Estimating age specific mortality: a new model life table system with flexible standard mortality schedules

Author affiliation:

Haidong Wang (Ph.D.)
University of Washington, Seattle, United States
haidong@uw.edu

Alan D. Lopez (MS,Ph.D.)
University of Queensland, Australia
a.lopez@sph.uq.edu.au

Christopher JL Murray (MD, DPhil)
University of Washington, Seattle, United States
cjlm@uw.edu
To estimate the age pattern of mortality for populations where vital registration information is not available, a common procedure is to estimate the mortality rates at one or several ages and to extend the estimate to all ages through the use of a “model” life table. A model life table is a set of age-specific death rates that together are believed to represent a particular “level” of mortality. Sets of model life tables are collections of the models arrayed by level of mortality. The three most widely used model life table systems are the Coale-Demeny model life tables, the United Nations model life tables for less developed countries, and the Modified Life Table system as used by the World Health Organization. However, most of the current models have shortcomings that limit their applicability. The principal potential weakness is inaccurate or unrepresentative data in the observed life tables that form the empirical basis of the models. This weakness is particularly serious when dealing with countries with a high prevalence of HIV/AIDS, since this disease has manifested itself only recently and primarily in countries that lack adequate systems of measurement. As a result, most current model life table systems often yield implausible results in high mortality situations. For example, when applying the Modified Logit Life Table system to a set of child and adult mortality estimates for 187 countries between 1970 and 2010, we find that 0.11% of the age specific mortality estimates are below zero. In addition, although most current model life table systems are multi-dimensional and use multiple entry parameters such as $s_0$ and $e^{10}$ (life expectancy at age 10), few can directly accommodate broad indicators of adult mortality such as $45q_{15}$, the probability of dying between ages 15 and 60. As both $s_0$ and $45q_{15}$ become more readily available for most countries, a model life table system that utilizes both indices directly is an urgent need.

We propose that an ideal model life table system should have the following desirable attributes. First, a model life table system should be parsimonious and require only a few entry parameters to generate a full life table. Second, it should adequately capture the range of age patterns of mortality observed in real populations and yield high predictive validity, not just measured by summary indices such as life expectancy at birth, but more importantly by age specific mortality rates. Third, it should provide satisfactory estimates of age specific mortality for countries with high levels of mortality, especially those plagued by the HIV/AIDS epidemic. Finally, a model life table should generate age specific mortality with a plausible time trend, and the partial derivative of entry parameters such as $s_0$ and $45q_{15}$ should be positive with respect to age specific mortality. These are the principles that our new model is built upon.

Empirical life table database

The performance of a model life table system depends partly on the empirical life table database it is based upon. Our new model life table system is built upon a database with 7,294 empirical life tables. Two major data sources provide about 95.4% of these life tables. The Human Mortality Database (HMD) (www.mortality.org) has over 8,400 life tables from 37 countries or areas dating back to 1751. 4,163 life tables from this database dated between 1950 and 2008 are included in our database. We excluded life tables of poor quality as assessed by the Human Mortality Database. The other major source of empirical life tables is the collection of vital registration data adjusted using death distribution methods revised by Murray and colleagues (2010). In total, we include 2,799 life tables from various vital
registration sources adjusted by the death distribution methods. We also include 14 life tables from Matlab, Bangladesh, and 19 life tables from US counties grouped by HIV seroprevalence. Provincial level data from South Africa, 1999-2005, are also included in our database. These 115 life tables have HIV seroprevalence ranges from 0.5% to 18.4% based on estimates provided by Actuarial Society of South Africa.

Our empirical life table database covers 84 countries and areas over the period of 1950 to 2008. Almost half (46.4%) of our life tables come from developed countries of Western Europe, North America, Australia, and Asia. A significant advantage of our database is the inclusion of more life tables from developing countries than any other model life table system. We have included 532 life tables from developing countries in Asia, among which 42 life tables are from the India Sample Registration System (earlier life tables from SRS only go up to age 70+; we apply a model, to be described in the method section, to extrapolate age specific mortality rates up to age group 80-84). Another 1,021 life tables from Latin American countries are also included. Perhaps most importantly, we have included 237 life tables from Sub-Saharan Africa. These life tables come from Botswana (2 life tables), Mauritius (90 life tables), South Africa (143 life tables which include 115 province level life tables), and Zimbabwe (2 life tables). Among the life tables included in our database, 1,153 (15.8%) are from populations with over 0.1% HIV seroprevalence among adults with ages 15 to 59.

Method

Our model is described by the following equation:

\[
\log it\hat{q}_t^{5g} = \log it\hat{q}_t^{5c} + \hat{\beta}_1 \cdot (\log it\hat{q}_0^{5g} - \log it\hat{q}_0^{5c}) + \hat{\beta}_2 \cdot (\log it\hat{q}_{15}^{5g} - \log it\hat{q}_{15}^{5c}) + \hat{\gamma}_1 \cdot (\log it\hat{q}_0^{5g} - \log it\hat{q}_0^{5c}) + \hat{\gamma}_2 \cdot (\log it\hat{q}_{15}^{5g} - \log it\hat{q}_{15}^{5c})
\]

(1)

In the above equation, \( s_5q_0 \) is the probability of dying from birth to age 5; \( 45q_{15} \) is the probability of dying from age 15 to 60; \( g \) represents sex; \( x \) represents age \( (0,1,5,10,...,80) \); \( n \) is the length of the age group \( x \) to \( x+n \); and \( nq_x \) indicates probability of dying in the age group \( x \) to \( x+n \). This equation proposes that the logit transformed age specific probability of dying in a target life table \( p \) can be represented as a function of the corresponding logit transformed age specific probability of dying in a standard life table \( s \) and the differences in probability of dying from age 0 to 5 in logit scale and the difference in probability of dying from age 15 to 60 in logit scale among two pairs of life tables: \( c \) & \( s \), and \( p \) & \( c \). Life table \( c \) is the estimated counter-factual life table when HIV seroprevalence is zero. When there is no HIV/AIDS epidemic in a population, life tables \( p \) and \( c \) are identical. The model is based on an empirical observation where the differences in age specific probabilities of dying in logit scale between two life tables are highly correlated with differences in \( s_5q_0 \) or \( 45q_{15} \) in logit scale. Coefficients \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are estimated using equation 2 with zero-HIV life tables in our database.

\[
\log it\hat{q}_t^{5g} - \log it\hat{q}_t^{5c} = \beta_1 \cdot (\log it\hat{q}_0^{5g} - \log it\hat{q}_0^{5c}) + \beta_2 \cdot (\log it\hat{q}_{15}^{5g} - \log it\hat{q}_{15}^{5c}) + \xi_t
\]

(2)

Proper selection of the standard life table is crucial. Instead of using a single standard life table for each sex, we use country-time specific and region (i.e. global burden of disease region) specific standard for
each life table not affected by HIV/AIDS in our database (other aggregated standard life tables by different geographical or epidemiological clustering criteria are also possible). Country-time specific standard life tables are used whenever an empirical life table from the same country within a 15-year time frame is available in our database. 69,265 unique pairs of life tables are matched using our life table database with time lags between 1 and 15 years. Region specific standard life tables are generated by collapsing all zero-HIV life tables in our database from the same global burden of disease region by sex. We then pair up all zero-HIV life tables in our database with the generated region specific life tables. This gives us an additional 5,976 pairs of life tables.

The values of $s_0$ and $s_{15}$ serve as points of entry (or entry parameters) for this new model life table system. To generate a standard life table (model age pattern of mortality) from our empirical life table database, we calculate the Mahalanobis distance between the target life table and all zero-HIV empirical life tables of the same sex in our database based on $s_0$ and $s_{15}$ (in logit scale). The Mahalanobis distance between two sets of $s_0$ and $s_{15}$ are defined as:

$$D_{ij} (Q') = \sqrt{(O' - O)^T S^{-1} (O' - O)}$$

where $O$ is a multivariate vector representing entry parameters $s_0$ and $s_{15}$ in logit scale. $Q' = (logit(s'_0), logit(s'_{15}))$ is a multivariate vector that corresponds to an empirical life table $i$ in our life table database. We choose Mahalanobis distance over Euclidean distance due to the fact that $s_0$ and $s_{15}$ are highly correlated in logit space (the correlation coefficients are 0.58 and 0.87 for males and females, respectively), and Mahalanobis distance takes the covariance matrix of $s_0$ and $s_{15}$ in logit scale into consideration when calculating the distance between any pair of life tables. We simply average the $l_x$ series (probability of surviving from age 0 to age $x$) from the 30 most similar life tables as evaluated by the Mahalanobis distance to generate a standard life table. Priority is given to life tables from the same country or global burden of disease region. The counterfactual life table, $c$, or the final life table, $p$, when HIV/AIDS is non-existent in the population, is estimated using a standard life table, $s$, which is generated using the aforementioned.

Currently, perhaps the most challenging issue for all existing model life table systems is providing plausible age specific mortality estimates given an HIV/AIDS epidemic. Both widely used model life table systems, the Coale-Demeny model life tables and the Modified Logit Life Table system, are largely based on empirical life tables from the pre-HIV era. In addition, they do not provide an integrated solution to the problem of incorporating HIV/AIDS. The 1,153 empirical life tables in our database with at least 0.1% HIV seroprevalence provide a unique opportunity for us to “directly” estimate the impact of HIV/AIDS on age specific mortality. Here, the impact of HIV/AIDS on age specific mortality, as measured by coefficients $\hat{\gamma}_{50}^{15}$ and $\hat{\gamma}_{50}^{0}$, is estimated using Equation 3, where we model the difference in age specific mortality in logit scale between an observed life table, $p$, affected by HIV/AIDS and a counter-factual life table, $c$, using the difference in $s_0$ and $s_{15}$ (also in logit scale) between the same pair of life tables. We generate an “HIV-free” base life table, $c$, using counterfactual $s_0$ and $s_{15}$ values and estimated parameters from Equation 2.
The counterfactual child mortality and adult mortality indices, $s_0$ and $s_{45}$, respectively, are estimated using a set of hierarchical mixed effects models as described in Equations 6 and 7.

In Equations 6 and 7, $j$ refers to country, $g$ refers to sex, and $t$ refers to time between 1970 and 2010. $Maternal.Edu_j$ is the average maternal education level for country $j$ at time $t$; $Ave.Edu_j$ is the average education level between age 15 and 60 for sex $g$ of country $j$ at time $t$; $GDP_j$ is the gross domestic product measured in international dollars for country $j$ at time $t$; and $HIVsero_j$ is the HIV seroprevalence for country $j$ at time $t$. We compiled a database for 191 countries with estimates of $s_0$ (Rajaratnam et al. 2010a), $s_{45}$ (Rajaratnam et al. 2010b), GDP in international dollars (Institute for Health Metrics and Evaluation 2010), maternal education (Gakidou et al. 2010), age specific level of education (Gakidou et al. 2010), and HIV seroprevalence (the Joint United Nations Programme on HIV/AIDS (UNAIDS) 2008) for years between 1970 and 2010. A 3-year lag is applied to the HIV seroprevalence variable to better capture the effect of HIV/AIDS. With the estimated coefficients, $\hat{\beta}_j$ and $\hat{\eta}_j$, the HIV-counterfactual $s_0$ and $s_{45}$ could easily be calculated using equations 5 and 6.

To illustrate the effect of HIV/AIDS on the age pattern of mortality, we use the life table for South African males in 2004 as an example. Using our new model life table system described in Equation 1, we first generate a base life table (c) using the counterfactual entry parameters where HIV/AIDS prevalence is set to zero. We then estimate age specific mortality given different levels of HIV/AIDS epidemic. The output from the Modified Logit Life Table system for the same life table is also included. Our estimated age specific mortality closely resembles both level and age pattern of the observed mortality. The age differentials of HIV/AIDS impact are precisely captured.