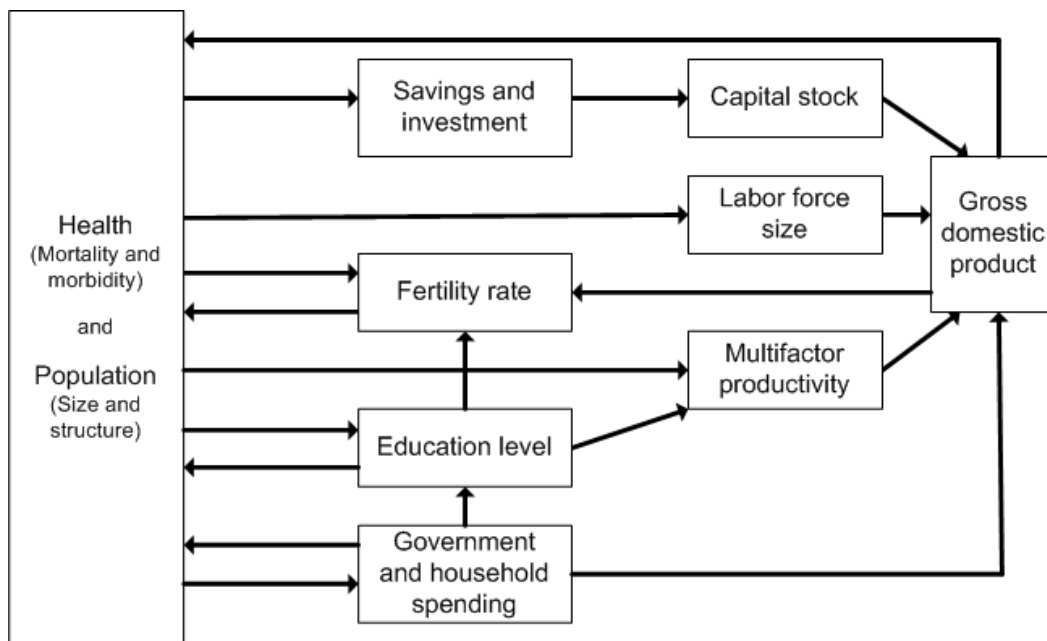
The figure below shows the approach for forecasting annual SAM prevalence in the under 5 population (MALNCHPSAM). The two key variables in the distal driver formulation at any point in time (ignoring the technology factor that adds dynamics over time) are GDP per capita at purchasing power parity and years of adult education. They are used in a cross-sectionally estimated function to calculate an implicit body mass index that then produces the associated implicit PAF. IFs uses an alternative and more risk-factor specific formulation to forecast values of SAM over time. The PAF associated with this explicit representation of MALNCHPSAM is compared with the PAF from the implicit calculation and the comparison alters the actual mortality pattern.

to forecast admissions in three different ways: (0) forecasts all countries according to the relationship with GDPPCP, (1) holds all initial zero values constant over time (default value), and (2) holds all initial values constant.

Success of treatment as a percent of all admission (SAMAVERTED) is initialized from the same data source (UNICEF, 2016), and assumed to be 67 percent (the global average) for any country with missing data. A multiplicative parameter (*samavertedm*) can be used to change SAM in scenario analysis. The model also assumes the success rate to converge to 100 percent over a 100-year time horizon, with a parameter (*samavertedconvyr*) that allows users to change the years to converge to 100.

2.4 Forward Linkages from Health

Chapter 7 of Hughes, Kuhn, Peterson, Rothman, and Solorzano (2011) elaborated the forward linkages of the health model to other parts of the IFs system at the time of that volume's completion. It begins by discussing a controversy in the literature about whether the effects on economic well-being (as indicated by GDP per capita) of improvements in life expectancy are positive or negative. It goes on to devote much attention to three major and general pathways of impact between health and GDP, each of which corresponds to an element in standard production functions and that in IFs. The diagram below shows the major pathways between health/demography and GDP, each of which requires elaboration by showing the variables and logic of the IFs system; those three are labor, capital, and multifactor productivity.

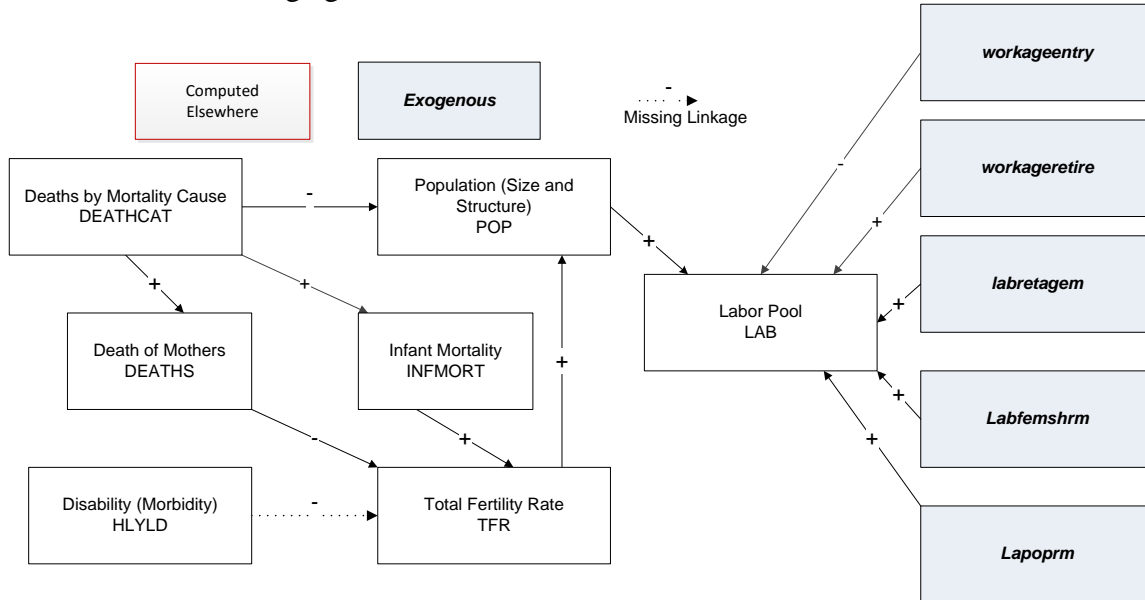


There are other potential forward linkages of health in the IFs system, many of which would have additional implications for economic production. Those other possible forward linkages include linkages of health (or lack of it) to public spending on health

and to education years and quality of it. Potentially there could also be a linkage in the system from health to economic inequality.

2.4.1 Forward Linkages of Health to Population and Labor Supply

The IFs demographic model captures the mechanical or accounting effects of mortality on population (see the solid paths in the figure below). A key pathway passes from mortality through adult age population to labor supply (including aging-related lags).⁵ Similarly, IFs captures the mechanical effect of mortality on fertility through the death of women of childbearing age.



The most important non-mechanical linkage is almost certainly the relationship between child mortality and fertility. IFs forecasts fertility as a relationship with infant mortality, the log of educational level of those aged 15 and older (neither the education of women alone nor the education of those 15-24 work as well), and the percentage use of modern contraception.

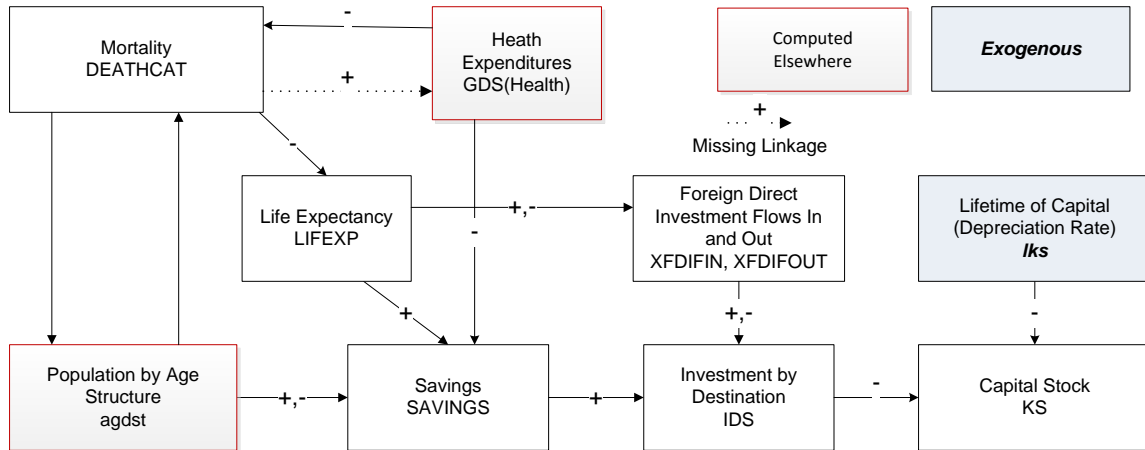
2.4.2 Forward Linkages of Health to Capital Stock

The figure below sketches the primary paths between health (morbidity and mortality) and capital stock. Most capital stock consists of buildings and machinery for producing goods and services; some representations may include land also, but most treat land separately and largely as a constant (although land developed for crop production or grazing can, in fact, be highly variable, as it is in the IFs agricultural model). Most immediately, investment increases capital stock and depreciation reduces it. Although there is certainly some impact of morbidity and mortality on the rate of depreciation of both built physical and natural capital, the relationship may not be substantial and we do

⁵ IFs also includes income-based formulations for changing the female participation rate.

not understand it well enough to model it. Investment is responsive to both domestic savings and foreign flows.

Turning our gaze to the paths by which health affects investment, the three major ones run through health spending, which can crowd out savings and investment, through the age-structure of societies, which affects the savings rate, and through investment from abroad, which can augment that generated domestically.



With respect to health spending, to which we return later, the IFs model uses a social accounting matrix (SAM) structure. Thus the flow of funds into health spending automatically competes with other consumption uses and with savings and investment. The major current weakness of the model with respect to this path is that there is no linkage from morbidity (associated in IFs with mortality) and health expenditures. (There is a linkage in IFs back from health spending to mortality –all categories except for AIDS).

The paths in IFs that link age structure most directly to domestic savings have two important elements. The most fundamental one represents the understanding of life-cycle dynamics in income, consumption and savings. The cycle for income is fairly clear-cut with a peak in the middle to latter periods of the working years. Workers set aside some portion of income as savings and that portion, too, tends to peak in the middle and late period of working years. Society-wide savings themselves become negative after retirement age (65 in the Base Case scenario but variable in scenarios) even though some portion of the population will continue to work. The second fundamental element is that both the horizon of life expectancy and the average income level of a society can have an impact on the portion set aside for savings and the degree to which it rises and then falls. Thus, for example, the life-cycle “bulge” of savings may be earlier and flatter in developing countries.

We implemented the representation of savings and investment in accord with that understanding. Relying upon analyses of selected countries that Fernández-Villaverde and Kruegger (2004 and 2005) and Deaton and Paxson (2000) undertook, we extracted

general stylized patterns of the savings life cycle to represent more and less developed (and lower life expectancy) countries. In forecasting we use the pattern for less developed countries when life expectancy falls below 40 years, use that for more developed countries when life expectancy exceeds 80 years, and interpolate in between for all other countries. The result of this largely algorithmic approach⁶ is an adjustment factor (SavingsAgeAdj) that augments or reduces investment.

In addition, investment is somewhat augmented or reduced as a direct result of changing life expectancy. Life expectancy is compared over time with an expected value (tied to cross-sectional estimation with income). That difference is compared to the difference in the initial year and, if it rises, augments investment.

Although conceptually tied to savings rates, neither the life-cycle analysis nor the life-expectancy term directly affect savings in IFs. Instead, they affect investment directly and savings indirectly via the dynamics in IFs that balance savings and investment over time.

The path linking health to foreign direct investment is potentially quite important. Alsan, Bloom and Canning (2006: 613) reported that one additional year of life expectancy boosts FDI inflows by 9 percent, controlling for other variables. We have implemented that relationship in IFs. The representation of FDI in IFs captures the accumulation over time of FDI inflows in stocks of FDI, as well as the accumulation of FDI outflows in stocks. In addition, the stocks set up their own dynamics, including the tendency for stocks to reinforce flows. For that reason, we have set the base case parameter for the impact of each year of life expectancy on FDI flows to 0.05 (5 percent), lower than the estimate of Alsan, Bloom and Canning (2006).

2.4.3 Forward Linkages of Health to Economic Productivity

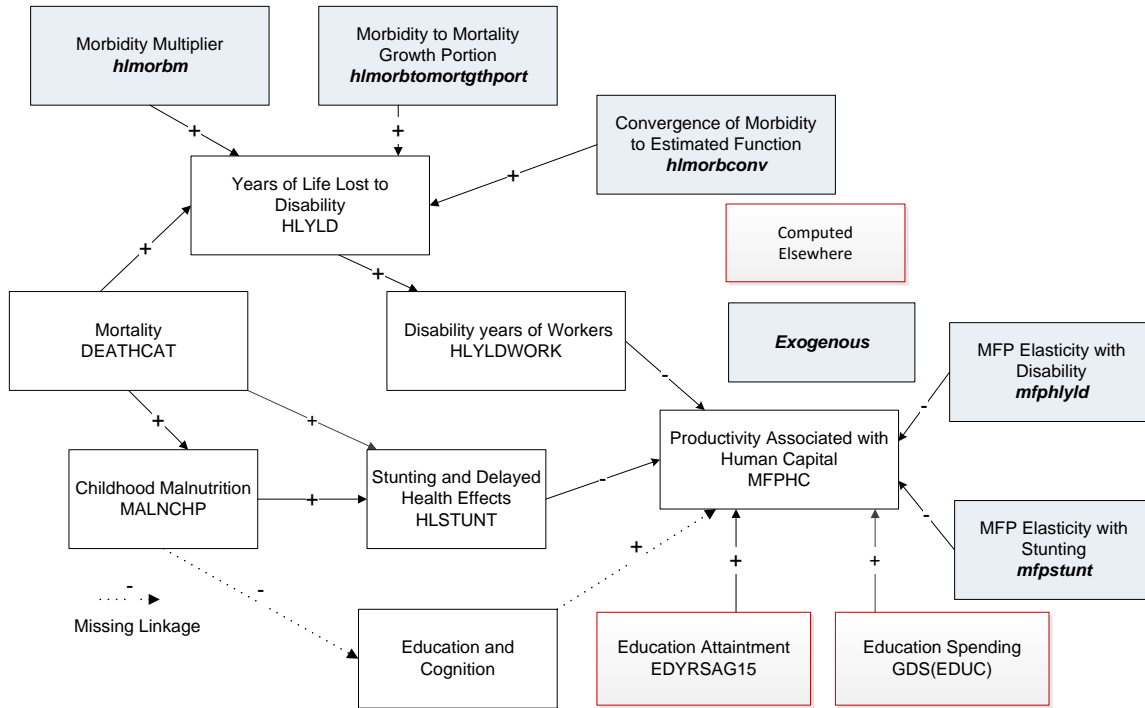
Health outcomes impact productivity through a variety of pathways (see the figure below). Overall the function for multifactor productivity from human capital (MFPHC) is a sum of terms linked to educational expenditures (GDS(EDUC)) as a portion of GDP and to educational attainment of adults in society (EDYRSAG15) with two more directly health-related terms of interest to us here, respectively from adult stunting (STUNTCONTRIB) and disability of those in their working years (HLYLDWORK).

In the IFs health module, the prevalence of adult stunting (HLSTUNT) relates negatively to overall productivity via an elasticity (*mfpstunt*). In extreme cases, stunting could cost as much as 1 percent of economic growth.

We compute HLSTUNT in the health model itself. We initialize adult stunting in a long-term lagged relationship (using a moving average of 25 years) with child malnutrition

⁶ See the subroutine SavingsDemogAdj in routine Populat.bas, which draws upon table IncConSav in IFs.mdb with different patterns of income, consumption, and savings for more developed countries (MDCs) and less developed countries (LDCs) across age categories; in general, peaks of income, consumption, savings occur the in late 40s and savings turn negative at 65.

(MALNCPH) and forecast it as a function of both malnutrition and child mortality as a proxy for morbidity



Turning to disability (which is driven by mortality rates), childhood malnutrition and morbidity do not give rise to all of disability in working years; much also comes from disabilities arising during the working years. IFs therefore also calculates millions of years of living with disability related to mortality rates specific to the working aged-population (HLYLDWORK).

Turning to the forecasting relationship between disability and productivity, the IFs approach drives changes in the growth of productivity from the changing difference between computed and expected values of disability. We used the world average disability rate as an “expected” value. Because we have replicated the practice of the GBD project and kept mental health disability rates constant over time, and because mental health generally dominates disability, forecasts of this disability term are relatively stable over time. Thus analysis with respect to this variable will depend on scenarios that increase or decrease those disability rates. Changes in disability levels (relative to expected ones) relate to change in productivity via a parameter, *mfphlyd*.

3. Health Model Equations

The hybrid IFs model for forecasting health (using distal and proximate drivers) provides forecasts of age, sex and country-specific mortality rates for most of the 15 cause clusters it represents. The model within IFs for mortality in those cause clusters builds on the Global Burden of Disease (GBD) methodology, which uses mainly distal (more distant) drivers to project mortality.⁷ IFs then extends that methodology in many cases by adding attention to a selected set of proximate drivers.

For several other causes or cause clusters (*), most of those related to communicable disease, the IFs system uses a variant of the distal driver approach, one that looks to a forecast of all communicable disease except HIV/AIDS and then subdivides that total by more specific disease type.

For still other causes of mortality (**), it uses more specialized models totally unrelated to the distal driver approach.

Cause Clusters in IFs⁸

1. Other Group I diseases (excludes AIDS, diarrhea, malaria and respiratory infections)*
2. Malignant neoplasms
3. Cardiovascular diseases
4. Digestive diseases
5. Diabetes
6. Chronic respiratory diseases
7. Other Group II diseases (excludes malignant neoplasms, cardiovascular diseases, digestive diseases, diabetes, chronic respiratory diseases, and mental health)
8. Road traffic accidents**
9. Other unintentional injuries (excludes road traffic accidents)
10. Intentional injuries
11. HIV/AIDS**
12. Diarrhea*
13. Malaria*
14. Respiratory Infection *
15. Mental Health**

⁷ See Mathers and Loncar (2006) for a more complete description of GBD projection methods and rationale.

⁸ The first 10 causes in this ordering were specific ones from Groups 1, 2, and 3, associated with a residual for the remaining causes in each group—that is, the ordering was systematic. The subsequent addition of HIV/AIDS and then of diarrhea, malaria, respiratory infections, and mental health pulled additional specific causes from the various groups and caused the ordering of those additions and the entire list to be irregular with respect to groups. At some point, we will likely re-order the causes of death into the 3 groupings.

3.1 Distal Driver Formulation

For the basic forecast of mortality related to most of the major cause clusters (exceptions are deaths from HIV/AIDS and traffic accidents) we use the regression models and associated beta coefficients prepared for the Global Burden of Disease project (Mathers and Loncar 2006). Age, sex, cause, and country-specific mortality rate is a function of income, adult education, technological progress, and (in specific cases) smoking impact:

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

M is mortality rate in deaths per 100,000 for a given age category c, sex p, cause of death d and country or region r.

Y is GDP per capita at PPP

HC (human capital) is Years of Adult Education over 25

T is time

SI is Smoking Impact

Income and education (IFs variables GDPPCP and EDYRSAG25, respectively) are forecast endogenously in IFs. Time, a proxy for technological progress, is calculated as calendar year minus 1900 (for example, T for the year 2001 equals 101). Smoking impact, a variable meant to capture historical smoking patterns, is included only in the forecasts of mortality related to malignant neoplasms, cardiovascular disease, and respiratory disease.⁹ As described in section 6.3.4 of this document, IFs uses both historical smoking rate estimates and SI projections to 2030 (as provided by GBD authors) to forecast the SI variable.

Using an historical database representing mortality data from 106 countries for the years 1950-2002, the GBD calculated sex-specific regression coefficients for seven age groups (<5, 5-14, 15-29, 30-44, 45-59, 60-69, and 70+) and ten major cause clusters—the first ten in the list above (Protocol S1, 1-3).¹⁰ GBD estimations using the data from the 106 countries created separate low- and high-income regression models (not coefficients for each country separately), with low income defined as GDPPCP < \$3,000 in the initial year. Both sets of coefficients are publicly available online.¹¹ In IFs we spread the coefficients for the seven age groups across 5-year subcategories; that is, we use the same coefficients for each subcategory within the larger GBD ones—normalization of mortality within each 5-year subcategory across causes and to total mortality rates for each subcategory (taken from UN Population Division data) does, however, create differences in mortality rates across those 5-year groupings.

⁹ See Protocol S1, Mathers and Loncar 2006 for more detail on the use of smoking impact in GBD projections.

¹⁰ See Table 1, Protocol S1, for the cause clusters used in the GBD 2002 and 2004 projections. IFs does not use GBD coefficients for HIV/AIDS, relying instead on a structural model; mental health mortality is kept at a constant rate and coefficients for the other three communicable diseases (diarrhea, malaria, and respiratory infection) come from other sources.

¹¹ For regression results, see Tables S3 and S4 at <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0030442#s5>.

We generally use the beta coefficients provided by GBD authors to forecast mortality related to six cause groups: Group I excluding detailed communicable causes, malignant neoplasms, digestive diseases, Group II excluding diabetes and mental health, other intentional injuries, and intentional injuries. However, for a few age and Group III cause groups where regression models provided low predictive value, we also follow the GBD in keeping mortality rates constant over time instead of using the regression equations. Affected groups include: unintentional injuries for males older than 70; unintentional injuries for females older than 60; intentional injuries for males and females under 5; intentional injuries for males older than 60; and intentional injuries for females older than 45.

Although we forecast mortality by age, sex, cause, and country as in the general GBD equation above (and the details can be seen in the specialized displays of the model on mortality by age, sex, and cause and the mortality J-curve), the major model variable for display is DEATHCAT, which is total deaths by country/region, cause, and sex. The equation for it, using the IFs variables for GDP per capita at PPP (GDPPCP), average years of education for adults aged 25 and older (EDYRSAG25), time (IY%), and smoking impact (HLSMOKINGIMP) is

$$\begin{aligned}
 DEATHCAT_{r,d,p} &= F \left((Constant_{r,d,p,c} + \beta_1 * \ln GDPPCP_r + \beta_2 * \ln EDYSAG25_r + \beta_3 \right. \\
 &\quad * (\ln GDPPCP_r)^2 + \beta_4 * IY\% + \beta_5 * \ln HLSMOKINGIMP_{r,p}) * mortm_r \\
 &\quad \left. * hlmortm_{r,d} \right)
 \end{aligned}$$

The betas in the equation, as indicated earlier, are from the GBD work and are dimensioned also by country/region r (only as high income or low income), cause of death d, sex p, and age category c. The entire equation for mortality is adjusted in an algorithmic process so that the total across all causes of death equal the mortality rates from the UN Population Division's data (using a normalization factor), while the relative weights for each disease match WHO data (using a scaling factor). The normalization and scaling factors are multiplicative, affecting everything in the equation. In the Base Case scenario we keep those factors constant, but we can control convergence of them (see Section 5).

The equation allows scenario modification with multiplicative parameters that change mortality overall (*mortm*) or by cause of death (*hlmortm*). Not shown in the equation, *hlmortcdchldm* changes the rates of all communicable diseases for children aged 5 and younger, while *hlmortcdadltm* affects rates of death from communicable diseases for adults aged 15-49.

3.1.1 Forecasting income and education for the distal driver formulation

For the basic distal driver formulation, GDP per capita at PPP (GDPPCP) and years of adult education (EdYrsAg25) are forecast endogenously within IFs. GDP per capita is computed in the Economic Module as an annual flow variable (that is, it is generated anew each year), driven in part by underlying stocks such as capital supply and, in fact, years of adult education (both of which accrete or deplete very slowly over time). Adult

education is computed in the Education Module using government spending on education as the main driver.¹²

3.1.2 Forecasting technology for the distal driver formulation of mortality

Although two IFs variables, namely GDP per capita at purchasing power parity (GDPPCP) and years of adult education (EDYRSAG25) drive the distal formulation in most of our forecasting and scenario analysis, the technology parameter (the beta on time) in the distal equation is very powerful. We therefore want some control over it, ideally with ability to differentiate that control with respect to level of income of countries and with respect to the age structure of mortality. For a basic approach to providing such control, we follow the Global Burden of Disease (GBD) project in modifying the regression models for child mortality low-income countries. But we have extended that GBD approach to allow some additional parametric control.

The control system in IFs uses a switching parameter (*hlmortmodsw*) in interaction with three other parameters (with their default values those are *hltechbase*=1, *hltechlinc*=0.25, and *hltechssa*=0); see the table below for a summary of the application of those parameters.

In the default mode (*hlmortmodsw* = 1), IFs uses the GBD approach to modifying the technology (time) coefficients for children under 5 in recognition of slower than expected historical progress in many countries.¹³ Specifically, for children under 5 in low-income countries in four regions (Africa, Europe, SE Asia and West Pacific) the time variable is held constant (zero, or no technological advance, using *hltechssa*); in low-income countries in the Middle East and North Africa the coefficient on time is reduced to 25 percent of its original value using the parameter *hltechlinc*.

In the IFs implementation of the GBD approach to treatment of technology, we wished to change the patterns not just for children under 5, but also for older children and adults. We decided, however, to regularize the somewhat ad hoc assignment of countries by the GBD to the low-income category by defining low-income as being less than \$3,000 per capita and high-income as being above that level. We use the parameter *hltechlinc* to control technological change for older children and adults in low-income countries regardless of geographical region; at its default setting that parameter reduces the coefficient on time to 25 percent of its original value,

¹² For a more complete description of driver forecasting in IFs, please see www.ifs.du.edu.

¹³ The Global Burden of Disease (GBD) project made low-income modifications after recognizing that historical child mortality data did not match back projections of the model (Mathers and Loncar 2006b: 9). Note that the GBD approach to these modifications changed from the 2002 revision to the 2004 revision of the project. In the 2002 revision, the human capital (education) beta was reduced to half of its magnitude for sub-Saharan countries, and to 75 percent of its original magnitude for other low-income countries. This was done only if the beta on the human capital (education) term in the distal model formulation was negative (reducing mortality with increases in education). Technological advance factor (time) was left constant (no advance) for sub-Saharan Africa and reduced to 25 percent for other low income countries. The 2004 revision dropped the human capital modifications, but continued to reduce the coefficient on time.

	Age and Geographic Impact of the Parameters	
Base or default values	For Children under 5 (GBD Geographic Classification)	For older children and Adults (IFs Geographic Classification)
<i>hltechssa</i> =0	Low Income Countries in mostly 4 regions (Africa, Europe, SE Asia and West Pac; also selected countries such as Haiti)	Not used
<i>hltechlinc</i> = 0.25	Low Income Countries in the Middle East and North Africa	Low Income countries (GDPPCP < \$3k in 2010)
<i>hltechbase</i> = 1	All other countries, mostly High Income and including most of Latin America	High Income countries (GDPPCP >= \$3k in 2010)

For children older than 5 and adults in what IFs classifies as high-income countries (countries with GDP per capita at PPP in 2010 above \$3,000), IFs uses the parameter ***hltechbase***. Thus in the default situation, technological change is unchanged from the basic value.

These GBD technology factor modifications (and their extensions by IFs to adults and older children in our definition of low- and high-income countries) can be turned off in the model (***hlmortmodsw*** = 0).¹⁴ When the switch is turned on, adjustments can also be made to ***hltechbase***, ***hltechlinc***, and ***hltechssa*** to build new scenarios.¹⁵ It is important for the user to know, however, that regardless of age or income level of countries, the model uses ***hltechbase*** for mortality from cardiovascular causes (which uses a different regression model for forecasting).

Changes to ***hltechbase*** can also be adjusted by using a shift parameter called ***hltechshift*** (0 by default), which adjusts the technology factor depending on the level of initial GDP per capita at PPP.

$$hltechbaseadj = hltechbase + Min(40, GDPPPCi_r * hltechshift)/40$$

This adjustment increases the technology factor for high income countries more quickly than for middle or low-income countries

¹⁴ Although not recommended, IFs also allows the user to use the original GBD 2002 modifications (as described in a previous footnote) by specifying ***hlmortmodsw*** = 2.

¹⁵ Given the results found for Intentional Injuries, where mortality was reaching unrealistic levels, we have limited the changes to ***hltechbase*** and ***hltechlinc*** to at most 1.5 for this particular cause of death in the 2004 revision (***hlmortmodsw*** = 1).

3.2 Diabetes and Chronic Respiratory Diseases

Two chronic cause groups, diabetes and respiratory, are so strongly influenced by specific risk factors that estimates based on distal drivers alone fail to accurately represent expected mortality rate trajectories. In the case of diabetes, rising population levels of overweight and obesity contradict suggestions that diabetes-related mortality will fall over time in line with other Group II causes. Conversely, declining smoking rates in many high income countries may temper projections of increasing chronic respiratory-related mortality (Protocol S1, 5-6). Therefore, IFs follows the GBD methods in modifying the distal driver formulation by adding proximate risk factors (BMI and SI, respectively) to forecast base diabetes- and chronic respiratory-related mortality rates.

3.2.1 Diabetes

To forecast diabetes, IFs uses the following formula:

$$M_{r,d=Diabetes,c,p} = HLDIABETESRR_{r,c,p} * ONCD_{r,c,p}^{0.75}$$

$M_{r,c,d=Diabetes,p,r}$ is diabetes-related mortality by country/region r , age category c , and sex p . $ONCD_{r,c,p}$ is other Group II (non-communicable disease) mortality, derived using the basic distal driver equation. $HLDIABETESRR_{r,c,p}$ is a “Diabetes Relative Risk” factor, explained below.

In a population at the “theoretical minimum” level of body mass index (BMI), where BMI is 21, diabetes-related mortality is expected to fall at 75 percent of other Group II mortality.¹⁶ The diabetes relative risk factor ($HLDIABETESRR$) captures the increased risk represented by a population above the theoretical BMI minimum level. For example, the factor is about 1 for young females in Vietnam (where BMI is close to the theoretical minimum level of 21). Comparatively, the RR is approximately 28 for middle-aged women in the United Kingdom where population BMI is much higher.¹⁷

The GBD project projected the RR variable for diabetes out to 2030 using fairly involved estimates of age and sex-specific levels (plus standard deviations) of population BMI. Our estimates in IFs of future BMI ($HLBMI$) are less sophisticated, and we only forecast country/region (r) and sex-specific (p) mean BMI (see section 6.3.2 for a description of our forecasts of BMI). As such, while we endogenize the RR variable by tying it to our forecasts of BMI, we also adjust our forecast by initializing RR using the GBD estimates for the year 2010 and computing an age-category specific shift factor ($HLDIABSHIFT$) in order to tie our forecast of expected RR with GBD estimates.

¹⁶ The slower decrease in diabetes-related mortality reflects assumptions that risk factors for diabetes will improve more slowly than risk factors for other Group II diseases (Protocol S1, 6).

¹⁷ All RRs available in the IFs system, variable name $HLDIABETESRR$.

The RR for diabetes forecast in IFs (HLDIABETESRR) assumes that country-specific BMI is distributed normally, and also assumes a standard deviation of 10% of the mean:¹⁸

$$\begin{aligned}
 &HLDIABETESRR_{r,c,p} \\
 &= HLDIABSHIFT_{r,c,p} \\
 &\quad * \int e^{\left(\left(HLBMI_{r,p} - \left(\frac{HLBMI_{r,p} - avgBMI}{StdDev} + 21 \right) \right) * LogRR_{c,p} \right)} * P(HLBMI) * dBMI
 \end{aligned}$$

LogRR is the change in log of RR per 1 unit change in BMI.¹⁹ These values are age category (c) and sex (p) specific; the absolute relative risk of diabetes-related mortality in relation to a unit increase in BMI varies from between 1.47 (females under 45) and 1.2 (females over 80).²⁰ P(HLBMI) is a normal distribution function with mean of avgBMI; StdDev is a fixed 10% of avgBMI.²¹

3.2.2 Chronic Respiratory Disease

Again following GBD authors, IFs separately computes the two components of the chronic respiratory disease category—chronic obstructive pulmonary disease (COPD) (where smoking is the overwhelming related risk factor) and “other” respiratory disease (where smoking is somewhat less determinative). Both elements follow the same formulation:

$$\begin{aligned}
 &M_{r,d=chronic\ respiratory\ disease,c,p} \\
 &= \ln \left((SIR_{r,c,p} * CRDRR_{c,p,d} + 1 - SIR_{r,c,p}) \right) \\
 &\quad * (e^{ONCD_MORT_{d=Other\ Group\ II}})^{0.75}
 \end{aligned}$$

where

$$SIR_{r,c,p} = \frac{HLSMOKINGIMP_{r,c,p}}{SMOKIMPADJ_{r=three\ regions,c,p}}$$

SIR is the “smoking impact ratio,” that is smoking impact in IFs (HLSMOKINGIMP) divided by an adjustment factor that is specific to three big regions (1. China, 2. World, and 3. SearD (Bangladesh, Bhutan, India, North Korea, Maldives, Myanmar, Nepal,

¹⁸ We recognize, of course, that BMI is most likely not distributed normally in a population. However, we follow CRA authors in assuming normality in order to compare a given population with an ideal counterfactual population (James et al 2004).

¹⁹ WHO Comparative Risk Assessment Methodology, Kelly et al, 2009

²⁰ See associated data table, Kelly et al 2009.

²¹ avgBMI is our forecast of BMI, while BMI are the values from -3 standard deviations to +3 standard deviations away from that avgBMI. Cecilia Peterson determined the fixed 10 percent rate for StdDev from the literature.

Afghanistan, Pakistan)),²² age category c, gender p. CRDRR is the relative risk for chronic respiratory disease specific to age category c, gender p, and disease/death d type (COPD or other respiratory disease).²³ Chronic respiratory disease is assumed to be declining at 75 percent of ONCD_Mort, which is other Group II related mortality.

3.3 Diarrhea, Malaria, and Respiratory Infections

IFs added three additional communicable diseases, namely diarrhea, malaria, and respiratory infection, after it had developed the modelling approach for distal drivers discussed above. The model uses the more general Group I (communicable disease and maternal mortality excluding AIDS) forecast to project mortality related to all three additions:

$$\ln(M_{r,c,p,dg=1,d}) = C_{r,c,p,dg=1,d} + \beta_{r,c,p,dg=1,d} * \ln M_{r,c,p,dg=1}$$

M is mortality rate in deaths per 100,000 for a given region r, age category c, sex p, general (Group 1) cause dg=1 and specific disease d within general cause group dg=1. Here dg=1 refers to all communicable diseases (other than HIV/AIDS because Mathers and Loncar 2006 did not use the same approach to that particular communicable disease) and d refers to diarrhea, malaria or respiratory infections.

The constants and beta coefficients for the above equation for the three diseases come from Mathers and Loncar (2005 and 2006c).²⁴ For diarrhea and malaria we used their coefficients for infectious and parasitic diseases (Mathers and Loncar 2005: Table A-6 on page 115); for respiratory infections we used their coefficients for respiratory infections specifically (Mathers and Loncar 2006c, Table S5).

The results for these three subtypes are then subtracted from the mortality for the total Group I category (except HIV/AIDS). The reason for this is to make sure that the sum of all them does not exceed the total of Group 1 (excluding HIV/AIDS), which is a result the equation could theoretically produce. In fact, we want to be sure that there is room for the Other Group 1 cause of death, so IFs limits the sum for diarrhea, malaria, and respiratory infection to 95 percent of the total of Group 1 (excluding HIV/AIDS). If necessary, all three subcategories are reduced proportionally by a factor of $0.95 / (\text{SUM}(3 \text{ subtypes}) / \text{Tot}(\text{big type}))$. Note that, if this restraint needs to be imposed, the denominator

²² GBD authors provided the adjustment factor for SIR, and it is constant over the length of the IFs forecast. The computation is done with the hard-code value in a procedure called UpdateRespDisease. The name for SmokingImpactAdj in the model is sird. CRDRR is hard-code in the same procedure.

²³ RR ranges from approximately 10 for COPD to about 2 for other chronic causes. Again, GBD authors provided the relative risk estimates used in IFs.

²⁴ The extended process for using those is described in 2 working notes for the IFs project by Dale Rothman, Dealing with Diarrhoeal Diseases Including the Effects of Unsafe Water & Sanitation and Undernutrition (March 25, 2009) and Dealing with Effects of Indoor Air Pollution (October 8 2009). Titles of the files are Incorporating Diarrhoea 25 March 2009 and Incorporating Indoor Air Pollution 9 October 2009, respectively.

will always be higher than 0.95 and then the multiplicative adjustment factor will always be lower than 1.

3.4 HIV/AIDS

The ultimate objective of the calculations around HIV infections and AIDS is to forecast annual deaths from AIDS (AIDSDTHS) by age category and sex. We did not look to the forecast methodology of Mathers and Loncar (2006) for their approach on this particular communicable disease; in fact, they also used an approach that did not rely upon the general distal driver formulation.

The IFs approach begins by forecasting country-specific values for the HIV prevalence rate (HIVRATE). 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 sub-groups.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

²⁵ The HIV/AIDS data were being update in October, 2013. 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 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 the age-category level.

²⁷ Table 17 (pp 77-78) of the Annex to World Population Prospects: the 2002 Revision (UNPD 2003) 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-

$$HIVRATE_{r,t} = HIVRATE_{r,t-1} * (1 + hivincr_{r,t}) * hivm_r$$

where

$$hivincr_{r,t} = F(\mathbf{hivincr}_{r,t=1}, \mathbf{hivpeakyr}_r, \mathbf{hivpeakr}_r)$$

t is time, r is country or region. 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 a certain percent 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

$$HIVTECCNTL_{r,t} = HIVTDCCNTL_{r,t-1} * \left(1 + \mathbf{hivtadvr} * \frac{t}{100}\right) + F(\mathbf{hivincr}_{r,t=1})$$

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

$$HIVCASES_{r,t} = Pop_{r,t} * HIVRATE_{r,t}$$

The rate of death to those with HIV would benefit from a complex model in itself, because it varies by 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 the manifestation of AIDS and deaths from it both 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:

$$AIDSDRATE_{r,t}$$

$$= HIVRateMAvg_{r,t} * HIVAIDSR_{r,t=1} * \left(1 - \frac{\mathbf{aidsdrtadvr}_{r,t}}{100}\right) * \mathbf{aidsratem}_{r,t}$$

where

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.

$$HIVRateMAvg_{r,t} = F(HIVRATE_{r,t}, \text{last 10 years})$$

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 (*aidsdeathsbyage*) from those data.

3.5 Cardiovascular Disease

The regression models used in the GBD project did not differentiate between-subject from within-subject variation. Particularly for cardiovascular-related outcomes in some age/sex groups, this model produced a perverse finding: a negative relationship between cardiovascular-related mortality and smoking impact (HLSMOKINGIMP). However our further statistical investigation showed, as expected, a positive relationship between cardiovascular-related mortality and HLSMOKINGIMP within a given country over time.

As such, we completed a more sophisticated mixed model regression analysis (using SAS, version 9.1) to capture both within and between-subject effects. We used the GBD mortality database described in section 5, supplemented by our historical series of income per capita²⁹. All distal drivers were included as fixed effects, with random effects included for subject (country) and time (T). The revised coefficients (see Appendix Table 1) were used to forecast cardiovascular disease-related mortality. We created only one model for all countries (no separate low-income model) due to lack of data. Comparison with the original GBD models reveals fairly similar forecast outcomes overall. However, the positive change in the smoking/cardiovascular mortality relationship allows us to better examine how smoking intervention scenarios might impact cardiovascular-related mortality.

3.6 Road Traffic Accidents

In forecasting mortality related to road traffic accidents, IFs replaces the GBD regression model with a structural formulation designed to better capture relevant drivers for this cause group. Specifically, IFs projects deaths due to traffic accidents (DEATHCAT, Traffic) as a function of deaths in traffic per vehicle (DEATHTRPV) and vehicle numbers (VEHICLESTOT), both computed in the automobile module of IFs. We first need to compute the total size of the vehicle fleet.

Total vehicles per capita (VEHICFLPC) is based on a formula proposed in a paper by Dargay et al (2007) in which fleet size per capita is a function of GDP per capita at PPP

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

²⁹ Note that we did use historical estimates of education provided by the GBD project, instead of using the less complete historical series available through IFs. Future distal driver analysis may explore using alternate sets of education data, including those included in the IFs system.

(GDPPCP). Translating the Dargay et al (2007) equation into one using IFs variable names yields:³⁰

$$VEHICFLPC_r = (852 - RF) * e^{(-5.987 * e^{(-0.2 * GDPPCP_r)})} * \mathbf{vehicfpcm}$$

The parameter **vehicfpcm** allows scenario intervention. RF is an adjustment factor that compensates for different land densities, that is the ratio of population (POP) to land area (LANDAREA), taking the U.S. as the base:

$$RF = 38.8 * \left(\frac{POP_r}{LANDAREA_r} - \frac{POP_{r=USA}}{LANDAREA_{r=USA}} \right)$$

The computation was only used when country R had higher density than the US. The paper also describes another adjustment factor related to urbanization as percentage of total population, but we did not use this additional adjustment factor in our model.

Given fleet size per capita and the population, we compute the total size of the fleet.

$$VEHICLESTOT_r = VEHICFLPC_r * POP_r$$

The number of deaths per vehicle is based on Smeed's Law³¹, an empirical rule originally proposed by R.J. Smeed, which relates deaths to vehicle ownership. In the original conceptual form Smeed's Law is:

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

D is annual road deaths
n is number of vehicles
p is population

In terms of IFs variable names this would translate literally (ignoring some unit issues) as:

$$DEATHCAT_{r,d=TrafficAcc} = 0.0003 * (VEHICLESTOT_r * POP_r^2)^{1/3}$$

The actual representation in IFs involves two steps. First we calculate the death rates per vehicle, adding a division by a multiplicative term that is equivalent to total vehicle numbers VEHICLESTOT. One of the virtues of this first step is that we can add an exogenous multiplier for death rates per vehicle, **deathtrpvm**.

$$DEATHTRPV_r = \frac{0.0003 * (VEHICFLPC_r * 1E + 15 * POP_r^3)^{1/3}}{VEHICFLPC_r * POP_r} * \mathbf{deathtrpvm}$$

³⁰ Dargay, Gately, and Sommer 2007. "Vehicle Ownership and Income Growth, Worldwide: 1960-2030". Joyce Dargay, Dermot Gately and Martin Sommer, January 2007.

³¹ http://en.wikipedia.org/wiki/Smeed%27s_law

Smeed, RJ 1949. "Some statistical aspects of road safety research". *Royal Statistical Society*, Journal (A) CXII (Part I, series 4). 1-24.

Adams 1987. "[Smeed's Law: some further thoughts.](#)" *Traffic Engineering and Control* (Feb) 70-73

The second step is to use the death rate per vehicle, the vehicle fleet size per capita, and information on the age and sex distribution of deaths from vehicles to compute the mortality rate from vehicle accidents by age and sex, putting the results into a variable internal to model named **modmordstdet**. In a third step that variable is used with population by age and sex to compute the total deaths from vehicle accidents (DEATHCAT). These second and third steps stylistically yield

$$DEATHCAT_{r,d=TrafficAcc} = DEATHTRPV_r * VEHICLESTOT_r$$

After initialization in the base year (using GBD estimates of road traffic-related mortality and total vehicles from the automobile module in IFs), IFs calculates a multiplicative shift factor that is kept constant for the entire forecast horizon. If this initialization value is greater than 40 deaths per 1000 vehicles, we adjust the number of vehicles per capita to set 40 as our initialization value. We started using this limit after finding inconsistencies between estimates derived from Smeed's Law and those from initial estimates.³²

IFs also computes a ratio (in a variable internal to the model) of traffic accident mortality for males compared to females. The model converges that ratio to 1.5 over 100 years by preserving the total mortality for each age category but adjusting the distribution between males and females.

3.7 Mental Health

The IFs model assumes that the initial rate of mortality related to mental health remains constant across our forecast horizon. That rate is subtracted from the other Group II category. The scenario parameter **hlmortm** allows the user easy control over mortality from the cause.

³² The case of Bangladesh is illustrative, where the forecast calculation of 141 deaths/thousand vehicles contrasts with an expectation of 30 deaths/thousand vehicles using Smeed's Law. We concluded that our mortality figures were consistent with WHO estimates, but sometimes the total number of vehicles was too low. For example, for Bangladesh our data showed 1 vehicle per thousand people, which meant about 141,000 vehicles, when several reports indicate the real number is much higher (850,000) (<http://www.brta.gov.bd/pdf/Statistics%202005.pdf>).

4. Modifications to the Basic Health Model

After examining in IFs the long-term behavior of the regression model forecasts using the Global Burden of Disease (GBD) distal driver approach and coefficients, we made a limited number of modifications. One set of modifications was made to the treatment of the technological change term in the distal formulation (see Section 3.1.2). We also allow countries to transition from low-income status, given expected improvements in development status over the long-term. An adjustment for monotonicity ensures that the forecast population comports with well-known patterns of rising chronic-cause mortality rates with age. Finally, we include health spending in our model in order to better forecast potential outcomes.

4.1 Mortality Transition for Low-Income Countries

As described in the discussion of distal driver coefficients for low-income countries we use Global Burden of Disease (GBD) regression coefficients developed separately for low and high income countries. However, given the long forecast horizon of IFs, we recognize that many low-income countries eventually will reach high levels of income and thus should follow a similar pattern of mortality. Therefore, we allow low-income countries to transition gradually by computing two mortality rates for low-income countries—one using the low-income beta coefficients and the other using the high-income model. We start the transition when countries reach GDPPCP of \$3,000, and finish the transition when countries reach \$15,000. The transition is computed finding target mortality in between the two, interpolating depending on the current level of GDPPCP

Given the target mortality, we compute how much change we need from current mortality (low-income based), and slowly adjust using a moving average of 20 percent of current required change and 80 percent of change used in previous years:

$$\begin{aligned} \text{Change} &= \text{Target Mortality} - \text{Low Income Mortality} \\ \text{Smooth Change} &= 0.2 * \text{Change} + 0.8 * \text{Last Year Change} \\ \text{Final Mortality} &= \text{Low Income Mortality} + \text{Maximum}(\text{Smooth Change}, \text{Change}) \\ \text{where Last Year Change} &= \text{Maximum}(\text{Smooth Change}(\text{yr-1}), \text{Change}(\text{yr-1})). \end{aligned}$$

Note that most of the time target mortality is lower than low income mortality, and thus change is negative. Thus, when we find the maximum we are finding the smaller absolute number and smoothing change.

4.2 Maintaining Increase with Age in Non-Communicable Death Rates

In general, both in our initial conditions and forecasts, we try to maintain monotonicity in growth of death rates from chronic causes with increasing age (above 45) by adjusting deaths from a particular cause when initial computations do not illustrate increases with age and compensating in death rates from an alternative cause for which initial computations indicate room for mortality reduction while (a) maintaining monotonicity for that cause also and (b) not changing total mortality in the pair of causes.³³ If necessary

³³ The changes described here for monotonicity with age do not guarantee monotonicity of changes in rates over time within an age category. In fact, they could contribute to some small transients or irregularities in

we also work to make acceptable adjustments in one age category by readjusting the next age category for the same cause of death, decreasing deaths in the younger age category and increasing deaths in the older one, while doing the opposite for the compensating cause of death. Finally, in this overall and quite complicated algorithmic process we try to minimize the adjustments made to the initial calculations of mortality.

To elaborate this process further, when we find for a chronic cause of death a monotonicity problem for a country and a given age category relative to the next younger category, we find the type (H1) with the highest mortality rate (in the 100+ category among non communicable disease),³⁴ then we try to use H1, which often turns out to be cardiovascular disease, to compensate adjustments in other types in order to keep total mortality constant for the same age category.

For each 5-year age category starting at 45 to 49, we compute total mortality as the sum of all types, then for each non-communicable type with non-zero mortality we compute its growth G from the current age category j to the next $j+1$, for example in the first step from 45-49 to 50-54. Although our emphasis is on avoiding non-monotonicity, we also would like to see some regularity of progression of mortality increase with age, as we find in the quite high-quality data of Sweden. Thus we also look to that progression in Swedish data for a rate of increase of across age categories that we can use as a minimum. Specifically, across two adjacent age categories we find a Proxy growth P , where we use Sweden's mortality for each type, but we do not allow this P to be higher than $1/4^{\text{th}}$ of Total Mortality Growth. If G is smaller than P then we start the procedure for the given age category j and type of mortality d .³⁵

Once we start the adjustment procedure we check if there is room to reduce mortality in the current age and type, so we check growth from the previous age category to avoid breaking monotonicity. First we compute Proxy Growth $P1$ from the previous age category $j-1$ (40-44 for our example) to the current one j (45-49).

Second we compute the minimum acceptable value for current mortality:

$$Min = Mort_{r,a=j-1,d} * (1 + P1)$$

rates over time. In general, however, we believe that they will make such behavior less likely. For such irregularities, see the longitudinal curve in Afghanistan for cancer at 80-84 males; these are most likely to appear, as they would in the real world, when total mortality is not changing much over time.

³⁴ In specialized work looking at low senescent aging the last category is 200+.

³⁵An example can illustrate. Say male mortality is 0.5 for Cancer at 45-49, and 1.1 for 50-54, then P is 140% (these are numbers from Sweden). Say total male mortality is 10.1 for 45-49 and 14.5 for 50-54, so total growth is 43%. (These are numbers for any country, say Afghanistan). We can't use P of 140% on Afghanistan for male cancer, given that for those ages Total Mortality grows only 43%, so we use a Proxy (P) of 11% ($43 * .25$), for male cancer in Afghanistan between ages 45-49 and 50-59. G is the actual growth in mortality for male cancer in Afghanistan between ages of 45-49 and 50-54; say for example: 0.7664, 1.2876 respectively, so G is 68%. In this example, there is no need to change anything ($68 > 11$). Only if G is smaller than P does the adjustment process begin.

where $Mort_{r,a=j-1,d}$ is the mortality for country r, type d in age category j-1.

Third we compute maximum acceptable value for current mortality, we start with:

$$Max = \frac{Mort_{r,a=j+1,d}}{1 + P}$$

But we know that $Mort_{r,a=j-1,d}$ is also going to change to keep the number of deaths constant, so we also consider this adjustment:

$$Max = \frac{Mort_{r,a=j+1,d} + Adj * \frac{Pop_{r,a=j}}{Pop_{r,a=j+1}}}{(1 + P)}$$

And we know that:

$$Adj = Mort_{r,a=j,d} - Max$$

Solving for max, we have:

$$Max = \frac{\left(Mort_{r,a=j+1,d} + Mort_{r,a=j,d} * \frac{Pop_{r,a=j}}{Pop_{r,a=j+1}} \right)}{\left(1 + P + \frac{Pop_{r,a=j}}{Pop_{r,a=j+1}} \right)}$$

Where Mort is the original mortality for age category j and j+1, country r, and type d. Pop is population for age j and country r and P is the Proxy Growth computed as explained above.

If Min is smaller than Max, then we use Max as the new mortality in age j, in order to keep the adjustment as small as possible, if not that means that Max wouldn't keep monotonicity from age j-1, so we start trying to adjust going backwards, given that frequently there's more room in previous age categories. In order to start going backwards we keep track of the first age category that it's already saturated, i.e. that its growth is already the minimum possible without breaking monotonicity. If we find that the first saturated category is higher than the 45-49 that we started with, that means we have some room going backwards, so we take Max, otherwise we use Min as the new mortality in age j, and keep adjusting forward. The adjustment A is just the difference between original mortality in age j and the new chosen mortality.

Adjust backwards means that we will adjust mortality in age category j2 and j2-1, where j2 goes from j-1 to 10 (which corresponds to 45-49). While going backwards the formulas for min and max change a little bit, given that the adjustment is done in the previous age category.

$$Min = \left(Mort_{r,a=j2-1,d} + Adj * Pop_{a=j2} \right) * (1 + P1)$$

And substituting the adjustment computed earlier ($Adj = Mort_{r,a=j,d} - Max$), we end up with:

$$Min = \frac{\left(Mort_{r,a=j2-1,d} + Mort_{r,a=j2,d} * \frac{Pop_{a=j2}}{Pop_{a=j2-1}} \right)}{\frac{1}{1+P} + \frac{Pop_{a=j2}}{Pop_{a=j2-1}}}$$

Max gets simplified to:

$$Max = \left(\frac{Mort_{r,a=j2+1,d}}{(1+P)} \right)$$

We then check for room in type H1, and if there's enough room we adjust mortality for j2. If $Min \leq Max$ then we can stop, otherwise we keep going back until we reach the 45-49 category.

Fourth, we verify that, in doing compensation for type H, monotonicity is preserved too. In order to make this verification first we find the potential growth rate GH after applying adjustment A to type H. Then we compute the Proxy Growth PH for type H. If GH is greater or equal than PH then we can apply the adjustments if not we just leave mortality unchanged.

Fifth, applying the adjustments to type d by subtracting the adjustment from the original mortality in age j for type d, and adding it up adjusted for deaths to the original mortality in age j+1 for type d:

$$Mort_{r,a=j,d} = Mort_{r,a=j,d} - Adj$$

$$Mort_{r,a=j+1,d} = Mort_{r,a=j+1,d} + Adj * \frac{Pop_{r,a=j}}{Pop_{r,a=j+1}}$$

Sixth, applying the adjustments to type H by adding the adjustment to original mortality in age j for type H, and subtracting it adjusted for deaths from the original mortality in age j+1 for type H:

$$Mort_{r,a=j,d=H} = Mort_{r,a=j,d=H} + Adj$$

$$Mort_{r,a=j+1,d=H} = Mort_{r,a=j+1,d=H} - Adj * \frac{Pop_{r,a=j}}{Pop_{r,a=j+1}}$$

Seventh, if Min is greater than Max and we couldn't go backwards means that we took Min as the new mortality for age k, and it means that we still don't have monotonicity because we haven't changed age j+1 yet. Then we need to find the new mortality value for j+1 using Proxy Growth:

$$Newmort = Mort_{r,a=j,d} * (1+P)$$

$$Adj = Newmort - Mort_{r,a=j+1,d}$$

Eighth, we check that this new adjustment doesn't break monotonicity in death type H, if it doesn't we apply it as we did for age j, if does break it, we just leave mortality unchanged.

Ninth, applying this adjustment is the same as step 5 and 6, but using ages j+1 and j+2 instead of j and j+1. The only difference here is that when we get to the second to last age category (j = 20, 95-99),³⁶ then the compensatory adjustment for deaths is done in the first age category of the loop (j=10, 45 to 49), and we restart the process for a second and final check of monotonicity.

We have added check limits along the process to avoid mortality going above 1000 per 1000 and below 0 per 1000 at all times, and if the limits are reached then mortality is left unchanged.

4.3 Elasticity of Child Mortality with Health Spending

For countries that have a GDP per capita in the initial year of less than \$15,000 an elasticity factor with health spending (*elhmortspn*) of -0.06 will affect mortality of children under 5. That is, each 1 percent change in health spending as a percentage of GDP will lower mortality by 0.06 percent; an increase of 100 percent (doubling) would produce an automatic reduction of 6 percent in mortality. We have implemented a limit on the reductions to be at most 80 percent of mortality.

The GBD project's distal driver formulation does not take public health spending into account. However, we add a term to the basic GBD distal driver formulation to incorporate public health spending as a proximate driver to account for the relatively consistent inverse relationship between total public health expenditures and child mortality rates in poor countries (Anand and Ravallion 1993; Bidani and Ravallion 1997; Jamison et al. 1996; 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 (except HIV/AIDS) for the age 0-4 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

$$MortAdj_{r,j=0-4,d=1,t} = Mort_{r,j=0-4,d=1,t} * (1 + HExpFct_{r,t})$$

where

$$HExpFct_{r,t} = elhmortspn * \frac{\left(100 * \frac{GDS_{r,g=health,t}}{GDP_{r,t}}\right) - GDSHI_{r,t=1}}{GDSHI_{r,t=1}}$$

³⁶ In specialized work looking at low senescent aging the last category is 195-199.

where

$$GDSHI_{r,t=1} = \frac{GDS_{r,g=health,t=1}}{GDP_{r,t=1}} * 100$$

Elhmortspn = -0.06

where

GDS is government expenditure; elhmortspn is the elasticity of mortality with health spending, *j* is age category; *r* is country/region; *d* is cause (1 is other 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.

5. Data Initialization

Several important differences in our approach to forecasting health relative to that of the Global Burden of Disease (GBD) project required development of algorithms for computation of initial conditions and small multiplicative adjustments to formulations. Specifically, we forecast by country, we begin forecasts in the base year of 2010, and we maintain 5-year age categories. Moreover, we have our own sources of data for GDP per capita and education attainment level, which we forecast using our own models. The initial data we obtained from the GBD project provided country, sex, and cause specific mortality, but from the year 2008 (subsequently updated to 2010) and in slightly different age categories.³⁷ This section details our approach to reconciling differences in initial data.

5.1 Normalization and Scaling Factors

IFs initializes the base year (2010) data using age, sex, cause, and country-specific mortality data for 2010 provided by the Global Burden of Disease (GBD), courtesy of Colin Mathers at the World Health Organization. Those data are for infants and then for 5-year age categories up to 85+. To fill holes we used the same base rates for 1-4 as for infants and for all 5-year age categories above 85, subject to normalization to total mortality for the age category. The model then computes normalization and scaling factors which reconcile the results of the forecast regression models with these initial data.

The normalization factor helps match the sum of all mortalities in the health module to the mortality computed in the population module in the base year (2010). This process assures that we have initial conditions consistent with UNPD mortality data in our base year (i.e., the sum of all deaths will be the same as the UNPD mortality data for each 5-year age and sex category for the year 2010). The normalization factor uses all types of mortality except for AIDS in the numerator and denominator:

$$\text{Normalization Factor} = \frac{\text{Total Mortality data} - \text{Aids mortality data}}{\sum \text{Mortality data}_i}$$

The scaling factor sets the historic proportions across the different causes of mortality, assuring consistency of total deaths forecast using the GBD formulations and our 2010 values of driving variables with the cause-specific mortality data in the GBD's detailed death file. The scaling factor uses distal driver regression results for the denominator and GBD 2010 data for the numerator:

$$\text{Scaling Factor} = \frac{\text{Mortality data}_i}{\text{Mortality calculation}_i}$$

³⁷ Initial data from the GBD does not separate infants from under 5 mortality and puts together mortality for people over 85.

These adjustments mean that, except for the total mortality by age and sex of the UN, our numbers in the 2010 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.³⁸ In the forecasts themselves, we keep the multiplicative scaling and normalization parameters constant over time because there is no clear reason for changing them in the base, but have added parameters to control convergence in scenarios: *hlgbdconvdown*, *hlgbdconvup*, *hlscaleconvdown*, where the first parameter controls the normalization factor when is greater than 1, the second one controls the normalization factor when is smaller than 1, and the last one controls the scaling factor when is greater than 1.

5.2 Cause-specific Mortality: Infants and the Elderly

The detailed deaths data file (by cause, sex, and age) that we have obtained through the generosity of Colin Mathers at the World Health Organization does not include cause-specific infant or old-age (85+) mortality (greater than 85). Because IFs forecasts both infant mortality and 5-year age categories to 100 years, we incorporate detailed mortality data from Sweden (as a proxy, thanks mainly to availability of data) in order to initialize Group II (excluding mental health) cause-specific mortality for these missing populations.³⁹

The first step is to find the weights per age category for Sweden as follows:

$$Weight_{a=j,d,p} = \frac{Deaths_{r=Sweden,a=j,p}}{Deaths_{r=Sweden,a=JJ,p}}$$

Where *j* is the smaller IFs age category *a* (for example infants), *JJ* is the bigger corresponding GBD age category (for example children < 5, which implies the addition of infants plus children 1-4), *p* is gender, and *d* is mortality type.

The second step is to check the monotonicity of growth in the existing mortality data for each country and type of mortality (from age 45 forward). If monotonicity is not found (i.e., mortality rates do not rise in step with increasing age groups) then the initialization data is left unchanged for this country and mortality type combination. If initial mortality does increase monotonically, we further adjust mortality:

$$Mortality_{r,a=j,d,p} = Mortality_{r,a=j,d,p} * Weight_{a=j,d,p} * \frac{Pop_{r,a=JJ,p}}{Pop_{r,a=JJ,p}}$$

³⁸ Complicating initialization further, the UNPD presents its data in 5-year ranges, including 2005-2010 and 2010-2015. The age- and sex-specific survivor-table values in those ranges therefore do not correspond to specific years like our base of 2010. After correspondence with Kirill Andreev of the UNPD, which we acknowledge appreciatively, we decided to average the mortality values in the two 5-year ranges ending and beginning with 2010.

³⁹ This may be an issue also for Groups 1 and 3, but we have used the procedure with Sweden only for Group 2.

Where j , JJ , p and d are the same as in the previous equation and Pop is the population vector.

This option has been currently disabled because it was producing too much NCDs compared to CDs for countries like Mali for people over 90 years of age.⁴⁰

⁴⁰ Beginning with discovery of a problem for Mali, the use of this distribution procedure in IFs ran into some logic problems with poor behavior. As of September, 2013, the spread of initialization data for infants 1-4 and for ages 85+ is disabled. That spread, although desirable, is not necessary and at this point the code generates more problems than it solves. This section of the documentation is maintained should we want to revisit the issue.

6. Incorporating Proximate Drivers

Although the distal driver approach serves us well in forecasting health, thinking about intervention and leverage points in order to achieve alternate health futures necessarily involves the inclusion of proximate health drivers into the model. Therefore we need to have an approach that can layer specific risk analysis onto the underlying distal driver specifications. We need to discuss our general approach to doing that, which uses population attributable fraction or PAF formulations, and we need to provide equations of the models used to forecast proximate drivers in IFs and to build the relative risk estimates we used in the model.⁴¹

6.1 Adjusting mortality due to changes in a single risk factor

We build our approach on an understanding of two basic concepts used in the Comparative Risk Assessment (CRA) project (Ezzati et al. 2004), specifically *relative risk* (RR) and *population attributable fraction* (PAF). An RR 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”.⁴² 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, the WHO estimates that children under five who are moderately or severely underweight are almost nine times more likely to die of communicable causes than a population of “normal-weight” children (Blössner and de Onis 2005).

As its name suggests, a PAF or population attributable fraction 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 (Ezzati 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; and

P(x) is the population distribution in terms of exposure level, i.e. 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, the WHO defines an international reference population

⁴¹ See Chapters 5 and 6 of Hughes et al (2011) for a detailed discussion around the inclusion of proximate drivers into IFs.

⁴² Dictionary of Cancer terms, National Cancer Institute; accessed online, January 2010. <http://www.cancer.gov/dictionary/>.

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,⁴³ 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 such as undernutrition, obesity, smoking, and indoor air pollution.

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 (representing an “expected” value for a country given those distal drivers), and (2) estimates based upon a more complete set of drivers (representing our best forecast for a country using initial conditions and therefore path dependency, the additional and/or alternative drivers, and potential scenario interventions). We therefore calculate two versions of the PAF, namely PAF_{Distal} and PAF_{Full} . 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:

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

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 relative risk based on the full formulation implies that the distal drivers overestimate our anticipated improvement (or

⁴³ 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.

underestimate the deterioration) in the risk factor. Thus, the mortality forecast needs to be adjusted upwards. Alternatively, if the weighted-average relative risk is lower based on the full formulation than on the distal formulation, the mortality forecast will be adjusted downwards. Note that this property of the calculation actually obviates the need to know the theoretical minimum population.

6.2 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 and others 2004):

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

where

PAF^i is the PAF for risk factor i

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

Say that we have 2 risk factors:⁴⁴

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

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

$$\begin{aligned} ((1-PAF^{combined}_{Distal}) / (1-PAF^{combined}_{Full})) &= [(1-PAF^1_{Distal})(1-PAF^2_{Distal})] / [(1-PAF^1_{Full})(1-PAF^2_{Full})] \\ &= [(1-PAF^1_{Distal})/(1-PAF^1_{Full})] * [(1-PAF^2_{Distal})/(1-PAF^2_{Full})] \\ &= [\sum RR^1(x)P^1_{Full}(x)/\sum RR^1(x)P^1_{Distal}(x)] * [\sum RR^2(x)P^2_{Full}(x)/\sum RR^2(x)P^2_{Distal}(x)] \end{aligned}$$

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

⁴⁴ In the sequence of our calculations we decompose this equation in practice by finding the individual PAFs, computing their individual independent effects with $(1-PAF_{Distal})/(1-PAF_{Full})$, and multiplying mortality independently and cumulatively.

6.3 Specific Risk Factors

For each of the proximate risk factors used in the IFs model, we develop “full” model formulations (our best forecasts using IFs variables) and “distal” models (using just income and education) in order to compute the two PAFs described above. The IFs system includes full formulations and proximate risk analysis for selected risk categories: childhood undernutrition, child severe acute malnutrition (SAM), adult body mass index and related obesity, unsafe water and sanitation, indoor air pollution associated with solid fuel use, outdoor urban air pollution, and smoking.

6.3.1 Childhood Undernutrition

The population level of childhood undernutrition impacts IFs forecasts of under-5 mortality related to communicable causes. The “full” IFs forecast is based on estimated calories per capita⁴⁵ and access to safe water/sanitation:

$$z = 1.9212672 - 0.00069 * CLPC_r - 0.011589 \\ * (WATSAFE_{r,ss=2} + WATSAFE_{r,ss=3}) - 0.020278 \\ * (SANITATION_{r,ss=2} + SANITATION_{r,ss=3})$$

$$MALNCHP_r = 100 * \frac{e^z}{1 + e^z}$$

Where:

CLPC is calories per capita

WATSAFE(R, 2) is improved access to water

WATSAFE(R, 3) is piped access to water

SANITATION(R, 2) is shared access to sanitation

SANITATION(R, 3) is improved access to sanitation

MALNCHP is percent of children malnourished.

For each country/region r

Parameters for the distal regression were estimated in a mixed model regression analysis (using Proc Mixed in SAS Version 9.1) from historical data (1960-2005):

$$z = -0.714297 - 0.58979 * LN(GDPPCP_r) - 0.544938 * LN(EDYRSAG25_r)$$

$$7 \quad PMN = 100 * \frac{e^z}{1 + e^z}$$

8

⁴⁵ Calories per capita are calculated through the agricultural module in IFs. The number of available calories depends strongly on the interaction of two factors: income (including its distribution) and food price. Long-term trends in caloric availability reflect fairly rapidly-rising incomes in most parts of the world.

In the above equation GDPPCP is GDP per capita at purchasing power parity and EDYRSAG25 is average years of formal education for adults over 25. PMN is the percentage of children categorized as “moderately” or “severely” undernourished (≤ -2 standard deviations below the international standard of weight for age). Both the distal and full models incorporate an additive shift factor to match initial year (2010) model estimates to historical data. This additive shift factor converges to 0 in 100 years.

Assuming a normal distribution, we further categorize the under-5 population into four categories: severe (< -3 standard deviations below normal weight for age); moderate ($-3 \leq -2$ standard deviations below normal weight for age); mild ($-2 \leq -1$ standard deviations below normal weight for age); and baseline (> -1 standard deviations below normal weight for age). With the addition of severe acute malnutrition (SAM) and its treatment to the model, we have also included a mechanism by which successfully treated cases of SAM are taken out of the severe distribution of malnutrition and removed entirely from MALNCHP. The relative risks of mortality related to communicable disease category (compared to a baseline risk of 1) are listed in Table 3⁴⁶.

Table 3 – Relative Risks Related to Childhood Undernutrition (existing)

Cause	Mild	Moderate	Severe
Other Group I	2.06	4.24	8.72
Diarrheal Disease	2.32	5.39	12.5
Malaria	2.12	4.48	9.49
Respiratory Infection	2.01	4.03	8.09

Olofin et al, 2013, Weight-for-Age

Cause	Mild	Moderate	Severe
Other Group I	1.54	1.58	8.28
Diarrheal Disease	1.73	2.86	11.56
Malaria	1.26	1.65	1.29
Respiratory Infection	1.85	3.11	10.10

⁴⁶ Relative risk estimates from Gakidou et al. 2007, Table 3: 1880.

6.3.2 Childhood Severe Acute Malnutrition

The population level of childhood SAM impacts IFs forecasts of under-5 mortality related to communicable causes. The “full” IFs forecast is based on estimated average levels of education, a proxy for female empowerment, violent domestic conflict, and governance transparency (see Table X).

Driver	Justification	Data Information
Education	Linked to the quality of maternal health and infant feeding practices, which impacts nutrient intake (Ebenebe, 2012).	Average years of schooling for males and females aged 25 years or older (Barro & Lee, 2010).
Conflict	Conflict environments reduce available food supply by destroying arable land, displacing populations, and reducing social and physical infrastructure (FAO, 1995).	Political Instability Task Force dataset (2013).
Corruption	Corruption reduces the ability of governments and states to implement resources that combat malnutrition (Ngozi-Uchendu & Olatunbosun, 2015).	Corruption Perception Index from Transparency International (2012).

Coefficients for the full SAM equation were estimated using non-linear regression model (nl in STATA version 13) of a sigmoidal form on a pooled panel data set (years 2000 to 2015):

$$z = -2.455 - 0.103 * EDYRSAG25_{r,ss=3} - 0.306 * GOVCORRUPT_r + 0.114 * SFINTLWARMAG_r$$

$$MALNCHPSAM_r = 100 * \frac{e^z}{1 + e^z}$$

Where:

EDYRSAG25(R,3) is average years of education for adults over the age of 25 (male and female), GOVCORRUPT(R) is an index of government corruption (higher is less corrupt) based on the Corruption Perception index from Transparency International, and SFINTLWARMAG(R) is the forecast magnitude of conflict within a country for each country/region r.

This equation is used both in forecasts of MALNCHPSAM, and also to fill missing data. For the latter process, there is an additional step of setting any countries without data that have unimproved sanitation less than 0.5 percent to a seed value of 0.01. Similarly, coefficients for the distal SAM equation were estimated using non-linear regression model (nl in STATA version 13) of a sigmoidal form on a pooled panel data set (years 2000 to 2015):

$$z = -2.821 - 0.118 * LN(GDPPCP_r) - 0.45 * LN(EDYRSAG25_{r,3})$$

$$PMN = 100 * \frac{e^z}{1 + e^z}$$

In the above equation GDPPCP is GDP per capita at purchasing power parity and EDYRSAG25 is average years of formal education for adults over 25. PMN is the percentage of children categorized as severe acute malnourished (≤ -3 standard deviations below the international standard of weight for height). Both the distal and full models incorporate an additive shift factor to match initial year model estimates to historical data. This additive shift factor converges to 0 in 100 years.

Further development of the health sub-module is likely to build a distributional formulation of underweight by height, categorizing the under-5 population into four categories: severe (< -3 standard deviations below normal weight for height); moderate ($-3 \leq -2$ standard deviations below normal weight for age); mild ($-2 \leq -1$ standard deviations below normal weight for height); and baseline (> -1 standard deviations below normal weight for age). Currently, however, the model only supports the fourth distribution, which by default is not active. Using the switch parameter, *samrrsw*, users can substitute the severe distribution, and associated RR values, for MALNCHP with those from MALNCHPSAM. If the switch is not activated, then the identified relative risks of weight-for-age drive mortality/morbidity at every distribution below the mean. When the switch is activated, identified relative risks for weight-for-height (i.e. SAM) drive mortality/morbidity for the lowest distribution of undernourished children ($< -3z$ scores).

The relative risks of mortality related to communicable disease category (compared to a baseline risk of 1) are listed in Table 3⁴⁷.

Cause	MALNCHPSAM	Min	Max
Other Group I	11.21	5.91	21.27
Diarrheal Disease	12.33	9.18	16.57
Malaria	1.24	0.17	9.29
Respiratory Infection	9.68	6.07	15.43

⁴⁷ Relative risk estimates from Olofin et al. 2013, Table 5: 247.

Unlike child undernutrition (MALNCHP), SAM is an acute condition, and the core measure (MALNCHPSAM) represents a point estimate rather than an annual sum of those affected. Affected children may recover due to treatment (official or otherwise) or die within a year. Therefore, to understand the total annual burden of SAM and the upper limit on SAM admissions, it is necessary to estimate incidence, or the new cases of SAM over a given period of time (typically one year). In IFs this measure (MALNCHPSAMINC) is represented as the total annual number of new SAM cases as a percent of children under the age of 5.

In practice, incidence is often estimated from prevalence data using an incidence correction factor, most commonly given as 1.6. This factor is calculated according to the following logic:

$$\text{Incidence Correction (K)} = \frac{t}{(\text{average duration of disease})} = \frac{12}{7.5} = 1.6$$

7.5 months is a commonly accepted estimate for the average duration of SAM based on studies conducted on SAM populations in the DRC and Malawi (Garenne et al, 2009). Therefore, over a time period (t) of one year (i.e. 12 months), this yields:

While the relationships, or parameterization, of time period and duration do not exist in the model, users can adjust the incidence correction factor using the parameter *malnchpsamadj* (initialized to 1.6). Incidence and prevalence are mathematically related such that incidence can be estimated with data on prevalence and the incidence correction factor (K). Specifically:

$$\text{Incidence} = \text{Prevalence} * K$$

As SAM admissions happen continuously throughout the year, and successful treatment does not prevent children from becoming SAM again, cured cases should be estimated in the calculation of incidence. Since IFs' forecast of incidence is designed to capture all known and potential cases of SAM within a one-year time period, we extend the above formulation to include not just prevalence and the incidence correction factor, but also the number of successfully treated cases of SAM. The model captures incidence according to the following logic:

$$\text{Incidence} = \text{Pre-Averted Prevalence} * 1.6,$$

Where 1.6 represents the incidence correction factor, and Pre-Averted Prevalence represents an endogenous calculation in the model of prevalent cases of SAM absent treatment in a given year. In the model this is operationalized as:⁴⁸

⁴⁸ A maximum value of 5 has been exogenously imposed on the ratio between pre-averted incidence and post-averted prevalence, meaning that total incident cases can only exceed prevalence estimates by a factor of five for any given forecast year in the model.

$$MALNCHPSAMINC_r = \frac{MALNCHPSAM_r * malnchpsamadj}{1 - \left(\frac{SAMADMISSION_r}{100} * \frac{SAMAVERTED_r}{100} \right)}$$

The variable for SAM admissions (SAMADMISSION) is initialized using data on the total national number of admissions provided by ACF and its partners. SAMADMISSION is first calculated as a percent of total incident cases, and forecasted as a function of GDP per capita at purchasing power parity (whereby higher admissions rates are driven by increasing GDP per capita).⁴⁹ For OECD countries with no data on admissions the model assumes that all cases of SAM are treated. For all other countries, initial admissions are set to 100. A multiplicative parameter (*samadmissionm*) and an additive parameter (*samadmissionadd*) can be used to change SAM in scenario analysis. We have also included a switch parameter (*samadmissionsw*) that allows users to forecast admissions in three different ways: (0) forecasts all countries according to the relationship with GDPPCP, (1) holds all initial zero values constant over time (default value), and (2) holds all initial values constant.⁵⁰

Success of treatment as a percent of all admission (SAMAVERTED) is initialized from the same data source, and assumed to be 67 percent (the global average) for any country with missing data. A multiplicative parameter (*samavertedm*) can be used to change SAM in scenario analysis. The model also assumes the success rate to converge to 100 percent over a 100-year time horizon, with a parameter (*samavertedconvyr*) that allows users to change the years to converge to 100.

To calculate the final annual value of MALNCHPSAM, we use the following equation:

⁴⁹ This relationship can be found in the model as the bivariate function “GDP/Capita (PPP 2011) Versus SAM Admissions (Linear). Before fitting the regression line, all missing data for OECD countries were set to 100.

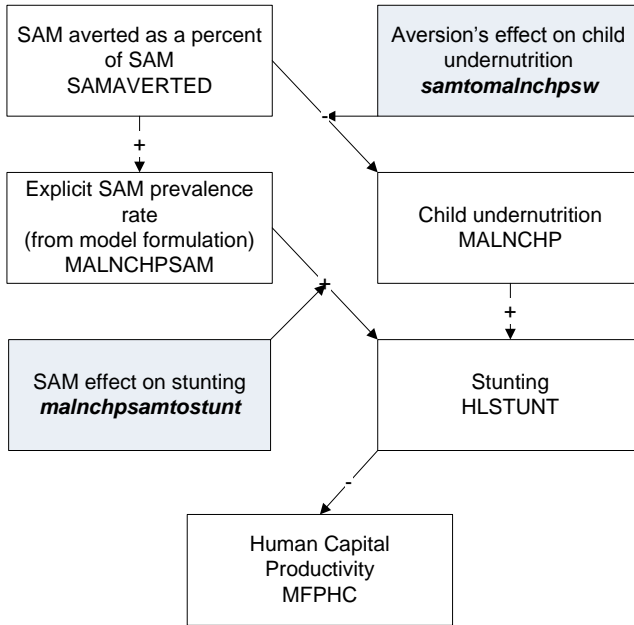
⁵⁰ In addition to parameters that directly affect forecast of admissions, we have included a parameter for the unit cost of treatment (*samadmissioncost*),⁵⁰ which allows us to find the total cost of treatment over the year as:

$$\begin{aligned} SAMADMISSIONTOTCOST_r &= MALNCHPSAM_r * malnchpsamadj / \left(1 - \frac{SAMADMISSION_r}{100} * \frac{SAMAVERTED_r}{100} \right) \\ &* SAMADMISSION_r * \frac{POPOT05_r}{1000} * samadmissioncost_r \end{aligned}$$

$$\begin{aligned}
& MALNCHPSAM_r \\
&= \left(1 - \frac{SAMADMISSION_r}{100} * \frac{SMAVERTED_r}{MALNCHPSAM_r} \right) \\
&* \frac{1 - \frac{SAMADMISSION_r - DiffAdmission}{100} * \frac{SMAVERTED_r - DiffAverted}{100}}{1 - \frac{SAMADMISSION_r - DiffAdmission}{100} * \frac{SMAVERTED_r - DiffAverted}{100}}
\end{aligned}$$

Where *DiffAdmission* is the difference in admission rates between the an unmodified SAMADMISSION value and SAMADMISSION after the additive and multiplicative parameters have been applied, and *DiffAverted* is the difference in averted rates between the an unmodified SMAVERTED value and SMAVERTED after the multiplicative parameter has been applied.

By default, it is assumed that all successfully averted SAM cases are rehabilitated to normal weight for age. Changing the parameter *samtomalnchpsw* allows users to specify scenarios in which successful treatment does not result in full weight (for age) rehabilitation, with a value of 1 resulting in all averted SAM cases remaining malnourished (by default this parameter is set to 0). The diagram below describes how this process connects to the broader economic model through stunting and on to human capital productivity (MFPHC). In addition, it shows a pathway directly from SAM to stunting. This is discussed in greater detail in section 8.3.



6.3.3 Adult Body Mass Index (BMI) and Obesity

Population levels of BMI (HLBMI) impact IFs forecasts of adult (over 30 years) mortality related to cardiovascular disease and diabetes. Both the distal driver and full model formulations are initialized using a multiplicative shift factor to match historic data; these shift factors are kept constant over time. Given the lack of historical data, all regressions were created using 2005 estimates.

Full model formulations for females and males, respectively, use calories per capita (CLPC) as the driver. The parameter *hlbmim* can be used to modify the result:

$$HLBMI_{r,p=2} = (18.73 + 0.00265 * CLPC_r) * \mathbf{hlbmim}_{r,p=2}$$

$$HLBMI_{r,p=1} = (16.54 + 0.00305 * CLPC_r) * \mathbf{hlbmim}_{r,p=1}$$

Distal driver formulations are based on GDP per capita at PPP (GDPPCP) and years of formal education for adults 25 and older (EDYRSAG25):

$$HLBMI_{r,p=2} = 22.7 + 0.46 * \ln GDPPCP_r + 1.36 * \ln EDYRSAG25_r$$

$$HLBMI_{r,p=1} = 21.3 + 0.95 * \ln GDPPCP_r + 1.17 * \ln EDYRSAG25_r$$

In order to use the PAF methodology, we assume a normal distribution of BMI around the mean with a standard deviation of 10% of the mean. BMI has a normal distribution where we forecast the mean, and assume a standard deviation of 10% of the mean.

Section 3.2.1 describes the use of BMI in forecasting diabetes-related mortality. For cardiovascular disease, the relative risk of mortality increases with every unit of BMI. The calculation of relative risk uses a continuous formulation based on BMI level:

$$RR(HLBMI_r) = \mathbf{CONSTANT}^{HLBMI_r - 21}$$

The constant (CONSTANT) depends on age category: 1.14 for 30-44 year olds, 1.09 for 45-59 year olds, 1.09 for 60-69 year olds, and 1.05 for 70-79 year olds.⁵¹

From BMI it is possible also to compute the portion of a population that is obese (HLOBESITY). The model does that as a function of HLBMI by using separate table functions for females (BMI Versus Female Obesity % (CRA) Quad) and males (BMI Versus Male) Obesity % (CRA) Quad). The parameter *hlobesitym* can be used for modification in scenario analysis.

6.3.4 Water and Sanitation

Forecasts of mortality related to diarrheal disease (all ages) depend on access to safe water and improved sanitation. The regression models for each were estimated using the most recent year of data.

⁵¹ Estimates derived from Kelly et al. 2009.

Full model formulations:

$$\begin{aligned}
 & \text{SANITATION}_{r,ss=other\ improved} \\
 & = 0.3809 - 0.0579 * \frac{GDS_{r,g=health}}{GDP_r} * 100 - 0.3309 \\
 & * \ln GDPPCP_r - 0.7968 * \ln EDYRSAG25_r + 0.0233 * \frac{INCOMELT1CS_r}{POP_r} \\
 & * 100
 \end{aligned}$$

$$\begin{aligned}
 & \text{SANITATION}_{r,ss=shared} \\
 & = -1.7235 - 0.0298 * \frac{GDS_{r,g=health}}{GDP_r} * 100 - 0.7834 \\
 & * \ln GDPPCP_r + 0.2591 * \ln EDYRSAG25_r + 0.0195 * \frac{INCOMELT1CS_r}{POP_r} \\
 & * 100
 \end{aligned}$$

$$\begin{aligned}
 & \text{WATSAFE}_{r,ss=unimproved} \\
 & = -0.936 + 0.02 * POPRURAL_r - 0.0891 * \frac{GDS_{r,g=health}}{GDP_r} * 100 \\
 & - 0.8896 \\
 & * \ln GDPPCP_r - 0.2384 * \ln EDYRSAG25_r + 0.0254 * \frac{INCOMELT1CS_r}{POP_r} \\
 & * 100
 \end{aligned}$$

$$\begin{aligned}
 & \text{WATSAFE}_{r,ss=other\ improved} \\
 & = -1.0409 + 0.0293 * POPRURAL_r - 0.1129 * \frac{GDS_{r,g=health}}{GDP_r} * 100 \\
 & - 0.7521 \\
 & * \ln GDPPCP_r + 0.082 * \ln EDYRSAG25_r + 0.018 * \frac{INCOMELT1CS_r}{POP_r} \\
 & * 100
 \end{aligned}$$

Distal Driver formulation:

$$\begin{aligned}
 & \text{SANITATION}_{r,ss=other\ unimproved} \\
 & = 0.9875 - 0.8841 * \ln GDPPCP_r - 0.6483 * \ln EDYRSAG25_r \\
 & \text{SANITATION}_{r,ss=shared} \\
 & = -1.1191 - 0.9798 * \ln GDPPCP_r - 0.1388 * \ln EDYRSAG25_r \\
 & \text{WATSAFE}_{r,ss=unimproved} \\
 & = 1.4998 - 1.4532 * \ln GDPPCP_r - 0.5593 * \ln EDYRSAG25_r \\
 & \text{WATSAFE}_{r,ss=other\ improved} \\
 & = 1.6681 - 1.3199 * \ln GDPPCP_r - 0.2643 * \ln EDYRSAG25_r
 \end{aligned}$$

INCOMELTICS/POP* 100 is the percentage of people living with less than \$1.25 per day
 $GDS_{r,g=health}$ is health expenditures as a percentage of GDP
 POPRURAL is the percentage of the total population living in rural areas

We use a logit formulation to manage the saturation of the 3 levels of access to either of these 2 services, so that the sum of the 3 levels never goes above 100 percent. In this logit formulation⁵² we compute the percentages using the regressions presented, then compute the final results following this method:

$$S1 = e^{(WATSAFE_{ss=unimproved})}$$

$$S2 = e^{(WATSAFE_{ss=oth improved})}$$

$$WATSAFE_{ss=unimproved} = \frac{S1}{S1 + S2 + 1} * 100$$

$$WATSAFE_{ss=other improved} = \frac{S2}{S1 + S2 + 1} * 100$$

$$WATSAFE_{ss=piped} = \frac{1}{S1 + S2 + 1} * 100$$

Where UnimpSWat% is the percentage of people with access to unimproved safe water, OthImpSWat% is the percentage of people with other improved access to safe water and PipedSWat% is the percentage of people with access to piped safe water. The same method is applied for estimating access to improved sanitation.

In order to compute the appropriate PAFs, IFs calculates the proportion of the population that falls into each of the following five categories:

- *Category II*: minimum of (share of population with piped connection for water supply, share of population with improved connection for sanitation)
- *Category IV*: minimum of (share with other improved or piped water supply not in category Vb or II, share with basic or improved sanitation not in category Va or II)
- *Category Va*: minimum of (share with basic or improved sanitation, remainder of those without other improved or piped connection for water supply that are not already in category VI)

⁵² For more detail on this formulation please refer to Rothman, Dale. 2009 (Feb). "Formulae for Predicting Shares 23 Feb 2009.doc", unpublished internal Pardee Center working note.

- *Category Vb*: minimum of (share with other improved or piped water supply, remainder of those without shared or improved access for sanitation that are not already in category VI)
- *Category VI*: minimum of (share without other improved or piped connection for water supply, share without shared or improved connection for sanitation)

Each category has a different Relative Risk associated with it:

Category II: 2.5

Category IV: 6.9

Category Va: 6.9

Category Vb: 8.7

Category VI: 11

The theoretical minimum or international reference is assumed to be 1, and thus the PAF equation gets simplified to:

$$PAF = 1 - \frac{1}{\sum RR(x) * P(x)}$$

6.3.5 Indoor Air Pollution

Indoor air pollution affects forecasts of under-5 mortality related to respiratory infection and adult (30+) mortality related to respiratory disease. IFs uses the percentage of people using solid fuels as their primary source of energy (ENSOLFUEL) as a proxy for indoor air pollution.

The full model calculation is:

$$ENSOLFUEL_{r,t} = \frac{100}{1 + e^{-(2.823 + 0.166 * GDPPCP_{r,t} + 0.032 * INFRAELECACC(national)_{r,t})}}$$

ENSOLFUEL = ratio of electricity use to total primary energy demand, in percentage
 GDPPCP = gross domestic product per capita at purchasing power parity in thousand constant 2005 dollars

INFRAELECACC(national, that is, not urban or rural but total) = percent of total population with access to electricity in percentage

- *multiplicative shift factor: ENSOLFUELShift; never converges*
- *multiplier: **ensolfuelm***
- *targeting parameters: **ensolfuelsetar, ensolfueltrgtyr, ensolfuelsetar, ensolfuelseyrtar***
- *hold switch: **ensolfhlldsw**, fixes value of ENSOLFUEL at initial year value*
- *cross-sectional data, GLM regression, R-squared = 0.81*

The distal driver formulation for ENSOLFUEL uses the following formula, which relies also on EDYRSAG25 and average years of formal education for adults over 25

$$z = 2.9538 - 1.0694 * \ln GDP_{PCP}_r + 1.0668 * \ln EDYRSAG25_r$$

$$ENSOLFUEL_r = 100 * \frac{e^z}{1 + e^z}$$

We use a multiplicative shift factor to match initialization data in the first year, and keep it constant in our forecast. Following work done through the WHO (Desai et al. 2004; Smith et al. 2004), IFs adjusts in the full formulation (not the distal one) the percentage of population exposed to indoor smoke from solid fuels by a ventilation coefficient (*ensfvent*) that ranges from 0 to 1. A coefficient of 0 indicates no exposure to pollutants from solid fuel use, whereas a coefficient of 1 indicates full exposure:

Table 4 - Recommended Ventilation Coefficients to use in Conjunction with Percentage of Population Exposed to Indoor Smoke from Solid Fuels	
Country	Ventilation Coefficient
Albania, Belarus, Bosnia & Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Moldova, Poland, Romania, Russia, Slovakia, Slovenia, Ukraine, Yugoslavia (Serbia and Montenegro)	0.20
China	0.25 for children; 0.50 for adults
All Others	1.0
• From Desai and others (2004).	

Since in the case of indoor air pollution there are only 2 categories—exposed or not exposed, in which case $RR = 1$ - the mortality effect can then be simplified to:

$$ME = \frac{(RR-1)P_{full} + 1}{(RR-1)P_{distal} + 1}$$

Where P is the percentage of population exposed to indoor smoke from solid fuel, adjusted for ventilation; and RR is the relative risk for the exposed population⁵³. Table 5 lists RRs used in IFs.

Table 5 - Relative risk estimates for Mortality from Indoor Smoke from Solid Fuels		
Health Outcome	Groups Impacted	Relative Risk
Respiratory Infections	Children under 5	2.30 (1.90, 2.70)
Respiratory Diseases	Females over 30	3.20 (2.30, 4.80)
	Males over 30	1.80 (1.00, 3.20)

⁵³ More information is available on Dale's documents: "Incorporating Indoor Air Pollution 9 October 2009.docx", unpublished internal Pardee Center working note.

- From Desai and others (2004)
- 95% confidence intervals in parentheses

6.3.6 Outdoor Air Pollution

IFs uses PM 2.5 concentration in urban areas (ENVPM2PT5) as a proxy for outdoor air pollution. Outdoor air pollution impacts mortality related to respiratory infections, respiratory disease, and cardiovascular disease for adults 30 or older.

The distal driver formulation for ENVPM2PT5 uses the following formula:

$$\begin{aligned}
 ENVPM10EXP_r &= 4.8929 + 0.2145 * \ln GDP_{PCP}_r - 0.0995 * (\ln GDP_{PCP}_r)^2 - 0.451 \\
 &* \ln EDYRSAG25_r - 0.0181 * (Year - 1989)
 \end{aligned}$$

In the above equation GDP_{PCP} is GDP per capita at purchasing power parity and EDYRSAG25 is average years of formal education for adults over 25

The full driver formulation for ENVPM2PT5 uses the following formula:

$$\begin{aligned}
 ENVPM10EXP_r &= 5.1238 + 0.2533 * \ln GDP_{PCP}_r - 0.056957 * (\ln GDP_{PCP}_r)^2 \\
 &- 0.4758 * \ln EDYRSAG25_r - 0.0137 * (Year - 1989) - 0.14 \\
 &* \frac{GDS_{r,g=health}}{GDP_r} * 100
 \end{aligned}$$

$$ENVPM10_r = e^{ENVPM10EXP_r}$$

$$ENVPM2PT5_r = ENVPM10_r * ENVPM2PT5ConvFct_r$$

T is time expressed as the current year

GDSHealth%GDP is the government expenditures in health as a percentage of GDP

The first formula returns PM10 concentration levels which then are converted to PM2.5 using a conversion factor. The WHO (Ostro 2004 and EBD spreadsheet) recommends the following conversion factors:

- 0.5 - developing countries outside of Europe
- 0.65 - developed countries outside of Europe
- 0.73 - European countries

In the case of outdoor air pollution we can assume that all persons in urban areas are exposed to the same level of air pollution and therefore the same relative risk. Therefore we can simplify the mortality effect as follows:

$$ME = \left(\frac{PM2.5_{Full} + 1}{PM2.5_{Distal} + 1} \right)^\beta$$

where the recommended value for β is 0.1161.⁵⁴

6.3.7 Smoking Rate and Smoking Impact

The ultimate purpose of forecasting smoking (HLSMOKING) by country/region r and sex p is to forecast smoking impact (HLSMOKINGIMP) by country/region, age category a , and sex p . We provide some background on the general approach surrounding smoking impact and the some specific elements of its implementation in IFs (some of the background comes directly from Hughes et al. 2011: 41-42).

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). Ezzatti 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 (Ezzatti 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 upon 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. Second, we constructed cross-sectional relationships that suggest expected rates of smoking based on GDP per capita at PPP for males and females separately. Third, we initialized a moving average rate of change in smoking rate with the compound rate of change between 1995 and 2005 and used that as the basis for forecasting longer-term. Finally, for forecasting smoking impact longer term we used the

⁵⁴ More Information on: Rothman, Dale. 2009 (Feb). “Incorporating Outdoor air pollution 5 October 2009.docx”

same process in reverse that we had earlier used to estimate the historical smoking series, that is, we calculated smoking impact from smoking rate using a 25-year lag.

In more recent work (beyond that supporting the Hughes et al. 2011 volume, we have introduced an alternative approach to forecasting change in smoking rate over time, one that uses a structural (and heavily algorithmic) smoking stages model. In the sections below we discuss the four steps of the original model and then the revised approach to forecasting smoking rates (work in progress).

6.3.6.1 Historical Smoking Rates

We found it necessary to compute historical smoking rates because we found historical smoking rate data (taken from WHO) to be exceptionally sparse, and we needed to understand the patterns and trajectory of smoking behavior over time as a subsequent basis for forecasting. We may revisit this in the future because we now have data from the WDI for 1977 and more recent years.

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 Global Burden of Disease (GBD). Those GBD forecasts of smoking impact cover the period from 2005 through 2030, provide considerable country coverage, and represent age in four quite large categories: 30-44, 45-59, 60-69, and 70 and older. These can be found in our tables SeriesHealthSmokingImpactMales30to44, SeriesHealthSmokingImpactMales45to59, SeriesHealthSmokingImpactMales60to69, SeriesHealthSmokingImpactMales70to100 in IFsHistSeries.mdb (and the same four tables for females).

Assuming a direct 25-year lag between smoking rate and smoking impact, we used year-to-year percentage changes in the smoking impact series to change smoking rates before and after our smoking data point. In spite of the simplicity of this approach, and the fact that smoking impact reflects more than smoking rates, we found that the constructed series tended to match relatively well when more than one historical point for smoking rate existed.

This is done in a procedure invoked under the IFs menu option Extended Features called Generate Historical Smoking Rate Estimates. The procedure starts by estimating historic smoking rates by age category (4 categories corresponding to the 4 smoking impact age categories of the GBD forecasts) and sex assuming a lag of 25 years (that is, filling in the historical smoking series from BaseYear – 25 to Base Year using GBD smoking impact data from Base Year to Base Year + 25). Then an all-age estimate for smoking rate is found with a simple average across the 4 smoking impact age categories. Next we compute an additive shift factor for each country to match the most recent WHO smoking rate data (from SeriesHealthSmokingPrevalenceWHOFemales% and the same table for males), and then we apply the same shift factor to smoking rate data for previous years. In cases where there are no smoking rate data we compute aggregated shifts using WHO Regions and apply the regional shift to the member country(ies) with no data. The final result of this process are 25 year-long series on smoking rates in the tables SeriesHealthSmokingMales%SI and SeriesHealthSmokingFemales%SI in IFsHistSeries.mdb.

6.3.6.2 Smoking as a Function of GDP per Capita

Once we have an historical rate of smoking, the next step is forecasting it so as to drive forecasts of smoking impact beyond the 2030 of the GBD estimates. Forecasting smoking in IFs is actually a two-step process beginning with the construction of cross-sectional relationships for expected rates of smoking for males and females separately, based on GDP per capita at PPP.

Female smoking rate cross-sectional relationship:

$$Smok_Rate_{r,p,t} = 5.6634 + 0.6893 * GDPPCP_r - 0.00573 * GDPPCP_r^2$$

Male smoking rate cross-sectional relationship:

$$Smok_Rate_{r,p,t} = 38.3996 + 0.3386 * GDPPCP_r - 0.00224 * GDPPCP_r^2$$

Where Smok_Rate by country/region r, sex p, and year t is an initial estimate of smoking rate (percent), and GDPPCP is GDP per capita at PPP (\$1,000). Regression results are kept constant after they reach a GDPPCP of \$30,000 (females) or \$50,000 (males).

The cross-sectional calculation is initial in part because it will not produce results consistent with the data for countries, even in the first or base year of a model run. In the first year an additive shift is computed for both male and female smoking rates to reconcile 2010 values with regression results. In future model years, the additive shift is evaluated based on regression results: if it is positive (thus producing the forecasted value to be above the expected value given by the regression) or for all non-high-income countries (initial GDPPCP <= 25k) the shift is converged to 0 over 100 years. If it is negative and the country is high income, the shift is kept constant for the entire run horizon. The resulting adjusted shift is added to the smoking rate produced by the regression.

6.3.6.3 Forecasting Changes in Smoking using Past Behavior and GDP per Capita Expectations

It is important to recognize not just the initial empirical (or estimated) value of smoking rate in our base year for each country, but the trajectory of country-specific change in smoking rates. Our approach to capturing the trajectory is a variation on a moving-average approach. Every year our first step is to compute a compound growth rate of the smoking rate over the last 10 years. In the second step we also compute the rate of change that one would expect based solely on applying the cross-sectional formulation in two consecutive years. The third step is to compute a slowly-changing moving average by combining those two growth rate computations, weighting the compound historical growth rate 90 percent and the expectation of growth from the cross-sectional formulation 10 percent.

In the first step described above we obtain every year the compounded growth rate (Comp_Gr_Rt) over the previous 10 years. That growth rate is based on historical (or constructed) smoking rate data (Smok_Rate).⁵⁵

$$Comp_Gr_Rt_{r,p,t=1} = \left(\frac{Smok_Rate_{r,p,t=1}}{Smok_Rate_{r,p,t=1-10}} \right)^{\frac{1}{10}} - 1$$

In the second step described above we compute the annual expectation of growth in the smoking rate (Exp_Gr_Rt) based on the cross-sectional formulation expectations (Smok_Rate_Exp) for the current and previous year.

$$Exp_Gr_Rt_{r,p,t} = \frac{Smok_Rate_Exp_{r,p,t} - Smok_Rate_Exp_{r,p,t-1}}{Smok_Rate_Exp_{r,p,t-1}}$$

In the third step, every year we compute what is effectively a moving average of change in smoking rates (Mov_Gr_Rt) by combining the compound growth rate for the last 10 years with the expected value from the preceding to the current year (with 90 percent and 10 percent weighting, respectively).

$$Mov_Gr_Rt_{r,p,t} = 0.9 * Comp_Gr_Rt_{r,p,t} + 0.1 * Exp_Gr_Rt_{r,p,t}$$

In the above computation 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 cross-sectional 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.

It is then possible to apply the moving average to obtain what is actually a preliminary forecast of smoking rate (although not used in the model, we can call it HLSMOKINGP) based on that in the previous year.

$$HLSMOKING_{r,p,t} = HLSMOKING_{r,p,t-1} * (1 + Mov_Gr_Rt_{r,p,t})$$

This preliminary smoking rate is then converged to match the result of the regression equation over a period of 100 years and yield a near final smoking rate (HLSMOKING)

$$\begin{aligned} & HLSMOKING_{r,p,t} \\ & = ConvergeOverTime(HLSMOKINGP_{r,p,t}, Smok_Rate_Exp_{r,p,t}, (BaseYear + 100) \\ & - t) \end{aligned}$$

⁵⁵ The constructed historical smoking data contributes to this computation of the compound growth rate for the first 10 years of its computation; thereafter, the computation will rely only on values generated by our forecasting.

High income countries (with initial GDP per capita at PPP or GDPPCPI > \$25,000) then are checked to avoid smoking rate growth after they have started to drop, i.e:

$$\begin{aligned} &\text{If GDPPCPI} > \$25,000 \text{ and } HLSMOKING_{r,p,t} > \\ &HLSMOKING_{r,p,t-1} \text{ and } HLSMOKING_{r,p,t-1} \leq HLSMOKING_{r,p,t-2} \text{ then} \\ &HLSMOKING_{r,p,t} = HLSMOKING_{r,p,t-1} \end{aligned}$$

Finally, a multiplier (*hlsmokingm*) by country and sex is available and applied to the smoking rate to compute the final value for the country-sex-year; the default value of the multiplier is 1 and alternative values introduce scenarios. Note that having a multiplier for a specific sex (e.g. 0.9 for males) and another for “total” or both sexes (e.g. 0.8) will produce a multiplicative effect on the forecast of 0.72. This ability to stack multipliers for individual sexes with total or both is not the standard practice in IFs; this may be the only example of it.

6.3.6.4 Computing Future Smoking Impact from Future Smoking Rates

With the year-to-year percentage change in smoking rate forecasts from 2010 forward, we change the year-to-year values of the smoking impact series 25 years later. Thus, we use the reverse process used earlier—rather than using changes in smoking impact to compute smoking rates, we use changes in smoking rates to compute smoking impact.

One complication is that we need smoking impact by age-category. We first compute smoking impact in the 4 big age categories of the GBD data. For purposes of affecting mortality we then apply the large age category values to each of the underlying 5-year categories; for example we compute smoking impact for 30-44 year olds, and then we apply the value for 30-34, 35-39, 40-44 year olds.

Even in the big categories, however, as well as in the smaller 5-year categories, a consequence of our approach is that the changes in smoking impact year-to-year are identical in each age category. This is an inevitable consequence of our not having smoking rates by age—we have no basis for positing different patterns of smoking over time by age and therefore for changing the smoking impact by age.

An unfortunate corollary consequence of this procedure is that, in the years between our initial forecast year and 2030, a period for which we have the smoking impact forecasts of the GBD but are also computing our own smoking impact forecasts, our forecasts will differ somewhat from those of the GBD. Remember, however, that we used GBD forecasts to create historical smoking rate series, and in those forecast years through 2030 are computing smoking impact from the historical smoking rates; thus our smoking impact forecasts through 2030 will not differ greatly from those of the GBD. It would be possible for us to actually use the GBD smoking impact forecasts for the 4 age categories through 2030 and then initiate our own changes to all 4 categories; we judged that the computational intensity of using theirs through 2030 offset the advantages of doing so.

Note that it is possible to stop changes in the Smoking Impact forecast in year 2030 and beyond by turning off the smoking impact switch, *hlsmimpsw* = 0.

6.3.6.5 Revised Approach: Forecasting Smoking Rates using a Stages Model

Smoking rate itself is computed in two different ways. The basic formulation uses only the initial condition, historical rates of growth in smoking, and a cross-sectionally estimated function linked to the simple and squared values of GDP per capita at PPP. The more extended formulation (still in development and testing and therefore not turned on as the default approach) is an algorithmic one based on the same general concept of a pattern that initially rises with GDP per capita, peaks, and then falls, but with a series of parameters that allow much more control over the stages.⁵⁶ This staged algorithmic approach (see Lopez et al. 1994; Shibuya et al. 2005; Ploeg et al. 2009) is turned on with a switch (*hlsmokingstsw*=1; the default =0). Based on country- and sex-specific trends, plus initial prevalence, in the first year each country is placed into one of four sequential stages: rising, peak, falling, late. During the forecast horizon they can move through other stages. The approach is heavily algorithmic and the logic is explained below.

Early to Rising (Stage 1). Conditions for being in group are: Male prevalence rises during historic estimate and current prevalence is $\leq 15\%$. Female prevalence rises at least until the last 5 years of history and current prevalence is $\leq 1\%$.

The forecasting strategy is: Ongoing increase. Use compound growth rate (10-year moving average from current year) when it is $< 5\%$ per year, increase growth rate to 5% /year over the next ten years. Continue 5% growth rate until country moves into next group (i.e. male smoking prevalence $> 15\%$, so country moves to stage 2). In the case of females, the previous year growth rate is kept constant until transition to next stage (female smoking prevalence $> 1\%$).

Rising to Peak (Stage 2). Conditions for being in this stage are: Male prevalence rises during historic estimate and current prevalence is $> 15\%$. Female prevalence has been rising at least until last 5 years of history and current prevalence is $> 1\%$ and $< 45\%$.

The forecasting strategy is: Exponential increase. If exponential rate of increase $< e^{-0.1x}$, where x =year of forecast (1,2,3,... end of forecast horizon), then increase = $e^{-0.1x}$, otherwise use constant growth. For females we use an exponential regression ($y=ab^x$), limit annual growth to a maximum of 0.5% . If country moves from stage 1 to stage 2 during forecast horizon, exponential rate of increase = $e^{-0.1x}$, where x =year of forecast (1,2,3,... end of forecast horizon - x begins again at 1 when country moves to stage 2 both for males and females). In base case forecast the ceiling is 65% male prevalence for countries with increasing smoking trends. If male prevalence rises to 65% during forecast, country moves to next stage (3). The limit for females is 45% prevalence.

⁵⁶ Cecilia Peterson developed this approach for IFs.

Peak to Falling (Stage 3). Conditions for being in stage are: Male prevalence peaks during historic estimate (20 years before the base year or later). Female prevalence peaks during the last 5 years of history and current prevalence is $\geq 45\%$.

The forecasting strategy is: If smoking rate peaks during historic estimate then do 10 years of exponential decrease from last year of peak ($y = ab^x$). In other words, regression based on historic estimate starting with last year of peak (not all years of historic estimate).

After 10 years, faster exponential decrease:

$Prev_{x+1} = Prev_x * e^{-0.01x}$, where $x = \text{year}$ (1,2,3,... end of forecast horizon - x begins again at 1 when country begins faster decrease).

If male prevalence falls to $\leq 20\%$, then flat rate to end of forecast horizon.

If smoking rate peaks during forecast slow exponential decrease for first 10 years:

$Prev_{x+1} = Prev_x * e^{-0.001x}$, where $x = \text{year}$ (1 to 10)

Faster exponential decrease for rest of the forecast:

$Prev_{x+1} = Prev_x * e^{-0.01x}$, where $x = \text{year}$ (1,2,3,... end of forecast horizon - x begins again at 1 when country begins faster decrease).

If male prevalence falls to $\leq 20\%$, then flat rate to end of forecast horizon.

For Females simply do a slow logarithmic decline:

$SmRt(t) = SmRt(t-1) - 0.05 \text{ LN}(\text{BaseYear} + \text{Yr} - 1 - \text{PeakYear})$

If female prevalence falls to $\leq 15\%$, then flat rate to end of forecast horizon.

Late (Stage 4). Conditions for being in the stage are: Male prevalence declines throughout at least the last twenty years of historic estimation (ie peak year $<$ BaseYear - 20). Female prevalence declines at least the last 5 years of historic estimation.

The forecasting strategy is: If current male smoking rate (prevalence) $\geq 30\%$ exponential decrease ($y = ab^x$) beginning with last year of peak. In other words, regression based on historical data starting with last year of peak (not necessarily all years of historical data). Trend continues through forecast horizon or until male prevalence $\leq 20\%$. If male prevalence $\leq 20\%$ then flat rate to end of forecast horizon

IF current male prevalence $< 30\%$ logarithmic decline ($y = a + b * \ln(x)$) beginning with last year of peak. In other words, regression based on historical data starting with last year of peak (not necessarily all years of historical data). Trend continues through forecast horizon or until male prevalence $\leq 20\%$. If male prevalence $\leq 20\%$ then flat rate to end of forecast horizon.

Female forecast simply uses a logarithmic decline ($y = a + b * \ln(x)$). Trend continues through forecast horizon or until female prevalence $\leq 15\%$. If female prevalence $\leq 15\%$ then flat rate to end of forecast horizon.

There are a number of parameters for the stages approach to forecasting smoking rates. Because control of tobacco is a major policy objective in many countries, a number of

these relate to the representation of a tobacco control score on a 100-point scale (*hlsmokingtcs*) with an associated parameter to control the elasticity of smoking with that score (*hlsmokingtcsel*), as well as a multiplier on the score (*hlsmokingtcsm*). The parameters related to tobacco control are:

- Tobacco Control Score: *hlsmokingtcs*, higher scores reduce growing trend or accelerate reduction.
- Elasticity of relationship between tobacco control score and smoking rate (*hlsmokingtcsel*).
- Multiplier on tobacco control score (*hlsmokingtcsm*)

Other parameters related to the stage approach are:

- Smoking Multiplier for Increasing Trend: *hlsmokingincm*, only affects countries in stage 2.
- Smoking Multiplier for Decreasing Trend: *hlsmokingdecn*, only affects countries in stage 3.
- Smoking Ceiling and Floor can be controlled using: *hlsmokingceiling* and *hlsmokingfloor*.
- Smoking Peak Year can be controlled using: *hlsmokingpeakyr*.

When the stages approach is turned on the computation of smoking impact from the approach is changed from that described for the basic smoking model based on the quadratic equation with GDP per capita. To compute smoking impact when using smoking in stages we use linear regressions between Smoking Prevalence and estimated Smoking Impact (smoking impact lagged 25 years) to estimate Smoking Impact by large GBD age category. For instance, one of those functions is “Male Smoking Prevalence (2005) Versus Male 30 to 44 Smoking Impact (2030) Linear”. The functions were computed using smoking rates in 2005 and GBD smoking impact forecasts for 2030. In forecasting we apply those functions to smoking rates 25 and 26 years earlier, and obtain two values for smoking impact. We compute the growth rate between those two values of smoking impact and apply it to the previous year’s value of smoking impact for each of the 4 large GBD age categories. We then reproduce the smoking impact value for each 5-year age category in the larger GBD categories.

For small values of SI (lower than 1) we restrict the annual change rate to be between -50% and 100%. If SI is 0 in the base year then it stays at 0. If the estimated value of SI reaches 0, and the next year is positive, then a 100% growth rate is used, but if the next year is negative then a -50% growth rate is used.

7. Outcomes

There are two main mortality-related outcomes from the Health Module: deaths by cause-category (DEATHCAT) and life expectancy (LIFEXPHLM).⁵⁷ From these, IFs calculates other relevant outcome variables including: years of life lost (YLL; HLYLL in IFs); years lived with disability (YLD; HLYLD); disability adjusted life years (DALY; HLDALY); YLLs and YLDs among the working-age population (HLYLLWORK and HLYLDWORK,⁵⁸ respectively, the sum of which would be DALYs for the working population); morbidity and the probability of mortality during a user-defined age-range.⁵⁹

The computation of life expectancy in the health module is a replica of the one in the population module, with the only difference being the mortality distribution. In the initial year, these match because of the normalization process but they grow apart as the model advances. DEATHCAT is computed by multiplying the mortality distribution from the Health Module by population age categories.⁶⁰

7.1 Years of Life Lost (YLL)

Years of Life Lost (YLLs) are computed using the number of deaths in each age category multiplied by the number of years they died prematurely (the potential life expectancy at the age category as determined by a stored vector of standard life expectancy):⁶¹

$$\text{Years of Life Lost} = \text{Deaths in age category} * \text{StdLifeExp for age}$$

There are two extra elements that are considered in the computation of YLL – discounting and age weighting. Discounting represents the preference of individuals for current, rather than future, benefits. Age weighting is an attempt to capture age-specific social roles, with working ages more heavily weighted than children or older adults. While controversial, both elements are used in GBD studies and are therefore useful for standard comparisons.

If we initially considered that the number of years lost per each death is the expected life expectancy minus the average age in a given category, then we now need to use the following formula to apply both discounting and age weighting:

⁵⁷ Note that LIFEEXPHLM replaces the old IFs variable LIFEEXP which was forecasted in the population module.

⁵⁸ As of October 2013 we were using the ages 20-64 as the range for computation of these two variables. We may want to make that dynamic now that we have dynamic working life years.

⁵⁹ The default is child mortality (0 to 5) and adult mortality (15 to 60).

⁶⁰ If the user turns off the health model switch (*hlmodelsw*), IFs stops computing all of the HL variables and DEATHCAT (also morbidity and the probabilities of mortality); they have values with the switch set to off because they take those of the base case, but they are not used to compute LIFEXP, which would then come from the older or legacy population model. It is important to note also that if the health model switch is off, it is impossible to expand the age categories beyond 100+.

⁶¹ Randall Kuhn provide the hard-coded vector for standard life expectancy.

$$years = \frac{-AWC * e^{(-\beta * Avg(j))}}{(\beta + r)^2} * \left(e^{(-(\beta+r) * (StdLE(j))) * (1+(\beta+r) * StdLE) - (1+(\beta+r) * Avg(j))} \right)$$

Where AWC is the age weighting correction factor (0.1658), this constant compensates so that the global estimated burden of disease remains the same as if not using age weighting. β is the parameter from the age weighting function (0.04). These two constants are related and if you want to change the form of the weighting function by changing β , then you would have to change AWC too. R is the Discount Rate, which we have set to 3 percent to match the GBD study.

YLLWORK uses the same formulations, but only includes the population of 20 to 64 year-olds.

7.2 Years of Life Lost to Disability (YLD)

External to the model, the IFs project calculated an initial ratio of YLD to YLL based on data from the World Health Organization. Those ratios are regional but specific to cause of death category. In the base year of the model, IFs initializes the value for YLDs by applying this ratio to YLL. In subsequent years IFs calculates the growth rate for mortality from the previous year and uses that to forecast the change in morbidity. The percentage decline in disability relative to decline in mortality can be controlled by the parameter (*hlmorbtomortgthport*). IFs estimates YLD from the ratio of regional morbidity to mortality (YLD/YLL). This ratio is used for initialization and then the growth rate for mortality is computed and applied to forecast morbidity. Morbidity results can then be adjusted with a multiplier (*hlmorbtomortgthport*) that can be handled by the user across time. This multiplier is specific for each subtype, with default values detailed in the table below. As an example, the value of 50 percent for cardiovascular disease (*hlmorbtomortgthport*=0.5) suggests that morbidity decreases at half the rate of mortality. Because malaria is set at 100 percent (*hlmorbtomortgthport* =1.0), morbidity will change at exactly the same rate as mortality; a higher value for the parameter could be used to represent a situation in which morbidity declined faster than mortality.

When the morbidity to mortality parameter is set to a value different than 1, we have an asymmetry in the growth applied, for example see the case mentioned above where cardiovascular disease is set to 0.5, a decrease of 10% in mortality will equal a decrease of 5% in morbidity, say the level of mortality was at 10 deaths per 1000 person, then mortality would go down to 9, and if morbidity was also set at 10 per 1000 person, morbidity then would reach 9.5. Now see what happens if mortality immediately jumps back up by the same amount, say from 9 to 10, that's an 11.11% increase, if we were to apply directly 50% of that increase to morbidity (5.56%) we would overdo the morbidity bounce back, note that $9.5 * 1.0556 = 10.0278$. In order to avoid this asymmetry in the up side of morbidity we are instead using the following logic:

Pd is the portion that mortality decreases

Pu is the portion that mortality increases to jump back to the original value

Then we have that:

$$1 + Pu = 1/(1 + Pd)$$

Then when mortality increases by a portion Pu , is like recovering from a previous decrease by a portion Pd . When Pu and Pd are applied as is to morbidity, then the 2 move together symmetrically, but when we start applying a factor f to Pu and Pd the problems shown above emerge.

If we have a decrease of Pd in mortality and we want a decrease of $Pd*f$ in morbidity, we need to consider that in order to recover from this decrease we would need to multiply morbidity by: $1/(1+Pd*f)$, which is not the same as $1 + Pu*f$, as you can see from the previous equation, so if we solve for Pd in the equation for mortality we get:

$$Pd = 1/(1 + Pu) - 1$$

Replacing in the equation for morbidity we then get the factor we need to apply to morbidity when mortality goes up by a portion of Pu :

$$1/(1+(1/(1 + Pu) - 1)*f)$$

Note how this reduces to simply $1+Pu$ when $f = 1$.

For the previous example we would have:

$$1/(1+(1/(1 + 0.1111) - 1)*0.5) = 1.0526$$

and when we apply this to the morbidity value of 9.5 we get the original 10 that we were looking for.

In conclusion, when the portion parameter is different than 1, like it is for Cardio Vascular disease in the base case, and we're increasing mortality by a portion Pu , then we multiply morbidity by: $1/(1+(1/(1 + Pu) - 1)*f)$

In the case where we're decreasing mortality, we do use the simple $1 + Pd*f$ formulation for morbidity.

Percentage decline in disability relative to decline in mortality

Group I	Percentage change in disability
Diarrhea	75 ($hlmorbto mortgthport=.75$)
Malaria	100
Respiratory	100
Other communicable	75
HIV/AIDS	Modeled separately
Group II	
Cardiovascular	50
Digestive disorders	100
Malignant neoplasms	100
Diabetes	100
Mental health	Not Applicable; given value of -1
Chronic respiratory	100
Other NCDs	50
Group III	
Intentional injuries	75
Traffic accidents	75
Other non-intentional injuries	75

Once morbidity is computed, the calculations to find YLDs (HLYLD) are identical to those for YLLs (HLYLL)—using morbidity instead of mortality to compute the number of people affected by a given disease. HLYLDWORK again focuses only on the population aged between 20 and 64.

Morbidity related to mental health is the one exception to the above methodology. IFs computes an initial ratio of YLD/POP based on WHO data, keeping the ratio constant over time.

It's also important to note that once mortality reaches 0, then morbidity stays constant for the rest of the horizon, unless mortality starts to increase again.

7.3 Disability adjusted life years (DALY)

DALYs are computed by summing YLLs and YLDs. YLLWORK and YLDWORK (the sum of which would be DALYs for the working population) include only the 20 to 64 year-old population.

7.4 Morbidity

We have added 3 parameters to control Morbidity.

NAME	DESCRIPTION	DIMENSION	DEFAULT	HORIZON
<i>hlmorbm</i>	Target Morbidity Multiplier	None	-1	All Years
<i>hlmorbconv</i>	Convergence to Target Morbidity	None	20	All Years
<i>hlmorbtomortgthport</i>	Morbidity to Mortality growth portion	Disease sub-types (15)	1	All Years

The first one works as both a switch and a multiplier; in terms of its functioning as a switch, a value of -1 will turn it off. When it is given a value above 0 it functions as a multiplier. The multiplier works as a portion of initial morbidity to compute Target morbidity, so that if you set it to 1, it will keep morbidity constant, and if you set it to 0 it will eliminate morbidity. The second parameter is used to indicate how many years to converge from the regular Health Module morbidity to the Target morbidity. Note that if you want to keep morbidity constant you need to change the convergence parameter to 0 to avoid using Health Module morbidity at all.

The last parameter controls the change rate of morbidity as described in section 7.2.

7.5 Mortality probabilities

IFs allows the user to compute the probability that a person of a given age (e.g. 15) will live to reach another age (e.g. 60)—to compute it look for under the Display Type Options on the specialized display for a J-curve (there are probabilities and rates for children and adults). In this example, the probability of a 15-year old dying before she reaches 60 equals 1 minus the cumulative probability of surviving (l_x) to 59 given that she has survived to 15:

$$P(15-59) = 1 - l_x(59) / l_x(15)$$

In order to compute l_x at age j we need to consider the cumulative effect of the previous age category and the probability of death in the current age category (nqx). $L_x(0)$ is assumed to be 1, given that we are only considering deaths for people that are born alive:

$$l_x(j) = l_x(j-1) * (1 - nqx)$$

The probability of death at the current category is computed based on the mortality of the age category nMx , the number of years in the given category N (5 for most of the IFs age categories), and the average years lived within the same category nax , which in most cases is 2.5, but can be lower for shorter age categories:

$$nqx = (N * nMx) / (1 + ((N - nax) * nMx))$$

This adjustment is necessary because nMx is the mortality rate of the 5 year period (which is not the same for each of the 1 year periods within it). The mortality rate nMx already considers that some people die in the middle of the period, which we don't need for the probability nqx , which is why in general probabilities are lower than mortality rates and more pertinent for older ages where people tend to die earlier in the age category.

Although the infant age category covers a shorter age range, the rate correction is in some ways a bigger issue because infants tend to die within the first days of life. The basic framework for understanding nax in this category is thus that the higher the mortality, the higher the average years lived nax . In a highly developed country such as Sweden, nearly all infant mortality takes place in the neonatal period (so nax is almost = 0). Alternately, in a country such as the Congo, infant mortality takes place throughout the year (though it is still concentrated in the neonatal period) and nax rises fairly consistently with nMx . This has been implemented in the following way:

Average years lived by those who die (per Keyfitz) for Infants

```

If  $nMx \geq 0.107$  Then
     $nax = 0.34$ 
Else
     $nax = 0.049 + 2.742 * nMx$ 

```

When estimating nax for children aged between 1 and 4, the logic becomes that child deaths between age 1 and 5 are a prolonged extension of infant mortality. Estimations have thus shown that nax is more directly tied to infant mortality than it is to nMx for its own age category. In other words, countries with very high infant mortality also experience elevated child mortality, mostly concentrated in the 1 - 2 age range. Thus as infant mortality rises so does child mortality, pulling nax away from 2:

Average years lived by those who die (per Keyfitz) for children 1 to 4

```

If  $infMort \geq 0.107$  Then
     $nax = 1.356$ 
Else
     $nax = 1.587 - 2.167 * infMort$ 

```

8. Forward Linkages

Chapter 7 of Hughes, Kuhn, Peterson, Rothman, and Solorzano (2011) elaborated the forward linkages of the health model to other parts of the IFs system at the time of that volume's completion. It began by discussing a controversy in the literature about whether the effects on economic well-being (as indicated by GDP per capita) of improvements in life expectancy are positive or negative. It went on to devote much attention to three major and general pathways of impact between health and GDP, each of which corresponds to a major element in standard production functions and that in IFs: labor, capital, and multifactor productivity. The equation documentation uses the same division and then provides some attention to potential forward linkages that do not involve those same three paths (notably the impact of health on public spending on health on the years of education attained by members of society and the quality of that education).

8.1 Population and Labor Supply

The IFs demographic model captures the mechanical or accounting effects of mortality on population. A key pathway passes from mortality through adult age population to labor supply (including aging-related lags).⁶² Similarly, IFs captures the mechanical effect of mortality on fertility through the death of women of childbearing age.

The most important non-mechanical linkage is almost certainly the relationship between child mortality and fertility. IFs forecasts fertility as a relationship with infant mortality, the log of educational level of those aged 15 and older (neither the education of women alone nor the education of those 15-24 work as well), and the percentage use of modern contraception. Adding the rate of infant mortality boosted the overall adjusted R-squared to 0.84.

$$TFR_{r,t} = 3.8812 + 0.21 * INFMORT_{r,t} - 0.8327 * LN(EDYRS15_{r,t}) - 0.009 * CONTRUSE_{r,t}$$

R-squared = 0.8345; other terms and algorithmic specifications modify the relationship, most importantly a term that slowly shifts TFR over time and the specification of a minimum level toward which the function slowly converges if it overshoots on the downward side.

8.2 Capital stocks

Most capital stock consists of buildings and machinery for producing goods and services; some representations may include land also, but most treat land separately and largely as a constant (although land developed for crop production or grazing can, in fact, be highly variable). Most immediately, investment increases capital stock and depreciation reduces it. Although there is certainly some impact of morbidity and mortality on the rate of depreciation of both built physical and natural capital, the relationship may not be substantial and we do not understand it well enough to model it. Turning our gaze to the paths that affect investment, the three major ones run through health spending, which can crowd out savings and investment, through the age-structure of societies, which affects

⁶² IFs also includes income-based formulations for changing the female participation rate.

the savings rate, and through investment from abroad, which can augment that generated domestically.

The IFs modeling system treats capital stock dynamically over time, investing in it and allowing it to depreciate. Investment is responsive to both domestic savings and foreign flows. Thus, the necessary elements for considering the impact of morbidity and mortality, via paths such as those across health spending and age structure, are part of the economic model's core structure.

With respect to health spending, to which we return later, the IFs model uses a social accounting matrix (SAM) structure. Thus the flow of funds into health spending automatically competes with other consumption uses and with savings and investment. In addition, health spending does affect mortality. We focus here on the paths versus the age-structure and foreign direct investment.

Age Structure and Domestic Savings. The paths in IFs that link age structure most directly to domestic savings have two important elements. The most fundamental one represents the understanding of life-cycle dynamics in income, consumption and savings. The cycle for income is fairly clear-cut with a peak in the middle to latter periods of the working years. Workers set aside some portion of income as savings and that portion, too, tends to peak in the middle and late period of working years. Society-wide savings themselves become negative after retirement age (65 in the Base Case scenario) even though some portion of the population will continue to work. The second fundamental element is that both the horizon of life expectancy and the average income level of a society can have an impact on the portion set aside for savings and the degree to which it rises and then falls. Thus, for example, the life-cycle “bulge” of savings may be earlier and flatter in developing countries.

We implemented the representation of savings and investment in accord with that understanding. Relying upon analyses of selected countries that Fernández-Villaverde and Kruegger (2004 and 2005) and Deaton and Paxson (2000) undertook, we extracted general stylized patterns of the savings life cycle to represent more and less developed (and lower life expectancy) countries. In forecasting we use the pattern for less developed countries when life expectancy falls below 40 years, use that for more developed countries when life expectancy exceeds 80 years, and interpolate in between for all other countries. The result of this largely algorithmic approach⁶³ is an adjustment factor (SavingsAgeAdj) that augments or reduces investment.

In addition, investment is somewhat augmented or reduced as a direct result of changing life expectancy. Life expectancy is compared over time with an expected value (tied to cross-sectional estimation with income). That difference is compared to the difference in the initial year and, if it rises, augments investment.

⁶³ See the subroutine SavingsDemogAdj in routine Populat.bas, which draws upon table IncConSav in IFs.mdb with different patterns of income, consumption, and savings for more developed countries (MDCs) and less developed countries (LDCs) across age categories; in general, peaks of income, consumption, savings occur the in late 40s and savings turn negative at 65.

Although conceptually tied to savings rates, neither the life-cycle analysis nor the life-expectancy term directly affect savings in IFs. Instead, they affect investment directly and savings indirectly via the dynamics in IFs that balance savings and investment over time.

Foreign Direct Investment. The path linking health to foreign direct investment is potentially quite important. Alsan, Bloom and Canning (2006: 613) report that one additional year of life expectancy boosts FDI inflows by 9 percent, controlling for other variables. We have implemented that relationship in IFs. The representation of FDI in IFs captures the accumulation over time of FDI inflows in stocks of FDI, as well as the accumulation of FDI outflows in stocks. In addition, the stocks set up their own dynamics, including the tendency for stocks to reinforce flows. For that reason, we have set the base case parameter for the impact of each year of life expectancy on FDI flows to 0.05 (5 percent), lower than the estimate of Alsan, Bloom and Canning (2006).

8.3 Productivity

Health outcomes impact productivity through a variety of pathways (see Figure 8.3). Overall the function for multifactor productivity from human capital (MFPHC) is a sum of a term linked to educational expenditures (EDEXPCONTRIB) and three terms of interest to us here because mortality and morbidity affect them. Those three of interest are adult stunting (STUNTCONTRIB), disability (DISABCONTRIB) and years of education (EDYRSCONTRIB). We detail these three contributions in turn.

$$MFPHC_r^t = EDEXPCONTRIB_r^t + STUNTCONTRIB_r^t + DISABCONTRIB_r^t + EDYRSCONTRIB_r^t$$

Stunting Contribution. In the IFs health module, the prevalence of adult stunting (HLSTUNT) relates negatively to overall productivity (an elasticity of -0.025, in *mfpstunt*). In extreme cases, stunting could cost as much as 1 percent of economic growth.

$$STUNTCONTRIB_r^t = (HLSTUNT_r^{t-1} - STUNTINGCOMP_r^t) * mfpstunt$$

where

$$STUNTINGCOMP_r^t = F(GDPPCP_r^t)$$

We compute HLSTUNT in the health model itself. We initialize adult stunting in a long-term lagged relationship (using a moving average controlled by the parameter *hlstuntyrs*)⁶⁴ with child malnutrition and child severe acute malnutrition and forecast it as a function of malnutrition, SAM, and child mortality as a proxy for morbidity. Initial

⁶⁴ The lag is initialized to 25 years, found as the difference from the midpoint of childhood (7.5) to the midpoint of adulthood (32.5).

values in 2005 range up to about 55 percent for India and Bangladesh and even over 80 percent for Somalia; in the base case these generally but not universally decrease.⁶⁵

To initialize HLSTUNT in the preprocessor, IFs must first estimate historic levels of childhood undernutrition (MALNCHP):⁶⁶

$$MN = 23.853 * GDPPCP^{0.6721}$$

First we find the result of this function with GDPPCP numbers from 2005, then we compute an additive shift factor to match initialization data for MALNCHP(2005). Second we compute the result of the function with GDPPCP from 1980 and apply the shift factor to estimate HLSTUNT in 2005. We use a limit of 80% for the maximum possible stunting value.

Since there lies some uncertainty as to whether SAM has a disproportionate impact on stunting compared with child undernutrition, we have updated the forecast of HLSTUNT to include MALNCHPSAM, with the parameter *malnchpsamtostunt*, which controls to what degree MALNCHPSAM affects HLSTUNT, set to zero initially. IFs forecasts HLSTUNT using an extremely slowly moving average:

$$HLSTUNT_{t,r} = \frac{HLSTUNT_{t-1,r} * (hlstuntyrs - 1) + MALNCHP_r + (MALNCHPSAM_r * malnchpsamtostunt)}{hlstuntyrs} + HLSTUNHeight$$

Childhood malnutrition, morbidity, and SAM do not give rise to all disability in working years; much also comes from disabilities arising during the working years. IFs therefore also calculates millions of years of living with disability related to mortality rates specific to the working aged-population.⁶⁷

8.4 Additional potential forward linkages

There are at least three additional forward linkages that could be usefully added to the model. First, we know that morbidity affects health spending, but the model does not have that relationship (change in health spending is linked instead to change in GDP per

⁶⁵ Global data on stunting among adults appear nearly nonexistent. UNICEF (2009:5) suggests that under-5 stunting exceeds that of malnutrition (200 versus 130 million) and that stunting is nearly irreversible with aging; these facts suggest very high percentages of stunting among global adults, concentrated in Africa and Asia.

⁶⁶ While IFs includes historic data series from WDI and WHO for child undernutrition, many countries do not have data for 1980 (25 years prior to 2005, our initial year).

⁶⁷ As a quick reality check on those numbers, dividing disability years for the working population by population aged 15-65 generates numbers in 2005 that range from around 0.20-0.27 at the top end of the range (Timor-Leste, Afghanistan, Montenegro, Puerto Rico, Cambodia, and mostly other African countries) to 0.05-0.06 at the bottom end (Kuwait, UAE, Cape Verde, Algeria, Japan and mostly other rich countries).

capita). Because health spending competes with other government spending (IFs does not represent private health spending separately from public), adding such a linkage would affect other spending and/or the total expenditure and revenue balance and therefore the level of taxation.

Second, we know that health affects the ability of populations to obtain education and the quality of that received. IFs includes enrolment terms to which health could be added. Third, health will affect distribution of income and economic well-being, although that relationship might be particularly difficult to represent.

Certainly there are other forward linkages of mortality and morbidity that the model might include.

9. Conclusion

The health module significantly adds to the overall utility of the IFs forecasting system by improving projections of mortality and other health related outcomes. Users can now explore age, sex, and cause-specific mortality and morbidity out to 2100 and consider how changing patterns might influence other demographic, social, and economic variables.

This document supports our intention to make the IFs model fully transparent to all users by providing technical documentation detailing model inputs, data sources, and regression equations. We welcome all comments and questions.

References

- Adams 1987. "[Smeed's Law: some further thoughts.](#)" *Traffic Engineering and Control* (Feb) 70-73.
- Alsan, Marcella, David E. Bloom, and David Canning. 2006. "The Effects of Population Health on Foreign Direct Investment Inflows to Low- and Middle-Income Countries," *World Development* 34(4): 613-630.
- Anand, Sudhir and Martin Ravallion. 1993. "Human development in poor countries: on the role of private incomes and public services," *Journal of Economic Perspectives* 7(1): 133–150.
- Ashraf, Quamrul H., Ashley Lester, and David N. Weil. 2008. "When Does Improving Health Raise GDP?" NBER Working Paper No. 14449. National Bureau of Economic Research, Cambridge, MA.
- Bidani, Benu and Martin Ravallion. 1997. "Decomposing social indicators using distributional data." *Journal of Econometrics* 77: 125–139.
- Bloom, David E., and David Canning. 2004. "Global Demographic Change: Dimensions and Economic Significance." NBER Working Paper No. 10817. National Bureau of Economic Research, Cambridge, MA.
- Blössner, Monika, and Mercedes de Onis. 2005. *Malnutrition: quantifying the health impact at national and local levels*. Geneva, World Health Organization. (WHO Environmental Burden of Disease Series, No. 12).
- Dargay, Gately, and Sommer 2007. "Vehicle Ownership and Income Growth, Worldwide: 1960-2030". Joyce Dargay, Dermot Gately and Martin Sommer, January 2007.
- Deaton, Angus, and Christina Paxson. 2000 (May). "Growth and Savings Among Individuals and Households." *The Review of Economics and Statistics* 82(2): 212-225.
- Desai, Manish A., Sumi Mehta, and Kirk R. Smith. 2004. "Indoor smoke from solid fuels: Assessing the environmental burden of disease." WHO Environmental Burden of Disease Series No. 4. Annette Prüss-Üstün, Diamid Campbell-Lendrum, Carlos Corvalán, and Alistair Woodward, series eds. World Health Organization, Geneva.
- Ezzati, Majid and Alan D. Lopez. 2004. "Smoking and oral tobacco use." In Majid Ezzati, Alan D. Lopez, Anthony Rodgers, and Christopher J.L. Murray, eds., *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva: World Health Organization, 883-957. Retrieved 4 Feb 2009, from <https://www.who.int/publications/cra/chapters/volume1/0883-0958.pdf>.
- Ezzati, Majid, Alan D. Lopez, Anthony Rodgers, Christopher J.L. Murray, eds. 2004. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva: World Health

Organization.

- Fernández-Villaverde, Jesús, and Dirk Kruegger. 2004 (September 14). "Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data," unpublished manuscript, University of Pennsylvania and University of Frankfurt. <http://www.dklevine.com/archive/refs450643900000000304.pdf>
- Fernández-Villaverde, Jesús, and Dirk Kruegger. 2005 (December 19). "Consumption over the Life Cycle: How Important are Consumer Durables?," unpublished manuscript, University of Pennsylvania and Goethe University. <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=8466457>
- Gakidou, Emmanuela, Shefali Oza, Cecilia Vidal Fuertes, Amy Y. Li, Diana K. Lee, Angelica Sousa, Margaret C. Hogan, Stephen Vander Hoorn, and Majid Ezzati. 2007. "Improving Child Survival Through Environmental and Nutritional Interventions: The Importance of Targeting Interventions Toward the Poor." *Journal of the American Medical Association* 298(16): 1876-1887.
- Hughes, Barry B. and Hillebrand, Evan E. 2006. "Exploring and shaping International Futures". Boulder, CO: Paradigm Publishers.
- Hughes, Barry B., Randall Kuhn, Cecilia Peterson, Dale Rothman, and Jose Solorzano. 2011. *Improving Global Health: Patterns of Potential Human Progress, Volume 3*. Paradigm Publishing and Oxford India.
- Hughes, Barry B. 2005. "Productivity in IFs." Pardee Center for International Futures Working Paper, University of Denver, Denver, CO. <http://www.ifs.du.edu/documents/reports.aspx>.
- James, W. Philip T., Rachel Jackson-Leach, Cliona Ni Mhurchu, Eleni Kalamara, Maryam Shayeghi, Neville J. Rigby, Chizuru Nishida, and Anthony Rodgers. 2004. "Overweight and obesity (high body mass index)." In Majid Ezzati, Alan D. Lopez, Anthony Rodgers and Christopher J.L. Murray, eds., *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva: World Health Organization, 959-1108.
- Jamison, Dean T., Jia Wang, Kenneth Hill, and Juan-Luis Londono. 1996. "Income, Mortality and Fertility in Latin America: Country-Level Performance, 1960 - 90." *Analisis Economico* 11(2): 219-261.
- Kelly, Christopher, Nora Pashayan, Sreetharan Munisamy, and Joshn W. Powles. 2009. "Mortality attributable to excess adiposity in England and Wales in 2003 and 2015: explorations with a spreadsheet implementation of the Comparative Risk Assessment methodology." *Population Health Metrics* 7(11): 1-7.
- Lopez, Alan D., Neil E. Collishaw, and Tapani Piha. 1994. "A descriptive model of the cigarette epidemic in developed countries." *Tobacco Control* 3(3): 242-247.

- Mathers, Colin D., and Dejan Loncar. 2005. "Updated Projections of Global Mortality and Burden of Disease, 2002-2030: Data Sources, Methods and Results." Evidence and Information for Policy Working Paper. World Health Organization, Geneva.
- Mathers, Colin D., and Dejan Loncar. 2006. "Projections of Global Mortality and Burden of Disease from 2002 to 2030." *PLoS Medicine* 3(11): e442, 2011-2030. Retrieved 13 March 2009. doi:10.1371/journal.pmed.0030442.
- Mathers, Colin D., and Dejan Loncar. 2006b. "New projections of global mortality and burden of disease from 2002 to 2030." Protocol S1. Technical Appendix to Mathers and Loncar 2006.
- Mathers, Colin D., and Dejan Loncar. 2006c. "Results of Regressions of Age-Sex-Specific Mortality for Detailed Causes on the Respective Cause Cluster Based on the Full Country Panel Dataset, 1950-2002." Technical Appendix to Mathers and Loncar 2006.
- Nixon, John, and Philippe Ulmann. 2006. "The Relationship Between Health Care Expenditure and Health Outcomes: Evidence and caveats for a Causal Link." *European Journal of Health Economics* 7: 7-18.
- Peto, Richard, Jillian Boreham, Alan D. Lopez, Michael Thun, and Clark Heath, Jr. 1992. "Mortality from Tobacco in Developed Countries: Indirect Estimation from National Vital Statistics." *Lancet* 339(8804): 1268-1278. doi:10.1016/0140-6736(92)91600-D.
- Ploeg, Martine, Katja K. H. Aben, and Lambertus A. Kiemeny. 2009. "The Present and Future Burden of Urinary Bladder Cancer in the World." *World Journal of Urology* 27(3): 289-293. doi:10.1007/s00345-009-0383-3.
- Shibuya, Kenji, Mie Inoue, and Alan D. Lopez. 2005. "Statistical Modeling and Projections of Lung Cancer Mortality in 4 Industrialized Countries." *International Journal of Cancer* 117(3): 476-485. doi:10.1002/ijc.21078.
- Smeed, RJ 1949. "Some statistical aspects of road safety research". [Royal Statistical Society](#), Journal (A) CXII (Part I, series 4). 1-24.
- Smith, Kirk R., Sumi Mehta, and Mirjam Maeusezahl-Feuz. 2004. "Indoor Air Pollution From Household Use of Solid Fuels," in Majid Ezzati, Alan D. Lopez, Anthony Rodgers, and Christopher J.L. Murray, eds., *Comparative Quantification of Health Risks*. Geneva: World Health Organization: 1435-1494.
- Smith, Lisa C. and Lawrence Haddad. 2000. "Explaining Child Malnutrition in Developing Countries: A Cross-Sectional Analysis." Washington, D.C.: International Food Policy Research Institute.
- Soares, Rodrigo R. 2007. "On the Determinants of Mortality Reductions in the Developing World." *Population and Development Review* 33(2): 247-287.
- United Nations Population Division. 2003. *World Population Prospects: The 2002 Revision, Highlight*. New York: United Nations. Department of Economics and Social Affairs.

United Nations Population Division. 2009. *World Population Prospects: The 2008 Revision, Highlights*. New York: United Nations. Department of Economics and Social Affairs.

Wagstaff, Adam. 2002. "Inequalities in Health in Developing Countries: Swimming Against the Tide?" Unpublished Manuscript

Appendix

Appendix Table 1 – Cardiovascular Beta Coefficients (IFs Project Re-estimation)

Sex	Age	Variable	Estimate
Female	30-44	Intercept	-0.56708
		LnGDPPCP	1.434976
		LnGDPPCP2	-0.08148
		LNSI	0.013229
		T	-0.02083
		LnEdYrsOver25	-0.26159
	45-59	Intercept	-4.14287
		LnGDPPCP	2.535495
		LnGDPPCP2	-0.14355
		LNSI	0.027787
		T	-0.0211
		LnEdYrsOver25	-0.06727
	60-69	Intercept	1.394147
		LnGDPPCP	1.557257
		LnGDPPCP2	-0.08679
		LNSI	0.01519
		T	-0.01879
		LnEdYrsOver25	-0.14311
	70+	Intercept	-2.42923
		LnGDPPCP	2.718893
		LnGDPPCP2	-0.16091
		LNSI	0.011117

		T	-0.01162
		LnEdYrsOver25	0.202314
Male	30-44	Intercept	-7.10273
		LnGDPPCP	2.706356
		LnGDPPCP2	-0.14545
		LNSI	0.024695
		T	-0.01319
		LnEdYrsOver25	-0.09248
	45-59	LnEdYrsOver25	-0.09248
		Intercept	-6.23361
		LnGDPPCP	2.829483
		LnGDPPCP2	-0.15151
		LNSI	0.061432
		T	-0.01513
		LnEdYrsOver25	0.021862
	60-69	Intercept	-1.48587
		LnGDPPCP	2.077141
		LnGDPPCP2	-0.11296
		LNSI	0.040964
		T	-0.01443
		LnEdYrsOver25	0.055058
	70+	Intercept	-0.41857
		LnGDPPCP	2.188262
		LnGDPPCP2	-0.12904

		LNSI	0.061957
		T	-0.01231
		LnEdYrsOver25	0.22355