

# **Methods for life expectancy and healthy life expectancy uncertainty analysis**

**Joshua A Salomon**

**Colin D Mathers**

**Christopher JL Murray**

**Brodie Ferguson**

*Global Programme on Evidence for Health Policy Working Paper No. 10*

**World Health Organization  
August 2001**

# 1. Introduction

WHO introduced healthy life expectancy as a summary measure of the level of health attained by populations in the World Health Report 2000 (1, 2). The World Health Report 2001 (3) reports estimates of total life expectancy (LE) and healthy life expectancy (HALE) by sex for the 191 Member States of WHO. Calculation of HALE for WHO Member States requires three inputs. First, life expectancy at each age is calculated using the standard life table approach (4, 5). Second, estimates of the prevalence of various states of health at each age are required (2, 6). Finally, a method of valuing this time compared to full health must be developed (6).

Part of the utility of LE and HALE is that each provides an easily interpretable summary of numerous different pieces of information – age-specific mortality rates in the case of LE, with the addition of information on non-fatal health outcomes in HALE. As such, they also represent the product of numerous sources of uncertainty. Presentation of the results in the World Health Report include intervals around the LE and HALE estimates in order to convey the levels of uncertainty in these estimates due to limitations in the data inputs. This discussion paper describes the methods used to quantify uncertainty in the various inputs to LE and HALE estimates, and how these uncertainties were incorporated in intervals around the reported measures.

Modern epidemiological techniques report confidence or uncertainty intervals around all estimates, and there is a growing literature on methods for capturing uncertainty in quantitative policy analyses (see, for example, Morgan and Henrion (7), King et al. (8), Vose (9)). Confidence intervals and uncertainty intervals provide explicit characterizations of the precision around estimates derived from limited information sources. For some simple quantities of interest, such as linear combinations of normally distributed random variables, uncertainty may be reasonably captured using analytic methods. For more complex quantities, however, a simulation-based approach offers a more feasible solution.

In the following sections, we introduce our general approach to propagating uncertainty in LE and HALE, then describe in detail the sources of uncertainty in each of the key components of these measures. We describe the methods used to combine the uncertainty from each input into the computations of HALE, and end with a discussion of continuing work and future directions in this area.

## 2. Background

There are a variety of types and sources of uncertainty. Uncertainty may arise from incomplete information (e.g., when we base estimates for a population on observations from a sample), from potential biases in information (e.g., how representative for a whole population are estimates from a study of a subgroup), from disagreements between information sources (e.g., when we have several studies giving different estimates for the same quantity), from model uncertainty (e.g., the type of function specified in a regression model), from uncertainty in preferences (e.g., in preferences for health states), or it may be inherent to the data generation process itself (e.g., we may only infer risks from event counts in a population, which means that we can never know the risks themselves with certainty).

Probability is often used to describe and quantify uncertainty. According to King, “what we commonly call *probability* is really a formal model of uncertainty” (10). However, there is

controversy about the nature of probability, and not all types of uncertainty may be appropriately expressed via probability (7). So we briefly describe here the interpretation and use of probability to describe uncertainty.<sup>1</sup>

The classical or *frequentist* view of probability defines the probability of an event occurring in a particular trial or experiment as the frequency with which it occurs in a long sequence of similar experiments. Thus, the probability of obtaining heads when tossing a coin is conceptualized as the limiting value of the sequence of frequencies generated in a long series of tosses of the coin. The problem with this approach is that for most quantities of real interest it is not clear what the (infinite) series of trials of similar events should be. For example, what is the probability that the coin in the example above is “unbiased.” We can examine the frequency of heads in a trial for any particular coin, but in generalizing to another coin the probabilities obtained, we must make assumptions about the probability the other coin is unbiased. To assess this probability, what is the correct set of coins from which to draw samples?

One approach has been to distinguish events whose probabilities are knowable through a series of experiments from those whose probabilities are “unknowable” or “uncertain” because there is no unique and operationalizable set of similar experiments, but this essentially limits the use of probabilities to games of chance. If, instead, we adopt a Bayesian view of probability, we interpret probability as the degree of belief a person has that an event will occur, given all the relevant information currently known to that person. Since different people will have different information, they may legitimately assign different probabilities to the same event: there is strictly no such thing as “the” probability of an event.

These subjective probabilities must obey all the same axioms and rules as frequentist probabilities, or they are not probabilities. In fact, the practice of statistical inference is usually unaffected by these conceptual distinctions and essentially the same models of probability may be applied (7, 8). Moreover, when an empirical series of data from trials become available, the Bayesian assessment of probability should converge to the frequentist assessment, assuming the Bayesian uses the data rationally to update the assessments.

Our general approach to describing and estimating uncertainty in quantities of interest is to express them as probability distributions using a Bayesian interpretation of probability as expressing uncertainty of an observed or hypothetical event given a set of assumptions about the world. Probability distributions can be used to express uncertainty about empirical or chance quantities representing information about states of the world (7). These might include epidemiological quantities such as the prevalence of depression in a particular population, or population preferences such as health state valuations or the weight to be given to different domains of health system responsiveness.

Estimates of quantities of interest to decision makers may also include other types of information where it may be inappropriate to express uncertainty by probabilities. These might include decision or control variables, which are typically under the control of a decision maker, or are specified in a regulatory framework. An example would be the theoretical minimum counterfactual distribution to be used in estimating the attributable burden from exposure to a risk factor such as alcohol, where there is uncertainty in the actual level of alcohol exposure which minimises adverse health outcomes, but where a particular counterfactual is chosen (without uncertainty) as a standard for analysis (11).

---

<sup>1</sup> This discussion of probability follows Morgan and Henrion (7).

It is also not usual or appropriate to take account of uncertainty in certain value parameters such as the rate of time discounting used in an analysis. Even if there is empirical evidence on population preferences for time discount rates, and uncertainty in these estimates, the choice of discount rate for use in analysis is essentially a social value judgement and should not include uncertainty (7). Although there is uncertainty about the social value judgment, or about its effect on decisions based on the analysis, it is usually preferable to vary the value deterministically in the analysis and to perform a *sensitivity analysis* to examine the impact on the outcomes of interest or on decisions. Thus, the Global Burden of Disease 1990 study (12) examined the sensitivity of the ranking of causes of burden of disease globally when discount rates and age weights were varied across a range of possible values.

One exception to this general principle relates to the health state valuations that are used to link mortality information with information on non-fatal health outcomes in summary measures of population health. We may conceptualize a health state in terms of levels on the various domains of health and a health state valuation as the result of weighing up these various domains in order to arrive at a summary assessment of the health level associated with this state. These valuations, unlike values such as time preference, do not have any clear normative basis; that is, while we might rely on philosophical arguments about intergenerational equity in choosing a discount rate, there are no obvious arguments as to the relative importance of mobility versus cognition in overall assessments of health levels. Thus, although health state valuations are often treated as value parameters without uncertainty (7), we have treated health state preferences as empirical estimates with uncertainty distributions in the estimation of HALE, where the uncertainty reflects uncertainty in valuations derived from health survey data. There are at least two sources of uncertainty that should be reflected in these values: (a) in calculating summary measures of health, WHO has chosen to base the health state valuations on average valuations derived from surveys in representative population samples, so that there is sampling and model uncertainty in what are the actual population preferences for health states; and (b) particularly for the weights used in burden of disease analysis, there is uncertainty in the distribution of health states associated with particular cause sequelae.

The most familiar and most commonly quantified kind of uncertainty arises from random error in direct measurement of a quantity. An estimate of an epidemiological quantity for a population will have uncertainty arising from the finite sample size used in the study as well as in random measurement error. The standard error of the sample mean for such a quantity specifies the distribution of uncertainty in our knowledge of the true mean value in the population (assuming there is no systematic error). There are standard statistical techniques for quantifying and reporting this uncertainty, such as standard errors and confidence intervals. Thus, it is well accepted in epidemiology that quantities should be reported with 95% confidence intervals. The nature of the uncertainty may be represented in parametric distributions such as the binomial distribution for repeated binary outcomes or the Poisson distribution for event counts. Based on the central limit theorem, the normal distribution often provides a reasonable approximation of the nature of uncertainty around different quantities.

Most measurement involves not only random (stochastic) error but also systematic error, arising from biases in the measurement instrument (e.g., unrepresentativeness of a sampling frame for a survey), or from inaccuracies in the assumptions used to infer the actual quantity from the available data (e.g., estimating prevalence of a disease for a country from studies of representative subpopulations). Estimation of the magnitude of systematic error is extremely difficult and of necessity involves a large element of judgement. Examination of historical measurements shows that there is a consistent tendency to underestimate systematic error,

perhaps because systematic error usually relates to sources of error which are unknown or about which little is known (otherwise it would have been addressed) (7).

It is common to ignore systematic error when estimating uncertainty, but this often results in substantial underestimation of the true uncertainty. It is often possible to put upper and lower bounds on the systematic error component, for example, where there are biological limits to a disease process, or where there is evidence from a range of populations of likely upper and lower limits to an epidemiological parameter such as prevalence or case fatality. In these cases, at the worst, the uncertainty distribution could be quantified by a uniform distribution between the upper and lower limits, or perhaps a more complex distribution that captures some information on our knowledge of the variation in probable values of the quantity. For these types of uncertainty, distributions such as the uniform distribution or the triangular distribution are often useful to describe the distribution of uncertainty.

### 3. General Methods

Our general approach to propagating uncertainty in estimates of LE and HALE relies on numerical simulation methods. The simulation approach uses multiple samples from probability distributions around uncertain inputs to allow estimates of the probability distributions around quantities of interest that may be complicated functions of these inputs, without the need to solve difficult (or in many cases, insoluble) mathematical functions.

The basic steps in simulation-based methods are as follows:

1. Define probability distributions around input parameters. More precisely, a joint distribution of all uncertain inputs should be specified in order to account for interdependencies of different input
2. Draw one value for each input from the joint distribution.
3. Calculate the quantities of interest based on this vector of sampled input values.
4. Repeat this procedure  $M$  times (where  $M$  is some number of draws that adequately reflects the information captured in the joint probability distribution).
5. Compute relevant summary statistics for the quantities of interest based on the set of  $M$  values for these quantities.

There are a number of different approaches to generating random samples from specified distributions (8, 9). The most straightforward method is Monte Carlo simulation, which starts from a random number generated from a uniform [0,1] distribution and uses the inverse cumulative distribution function at this value to map this random number into a draw from the specified probability density function. A commonly used alternative to Monte Carlo simulation is Latin Hypercube Sampling (LHS), which offers efficiency gains over Monte Carlo by using stratified sampling without replacement. The basis for LHS is to divide the distribution into equi-probable intervals and then to sample randomly from within each interval. More powerful methods based on Bayesian statistics allow the use of exact distributions rather than asymptotic normal approximations, but these methods are considerably more difficult to implement and will produce substantively similar results in many cases (8). Details on the different sampling methods used in the various steps of our calculations are described in the relevant sections below.

## 4. Life table uncertainty

Methods for estimating life tables for each WHO Member State are described elsewhere (4,5,13). In brief, a system of model life tables was developed based on the two-parameter Brass logit life table approach, with additional age-specific parameters to correct for systematic biases (13). This model life table system was used in order to update vital registration information on age-specific mortality rates for countries in which the most recent data were from years earlier than 2000, and to develop life tables for countries with no vital registration data available. For countries with time series information from vital registration, projections from the last available data point were made using regression methods for estimating time trends in the model parameters. For other countries with more limited vital registration information, these parameters were estimated based on trends in indirect information on child and adult mortality. For countries with no vital registration data, these indirect survey or census estimates were used to develop point estimates and ranges around child mortality (summarized as the probability of dying before the age of five, or  ${}_5q_0$  in the language of demography) and adult mortality (summarized as the probability that individuals who have lived to age 15 will die before reaching the age of 60, or in shorthand  ${}_{45}q_{15}$ ). These point estimates and ranges were then mapped through the model life table system into corresponding life tables.

For countries in each of the categories described above, there were several sources of uncertainty that produced uncertainty around the final life tables. For those countries with vital registration data projected using time series regression models on the parameters of the logit life table system, uncertainty around the regression coefficients has been accounted for by taking 1000 draws of the parameters using the regression estimates and variance covariance matrix of the estimators. For each of the draws, a new life table was calculated. In cases where additional sources of information provided plausible ranges around  ${}_5q_0$  and  ${}_{45}q_{15}$ , the 1000 draws were constrained such that each life table produced estimates within these specified ranges. The range of 1000 life tables produced by these multiple draws reflects some of the uncertainty around the projected trends in mortality, notably the imprecise quantification of systematic changes in the logit parameters over the time period captured in available vital registration data.

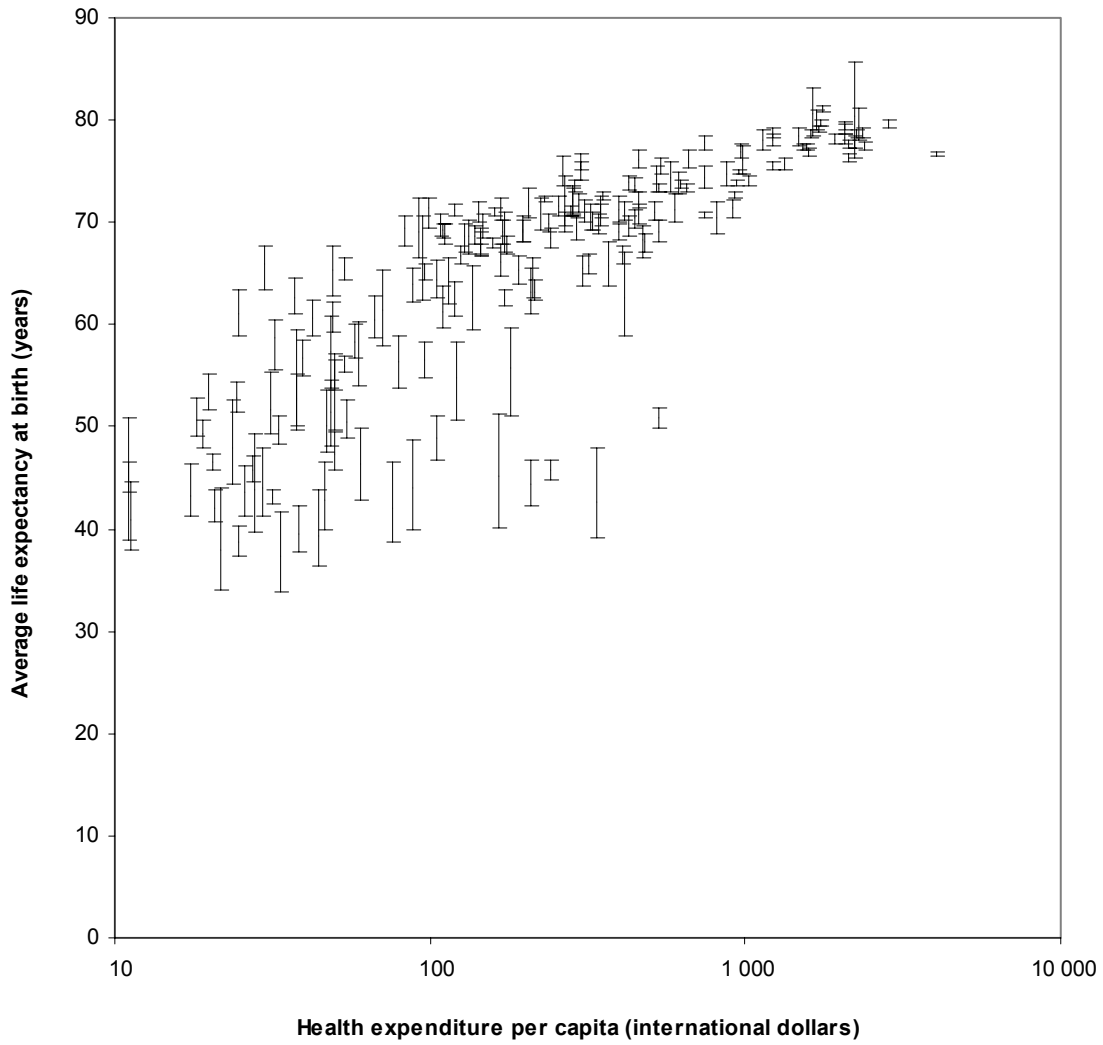
For countries that did not have time series data on mortality by age and sex, the following steps were undertaken. First, point estimates and ranges around  ${}_5q_0$  and  ${}_{45}q_{15}$  for males and females were developed on a country-by-country basis as described in Lopez et al. (5). In the modified logit life table system described by Murray and colleagues (13), values on these two parameters may be used to identify a range of different life tables in relation to a global standard life table. Using Monte Carlo simulation methods, 1000 random life tables were generated by drawing samples from normal distributions around these inputs with variances defined in reference to the defined ranges of uncertainty. In countries where uncertainty around  ${}_5q_0$  and  ${}_{45}q_{15}$  was considerable due to a paucity of survey or surveillance information we have sampled from wide distributions but then constrained the results based on estimates of the maximum and minimum plausible values for  ${}_5q_0$  and  ${}_{45}q_{15}$ . For each country, the results of this analysis were 1000 different simulated life tables which were then used to describe ranges around key indicators such as life expectancy at birth.

For a number of countries, concentrated particularly in sub-Saharan Africa, estimates of life tables were made by constructing counter-factual life tables excluding the mortality impact of the HIV/AIDS epidemic and then combining these life tables with exogenous estimates of the excess mortality rates attributable to AIDS. The estimates of AIDS-attributable mortality were based on back-calculation models (14) developed as part of collaborative efforts between

WHO and the Joint United Nations Programme on HIV and AIDS (UNAIDS) to derive country-level epidemiological estimates for HIV and AIDS. Given the dearth of data from which to estimate AIDS mortality directly and the uncertainties introduced by the back-calculation approach, it is important to try to quantify the level of uncertainty around the mortality estimates that result from these methods. Where enough data were available to undertake a maximum likelihood estimation approach in the back-calculation models (approximately 20 countries), the results included a measure of uncertainty around mortality estimates in each year. For the remaining countries, uncertainty intervals were derived based on an assessment of the coverage and representativeness of sentinel surveillance sites in each country. Probability distributions around the total number of deaths were then translated into distributions around age and sex-specific mortality rates using numerical simulation methods. By sampling 1000 draws from these distributions, uncertainty around AIDS mortality was incorporated into the uncertainty estimates in the life tables.

Annex Table 1 of the World Health Report 2001 gives 95% uncertainty ranges for adult ( $_{45}q_{15}$ ) and child ( $_{5}q_0$ ) mortality and for life expectancy at birth for males and females for WHO Member States (3). Average uncertainty in estimates of life expectancy at birth are shown for the 191 WHO Member States in Figure 1, plotted against health expenditure per capita (measured in international dollars for 1998 using purchasing power parity conversion factors).

**Figure 1. Uncertainty in average LE at birth (males and females combined) for the year 2000 versus average health expenditure per capita (1998) for 191 WHO Member States.**



## 5. Uncertainty in health state prevalences

The measurement of *time spent in poor health* is based on combining condition-specific estimates from the Global Burden of Disease 2000 study (GBD 2000), with estimates of the prevalence of different health states by age and sex derived from representative population health surveys (15). The health survey data derives from standardised surveys carried out using the WHO Health and Responsiveness Survey instrument and statistical methods to improve the cross-population comparability of the data (16, 17). Data from the GBD 2000 study have been combined with data from the surveys using Bayesian statistical techniques (6). The overall uncertainty in health state prevalences used to calculate HALE derives from uncertainty in the GBD-based prior estimates, uncertainty in the survey-based prevalence estimates, and uncertainty in the regression-based posterior estimates for countries without survey data.

### 5.1 Uncertainty in GBD prior estimates of health state prevalences

For the World Health Report 2001, burden of disease estimates have been updated for many of the cause categories included in the Global Burden of Disease 2000 study, and have been used to develop internally consistent estimates of prevalence YLD, for over 130 major causes, for 17 sub-regions of the 6 WHO regions of the world (15). These data, together with new and revised estimates of deaths by cause, age and sex, for all Member States, have been used to generate prior estimates of the severity-weighted prevalence of health states by age and sex for each WHO Member State (6).

The degree of uncertainty in the country-level weighted disability prevalences is mainly determined by levels of uncertainty in

- (a) epidemiological estimates for prevalence and/or incidence of disability associated with specific causes or cause groups,
- (b) disability weights arising from uncertainty in health state valuations and, in some cases, also in the disability severity distribution associated with a condition,
- (c) estimation of prevalence YLD at country level from the regional prevalence YLD rates,
- (d) estimation of prevalence YLD from incidence YLD, and
- (e) the approximate nature of adjustments for comorbidity.

For each WHO Member State, uncertainty in the country-level weighted disability prevalences was estimated for a subset of specific causes as summarised in Table 3. For the remaining specific causes, the overall level of uncertainty was assumed to be greater than that for the causes where detailed uncertainty estimates were made. For specific causes, where the uncertainty in the disability weights reflected uncertainty in health state valuations, this was also estimated. For some causes where the disability weight also reflects the distribution of health state severity, additional uncertainty in the severity distribution was also modelled.

High levels of uncertainty were specified for a number of residual categories, where there is a substantial burden of disease due to mortality. For these categories, provisional YLD estimates were based on the methods used in the 1990 Global Burden of Disease study (12), and a high level of uncertainty included (see Table 3).

A Monte-Carlo simulation (125 iterations) was run for the country-level weighted disability prevalences using @RISK<sup>®</sup> (18), an add-in software program to commercial spreadsheet packages. It allows the entry of uncertainty distributions instead of point estimates for input variables. It then recalculates the spreadsheet many times over, each time picking a value from

each of the specified distributions and produces an output dataset containing the resulting distribution of values for the output cells of interest.

**Table 3. Estimation of uncertainty in GBD prior estimates of comorbidity-adjusted prevalences by age and sex for WHO Member States in the year 2000.**

Category	Cause	Uncertainty estimation
Group I	Chlamydia	Uncertainty in regional prevalence and uncertainty in disability weight
	Gonorrhoea	Uncertainty in regional prevalence and uncertainty in disability weight
	HIV/AIDS	Uncertainty in country prevalence (AFRO D), other regional prevalences and uncertainty in disability weight
	Diarrhoeal diseases	Uncertainty in regional prevalence and uncertainty in disability weight
	Malaria	Uncertainty in regional incidence and uncertainty in disability weight
	Maternal causes	Uncertainty estimated from uncertainty in regional prevalence YLD rates together with estimated uncertainty in country/region YLD ratios
	Perinatal conditions	Uncertainty estimated from uncertainty in regional prevalence YLD rates together with estimated uncertainty in country/region YLD ratios
	Protein-energy malnutrition	Estimated from analysis of uncertainty in epidemiological estimates of prevalence of stunting, wasting and developmental disability
	Iron-deficiency anaemia	Uncertainty in regional prevalence and uncertainty in disability weight
Group II	Depressive episodes	Uncertainty in regional prevalence, uncertainty in disability weight, additional uncertainty in prevalences among children and at older ages
	Bipolar disorder	Uncertainty in regional prevalence and uncertainty in disability weight
	Schizophrenia	Uncertainty in regional prevalence and uncertainty in disability weight
	Dementia	Uncertainty in regional prevalence and uncertainty in disability weight
	Migraine	Uncertainty in regional prevalence and uncertainty in disability weight
	Hearing loss, adult onset	Uncertainty in regional prevalence and uncertainty in disability weight
	Ischaemic heart disease	Uncertainty in regional mortality estimates, uncertainty in case fatality rates and uncertainty in disability weights
	Stroke	Uncertainty in regional mortality estimates, uncertainty in case fatality rates and uncertainty in disability weights
	Osteoarthritis	Uncertainty in regional prevalence and uncertainty in disability weight
	Gout	Uncertainty in regional prevalence, severity and duration
	Low back pain	Uncertainty in regional prevalence, severity and duration
Group III	Road traffic accidents	For each of these injury causes, uncertainty estimated from uncertainty in regional YLD/YLL ratios, uncertainty in country mortality estimates and age-specific uncertainty in estimation of prevalence YLD from incidence YLD
	Falls	
	Other unintentional injuries	
	Violence	
Residual categories	Other perinatal causes	Relative uncertainty modelled by triangular distribution 0.9-1-2
	Other neuropsychiatric	Relative uncertainty modelled by triangular distribution 0-1-5
	Other cardiovascular	Relative uncertainty modelled by triangular distribution 0-1-5
	Other respiratory	Relative uncertainty modelled by triangular distribution 0-1-5
	Other digestive	Relative uncertainty modelled by triangular distribution 0-1-5
	Other genitourinary	Relative uncertainty modelled by triangular distribution 0-1-5
	Other musculoskeletal	Relative uncertainty modelled by triangular distribution 0-1-5
Other sources of uncertainty	Causes for which uncertainty not estimated	Aggregate relative uncertainty assumed to be equal to the aggregate relative uncertainty for the causes listed above multiplied by a triangular distribution 0.5-1-4
	Comorbidity adjustment	Uncertainty in level of dependent comorbidity and in adjustment to disability weights for comorbid conditions
	Adjustment to prior prevalence	Relative uncertainty modelled by triangular distribution 0-1-5

\*For a uniform distribution, every value in the specified range has an equal probability of being chosen in each iteration of the simulation. For a triangular distribution, the probability of being chosen rises linearly from zero at the minimum value, to a maximum at the most probably value, then falls linearly to zero at the maximum value.

These output datasets were used to estimate 95% uncertainty intervals for the age-sex specific weighted prevalences for WHO Member States.

The 95% uncertainty intervals are quite wide for any individual age-sex group, reflecting the total uncertainty in epidemiological estimates, health state valuations and in the adjustments for covariance and for residual cause categories. Because some of these sources of uncertainty are not dependent on age, and because uncertainty in age distributions of epidemiological estimates may be smaller than uncertainty in the overall level for the population, the uncertainty distributions for different age groups will have non-zero correlations. The correlation in age-specific uncertainty distributions was estimated for the age-sex specific prevalence YLD for selected WHO Member States in all regions and average correlation matrices estimated for Member States in three groups (an example for one group is presented in Table 4). Sensitivity analysis showed that the uncertainty distribution for HALE at birth was not strongly dependent on the level of correlation, so that it was acceptable to use the average correlation matrices.

**Table 4. Estimated correlation matrix for age-specific uncertainty distributions for prevalence YLD estimates for the AFRO and SEARO D regions**

Age	Males								Females							
	0-4	5-14	15-29	30-44	45-59	60-69	70-79	80-89	0-4	5-14	15-29	30-44	45-59	60-69	70-79	80-89
0-4	1.00								1.00							
5-14	0.13	1.00							0.09	1.00						
15-29	0.12	0.30	1.00						0.07	0.22	1.00					
30-44	0.15	0.29	0.38	1.00					0.09	0.22	0.29	1.00				
45-59	0.19	0.22	0.25	0.32	1.00				0.13	0.19	0.22	0.26	1.00			
60-69	0.17	0.17	0.18	0.26	0.38	1.00			0.13	0.15	0.16	0.21	0.29	1.00		
70-79	0.11	0.12	0.13	0.21	0.32	0.37	1.00		0.09	0.10	0.11	0.16	0.24	0.28	1.00	
80-89	0.05	0.09	0.10	0.15	0.21	0.28	0.36	1.00	0.04	0.07	0.07	0.11	0.16	0.19	0.25	1.00

## 5.2 Uncertainty in health survey data

To overcome the problem of cross-population comparability, the WHO survey instrument uses performance tests and vignettes to calibrate self reported health on selected domains such as cognition, mobility and vision. WHO is developing several statistical methods for correcting biases in self-reported health using these data, based on the hierarchical ordered probit (HOPIT) model (17). The calibrated responses, reflecting cross-population comparable estimates of levels on key domains of health, are then used in estimating the severity weighted prevalence of sub-optimal health by age and sex for 63 surveys in 55 Member States (6).

Levels along multiple domains of health are mapped into health state valuations based on empirical results from the same household surveys on the relative valuations of different health states reflecting combinations of different levels on the core health domains. In the surveys, individuals provide descriptions for a series of hypothetical health states along seven core domains of health, followed by valuations of these states using a simple thermometer-type (visual analogue) scale. Results on the visual analog scale are adjusted to correct for distortions in the scale using a series of more detailed surveys that include more abstract and

cognitively demanding valuation tasks. A valuation function has been estimated to describe the way in which levels on the core domains of health for a particular health state are translated into a valuation of that health state. This valuation function has been applied to the individual-level descriptions of their own health states, calibrated to enhance cross-population comparability as described above. The result of these computations is a set of estimates of the overall severity-weighted prevalence of sub-optimal health states each of the surveyed countries.

The estimated severity-weighted prevalences have been derived to reflect several different sources of uncertainty. First, multiple estimates of individual domain levels have been generated during the calibration procedure in order to capture uncertainties that arise in the mapping from categorical self-reported responses into continuous measures of domain levels. Estimation uncertainties in the fitting of the model parameters have been propagated by sampling from draws from the joint distribution of the coefficient estimators. The standard errors for the survey means also reflect the sampling uncertainty arising from the development of inferences about the entire population based on a random sample from it.

As discussed in Section 2 above, it is important to include uncertainty arising from systematic error so as not to underestimate total uncertainty in estimates. In the case of the surveys, the potential average systematic error was estimated by comparing the survey prevalences with the GBD-based prior prevalences for survey countries. Uncertainty arising from non-random causes such as sampling bias, or unknown systematic differences between the GBD disability weights and the health state valuation function, was estimated for postal surveys and for other surveys separately using least squares regression to estimate the root mean squared error of the survey estimates around the prior estimates.

### 5.3 Uncertainty in posterior health state prevalences

The prevalence estimates for Member States based on the GBD 2000 and the prevalence estimates for the Member States with health surveys were combined using Bayes' theorem to obtain posterior health state prevalences for all Member States (6). For the analysis of HALE for the year 2000, the uncertainty in both the evidence (survey mean severity-weighted prevalences by age and sex) and the prior means were assumed to be normally distributed. In that case, the posterior mean severity-weighted prevalence is given by the weighted sum of the survey mean and the prior mean as follows:

$$Pr\ ev_{Post} = w_1 * Pr\ ev_{Survey} + w_2 * Pr\ ev_{Prior}$$

where the weights are defined in terms of the standard deviation  $SD_1$  for the average survey prevalence (for a given age and sex) and the standard deviation  $SD_2$  for the prior prevalence (for the given age and sex) as:

$$w_1 = SD_2^2 / (SD_1^2 + SD_2^2)$$

$$w_2 = SD_1^2 / (SD_1^2 + SD_2^2)$$

The variance of the posterior estimate for countries with surveys is given by:

$$SD_{POST}^2 = SD_1^2 SD_2^2 / (SD_1^2 + SD_2^2)$$

The inter-age correlation for the survey estimates was assumed to be zero. Thus the inter-age correlation matrix for the posterior estimates was approximately estimated for each country by

adjusting the inter-age correlation matrix for the prior estimates (see example given in Table 4) by a country-specific factor reflecting the average prior weight across age groups.

Evidence from the surveys was also used to update the priors for non-survey countries. Least squares ordinary regression was used to fit the following model for the survey countries:

$$Prev_{Post} = \alpha + \beta * Prev_{Prior} + \delta_1 * POSTAL + \delta_2 * EUROBC$$

where *POSTAL* is 1 for postal surveys, 0 otherwise, and *EUROBC* is 1 for countries in the EURO region in mortality strata B and C (high adult mortality countries), 0 for countries in other regions. The fitted model was used to estimate posterior severity-weighted prevalences for all non-survey countries.

One of the assumptions of the classic linear regression model is that the independent variables are fixed in repeated measures. Stochastic error around the independent variables included in the regression will produce biased estimates of the regression coefficients, which is known as the “errors-in-variables” problem. Because practical analytical methods are generally lacking, it is almost universal to ignore this layer of uncertainty. We have addressed this problem, however, using a Monte Carlo procedure as follows.

Consider the regression model above for a specific age-sex group (subscripts suppressed for ease of presentation):

$$Y = X\beta + \varepsilon,$$

where *Y* is the posterior prevalence, *X* is the matrix of regressor variables, and  $\beta$  is the coefficient vector.

1. To account for the stochastic error in *X* (arising from the uncertainty in the priors) and the uncertainty in *Y* (arising from the combined uncertainties of the priors and the surveys), one observation of the prior value and the survey value was drawn for each survey country from the relevant uncertainty distributions (assumed to be normal). In drawing observations of prior values, it was assumed that one half of the uncertainty arose at regional level (so all countries in the same region shared the same draw) and the other half was country-specific. For each country, the posterior value *Y* was calculated and the regression model estimated. This process was repeated 10 times, resulting in 10 estimates of the coefficient vector and the stochastic error term.
2. For each of the 10 regressions, 10 draws of the coefficient vector were then made based on the variance-covariance matrix of the regression estimators, resulting in a set of 100 coefficient vectors.
3. For each of these 100 coefficient vectors, 10 random draws of the prior prevalence for each of the 191 WHO Member States were made and the regression model used to compute the posterior prevalences. This results in a sample of 1000 posterior estimates for the age-specific prevalences in each of the 191 WHO Member States. These 1000 posterior prevalences are sorted and the 2.5<sup>th</sup> percentile and 97.5<sup>th</sup> percentile of the distribution estimated, providing a 95% uncertainty interval for the posterior prevalence for each WHO Member State.

For the non-survey countries, inter-age correlation between posterior estimates was assumed to be the same as that for the prior estimates.

## 6. Uncertainty in HALE estimates

Sullivan's method was used to compute HALE for each Member State from the country life table and the severity-weighted prevalence estimates. Sullivan's method involves using the severity-weighted prevalence of health states (adjusted for comorbidity) at each age in the current population (at a given point of time) to divide the hypothetical years of life lived by a period life table cohort at different ages into years with and without disability. The method is illustrated in detail in Mathers et al. (19).

For each of the inputs to the Sullivan calculation, the measurement and estimation efforts have generated an uncertainty distribution. This uncertainty has been propagated forward into the computation of the overall healthy life expectancy. For each country, the distribution of each component was randomly sampled to generate 1000 draws. This process resulted in the compilation of component matrices (191 columns  $\times$  1000 rows), the column values of which represented all available information, for 191 countries, about the uncertainty surrounding each component of the HALE. Each component matrix was sampled a row at a time without replacement, and the overall HALE was computed 1000 times for each of the 191 countries. The HALE estimates for each Member State were then used to calculate for each age and sex, the median value and 95% uncertainty intervals (2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles) of HALE.

As discussed above, in sampling from the component uncertainty distributions, it is important to recognize inter-dependencies between multiple uncertain components. If we fail to model these inter-dependencies, the joint probability will be incorrect. Correlation between uncertainty in prevalences across age groups has been discussed above. These correlations were built into the prevalence draws by making random draws from a multivariate normal distribution with a correlation matrix appropriately specified for each country.

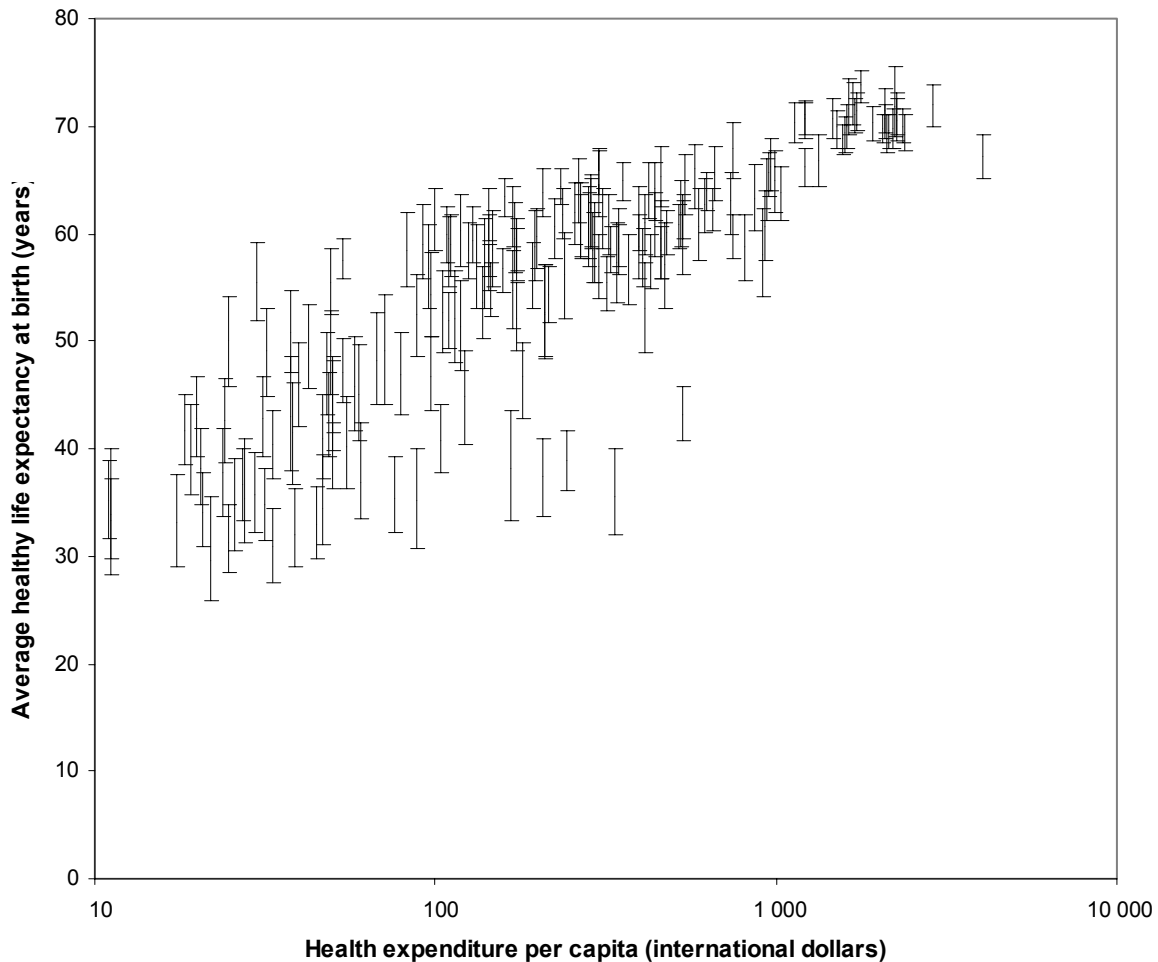
Because a proportion of the country-specific prior prevalences were derived making use of country-level cause-specific mortality information (6), there will be a correlation between uncertainty in life expectancy and the prevalences. This was taken into account by estimating the country-level prevalences for selected Member States using the complete set of life table draws for each, and computing the correlation between the age-standardized severity-adjusted prevalence and life expectancy at birth for males and females separately. A simple regression model was fitted to this data to estimate the corresponding correlation for other Member States as a function of life expectancy at birth. The correlation estimates ranged from around 0.2 in developed countries to around 0.3 to 0.4 in the African WHO region. For survey countries, the survey uncertainty was assumed to have zero correlation with life expectancy uncertainty, resulting in lower correlation estimates for these countries (typically around half the prior correlation).

The prevalence-mortality correlations were taken into account in selecting draws for input to Sullivan's method as follows. The prevalence draws were ranked in ascending order of age-standardized prevalence and the life table draws in ascending order of life expectancy at birth. We then drew from these two sets of draws to obtain paired prevalence and life table draws with the required rank order correlation using the methods of Vose (9). These pairs of draws were then used with Sullivan's method to compute a set of HALE estimates. These distributions were then used to estimate the 95% uncertainty intervals for HALE at birth and at age 60 years.

Annex Table 4 of the World Health Report 2001 gives 95% uncertainty ranges for HALE at birth and at age 60 for males and females for all WHO Member States (3). Average uncertainty ranges in estimates of healthy life expectancy at birth (both sexes combined) are shown for the 191 WHO Member States in Figure 2, plotted against health expenditure per

capita (measured in international dollars for 1998 using purchasing power parity conversion factors). Average 95% uncertainty intervals for Member States in each of the 17 epidemiological regions of the GBD 2000 are shown in Figure 3 together with average healthy life expectancy at birth for these regions.

**Figure 2. Uncertainty in average healthy life expectancy at birth (males and females combined) for the year 2000 versus average health expenditure per capita (1998) for 191 WHO Member States.**

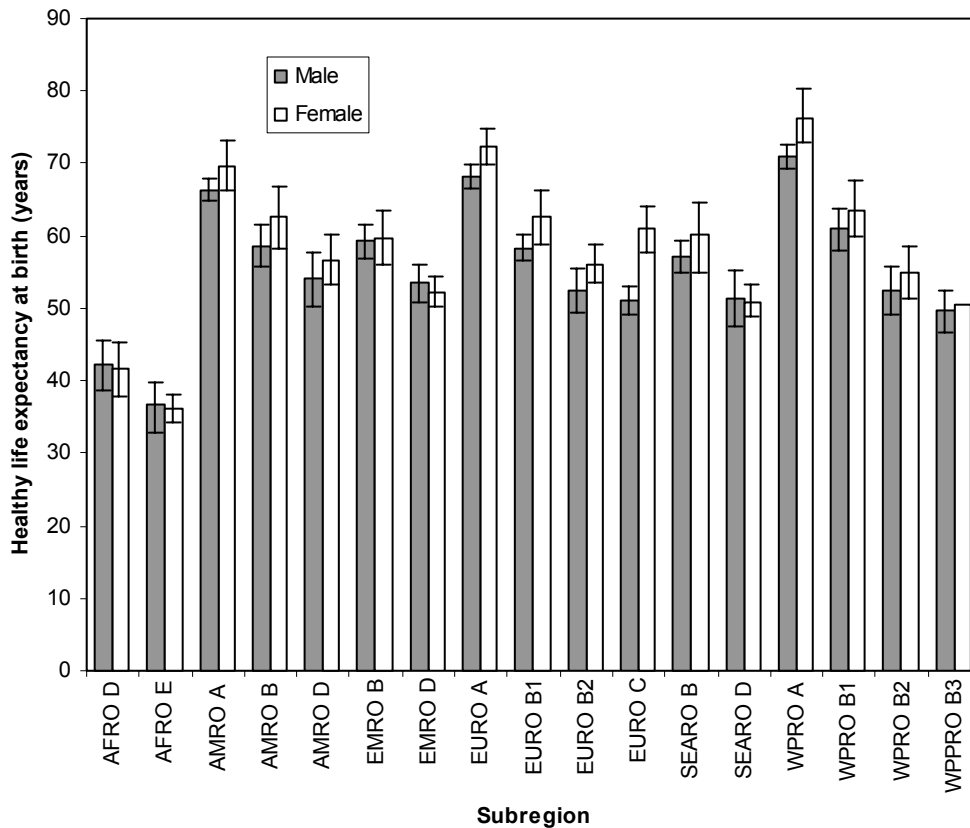


The uncertainty ranges for HALE in 2000 are larger than those estimated for DALE in 1999 as published in the World Health Report 2000 (1). There are two main reasons for this: (a) more detailed analysis of uncertainty in GBD-based priors for 2000, particularly for health state valuation uncertainty, uncertainty in residual categories and co-morbidity, and (b) estimation of non-random uncertainty due to survey data. Additionally, the World Health Report 2001 reports 95% uncertainty intervals rather than the 80% intervals reported in the World Health Report 2000.<sup>2</sup>

<sup>2</sup> In the Annex Notes of the World Health Report 2001, the ranges are erroneously described as 80% intervals.

Apart from the more detailed uncertainty analyses carried out for the 2000 estimates, which included more sources of uncertainty and improved estimates of its magnitude, the methods used to deal with correlation between components of uncertainty were also more sophisticated. For the 1999 analysis, a more conservative set of approximations was used as follows: we assumed 100% correlation between uncertainty at each age within broad age ranges 0-14, 15-29, 30-44, 45-59, 60-69 and 70+ years (so that for a given sample of the disability prevalence distribution, it is high at all ages or low at all ages within one of these ranges).

**Figure 3. HALE at birth with 95% uncertainty intervals, by sex, 17 epidemiological regions, 2000.**



## 6. Conclusions

Modern epidemiological techniques report confidence or uncertainty intervals around all estimates. Confidence intervals are based on the variation observed in sample data while uncertainty intervals are based on quantifying the possible uncertainty in the components of the estimates, including measurement uncertainty and other sources of uncertainty. This paper outlines the statistical methods used to estimate uncertainty intervals for life expectancies and healthy life expectancies in the World Health Reports 2000 and 2001 (1,3). Such uncertainty has also been carried through into estimates of the uncertainty in other quantities for which healthy life expectancies are inputs.

We reiterate that there is uncertainty in all estimates of epidemiological and demographic indicators at the population level. To the extent possible, WHO aims to utilise and synthesise

within a consistent and comprehensive framework, *all* relevant epidemiological evidence on population demography and health for the various regions of the world. Where the evidence is uncertain or incomplete, WHO attempts to make the best possible inferences based on the knowledge base that is available, and to assess the uncertainty in the resulting estimates.

A simulation-based approach to uncertainty propagation allows for an accurate representation of multiple sources of uncertainty, and also simplifies the process of accounting for correlations between the various sources of uncertainty. There is still substantial room for improvement in methods and estimates, and over the next year, efforts to extend and improve the analysis of uncertainty in both the GBD 2000 and in the survey program will continue.

While every effort is being made to reduce uncertainty in estimates, we are also attempting to ensure that we do not underestimate the true level of uncertainty by taking inadequate account of some non-stochastic sources of uncertainty. While the uncertainty intervals reported in this year's World Health Report may appear relatively large, it should be remembered that they include sources of uncertainty that will vary little from year to year (e.g., for many epidemiological quantities and for health state valuations), so that when auto-correlation is taken into account, the uncertainty in change from year to year in healthy life expectancy will be much less than the total uncertainty intervals for the annual estimates themselves.

## References

1. World Health Organization. *World Health Report 2000. Health systems: improving performance*. Geneva, World Health Organization, 2000.
2. Mathers CD, Sadana R, Salomon J, Murray CJL, Lopez AD. *Healthy life expectancy in 191 countries, 1999. The Lancet* 2001, **357**: 1685-1691.
3. World Health Organization. *World Health Report 2001. Mental Health: New Understanding, New Hope*. Geneva, World Health Organization, 2001.
4. Lopez AD, Salomon J, Ahmad O, Murray CJL. *Life tables for 191 countries: data, methods and results* (GPE discussion Paper No. 9). Geneva, World Health Organization, 2000.
5. Lopez AD, Ahmad O, Guillot M, Inoue M, Ferguson B. *Life tables for 191 countries for 2000: data, methods, results* (GPE Discussion Paper No. 40). Geneva, World Health Organization, 2001.
6. Mathers CD, Murray CJL, Salomon J, Lopez AD. *Estimates of healthy life expectancy for 191 countries in the year 2000: methods and results* (GPE discussion paper No. 38). Geneva, World Health Organization, 2001.
7. Morgan MG, Henrion M. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge, Cambridge University Press, 1990.
8. King G, Tomz M, Wittenberg J. Making the most of statistical analyses: improving interpretation and presentation. *American Journal of Political Science* 2000, **44**(2):341-355.
9. Vose D. *Risk analysis: a quantitative guide*, New York, Wiley, 2000.
10. King G. *Unifying political methodology: the likelihood theory of statistical inference*. Michigan, The University of Michigan Press, 1998.

11. Murray CJL and Lopez AD. On the quantification of health risks: lessons from the Global Burden of Disease Study. *Epidemiology* 1999, **10**(5): 594-605.
12. Murray CJL, Lopez, AD (eds.). *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020* (Global Burden of Disease and Injury Series, Vol. 1). Cambridge, Harvard University Press, 1996.
13. Murray CJL, Ferguson B, Lopez AD, Guillot M, Salomon J, Ahmad O. *Modified-logit life table system: principles, empirical validation and application* (GPE Discussion Paper No. 39). Geneva, World Health Organization, 2001.
14. Salomon JA, Murray CJL. Modelling HIV/AIDS epidemics in sub-Saharan Africa using seroprevalence data from antenatal clinics. *Bulletin of the World Health Organization* 2001, **79**(7): 596-607.
15. Murray CJL, Lopez AD, Mathers CD, Stein C. *The Global Burden of Disease 2000 project: aims, methods and data sources* (GPE Discussion Paper No. 36). Geneva, World Health Organization, 2001.
16. Üstün TB, Chatterji S, Villanueva M, Bendib L, Sadana R, Valentine N, Ortiz J, Tandon A, Salomon J, Yang C, Xie Wan J, Murray CJL. *WHO Multi-country Household Survey Study on Health and Responsiveness, 2000-2001* (GPE Discussion Paper No. 37). Geneva, World Health Organization, 2001.
17. Murray CJL, Tandon A, Salomon J, Mathers CD. *Enhancing cross-population comparability of survey results* (GPE Discussion Paper No. 35). Geneva, World Health Organization, 2000.
18. Palisade. *@RISK: advanced risk analysis for spreadsheets*. New York, Palisade Corporation, 1996.
19. Mathers C, Sadana R, Salomon J, Murray CJL, Lopez AD. *Estimates of DALE for 191 countries: methods and results* (GPE Discussion Paper No. 16). Geneva, World Health Organization, 2000.