BRIEF SUMMARY
of the
Methods Protocol for the Human Mortality Database

J.R. Wilmoth, K. Andreev, D. Jdanov, and D.A. Glei
with the assistance of
C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, P. Vachon[1]

The Human Mortality Database (HMD) contains uniform death rates and life tables (e.g., life expectancy) for various populations. It also includes the original raw data (i.e., births, deaths, census counts or official population estimates) from which they were derived. The following comprises a brief summary of the methodology used to calculate the HMD life tables. For a complete description, see the full Methods Protocol available at:
    www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf

For more information about the format of HMD data files, see the explanation given at:
    www.mortality.org/Data/ExplanatoryNotes

Steps for computing mortality rates and life tables

There are six steps involved in computing mortality rates and life tables for the HMD. Here is an overview of the process:

1. Births. Annual counts of live births by sex and by month are collected for each population over the longest time period available. At a minimum, a complete series of birth counts by sex is needed for the time period over which mortality rates and period life tables are computed. These counts are used mainly for estimating the size of individual cohorts (on January 1st of each year) from their birth until the next census, and for other adjustments based on relative cohort size. The birth-by-month data are used (whenever available) for more accurate estimates of population exposures.

2. Deaths. Death counts are collected by sex, completed age, year of birth, and year of death if available. The raw death counts for a given calendar year may be available by completed age (or age group) but not by year of birth. Before making subsequent calculations, deaths of unknown age may be distributed proportionately across the age range, and aggregated deaths are split into finer age categories.[2]

3. Population size. Below age 80, estimates of population size on January 1st of each year are either obtained from another source (most commonly, official estimates) or derived using intercensal survival[3] as depicted in Figure 1 (shown in red). In most cases, all available census counts are collected for the time period over which mortality rates and life tables are computed. The maximum level of age detail available in the raw data is always retained and used in subsequent calculations. When necessary, persons of unknown age are distributed

---

[1] This document is the direct result of a series of discussions held in Rostock, Germany, and Berkeley, U.S.A., beginning in June 2000. The individuals on this list not only participated in those discussions but also made important contributions to the set of methods described here. All graphs in the main text were drawn by Georg Heilmann.

[2] For more details regarding the methods for redistributing deaths of unknown age and splitting aggregated deaths into finer age categories, see pp. 10-14 of the Methods Protocol.

[3] For details about the method of intercensal survival, see pp. 15-29 of the Methods Protocol.

proportionately across age before making subsequent calculations such as intercensal population estimation and calculation of mortality rates. Above age 80, population estimates are derived by the method of extinct generations for all cohorts that are extinct[4] (shown in green in Fig. 1) and by the survivor ratio method[5] for non-extinct cohorts who are older than age 90 at the end of the observation period (shown in blue in Fig. 1). For non-extinct cohorts aged 80 to 90 at the end of the observation period, population estimates are obtained either from another source or by applying the method of intercensal (or postcensal) survival.

Figure 1. Methods used for population estimates

A - Official estimates / intercensal survival
B - Extinct cohorts
C - Survivor ratio, SR90+
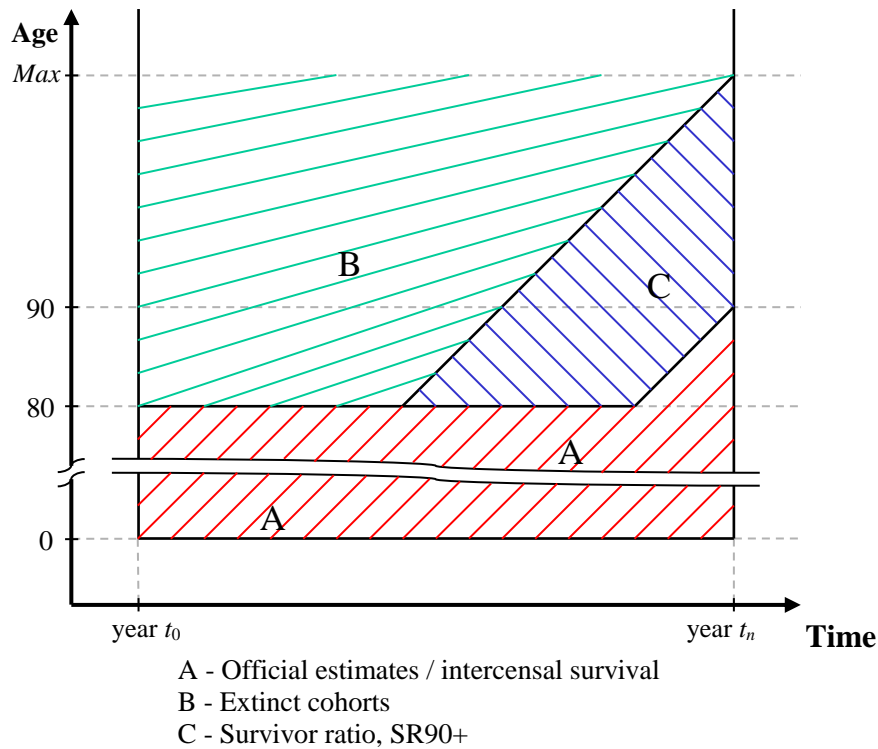
4. Exposure-to-risk. Estimates of the population exposed to the risk of death during some age-time interval are based on annual (January 1st) population estimates, with a small correction that reflects the timing of deaths during the interval and **variation in cohort's birthdays by month**.[6]

5. Death rates. For both periods and cohorts, death rates are simply the ratio of deaths to exposure-to-risk in matched intervals of age and time. For broader intervals of age and/or time (whether time is defined by periods or cohorts), death rates are always found by pooling deaths and exposures first and then dividing the former by the latter.[7]

6. Life tables. Period death rates are converted to probabilities of death by a standard method.[8] Cohort probabilities of death are computed directly from estimates of deaths and exposure,

---

[4] For details regarding the extinct cohort method, see paragraph 5.3, pp. 25-26 of the Methods Protocol.
[5] For more information regarding the survivor ratio method, see paragraph 5.4, pp. 26-29, in the Methods Protocol.
[6] See pp. 29-33 of the Methods Protocol for the formulas for calculating exposure estimates for periods (Equations 50-57) and cohorts (Equations 60-62).
[7] For details regarding the calculation of death rates, see paragraph 6.1, pp. 29-31 of the Methods Protocol.
[8] See paragraph 7.1, pp. 34-38 of the Methods Protocol for the formulas used to calculate period life tables.

but they are related to cohort death rates in a consistent way.[9] These probabilities of death are used to construct life tables.

Raw data versus estimates

Raw data (provided at the bottom of each country page) are the original data received from national statistical offices or other sources. The availability and format of these data varies somewhat across populations and over time. Therefore, it is important to note that the uniform data shown on country pages by single years of age (from age 0 to 109, with an open age interval for ages 110+) are sometimes estimates derived from aggregate data (e.g., five-year age groups, open age intervals such as 90+). As noted earlier, we frequently need to split the original data into finer age categories. In addition, the raw data often include persons of unknown age (in either death or census counts), which we redistribute (proportionally) across the age range. More information regarding the format of the raw data for a given population can be found in the *Background and Documentation* file (see *Appendix*), accessible from the country page.

Female / male / total

Raw data for women and men are always pooled when making calculations for the total population. Thus, "total" death rates and other quantities are not a simple average of the separate values for females and males, but rather a weighted average reflecting the relative size of the two groups at a given age and time.

Period versus cohort

Raw data are usually obtained in a period format (i.e., by the year of occurrence). In the HMD, most data are presented in a period format, but we also provide death rates and life tables in a cohort format (i.e., by year of birth) if the observation period is sufficiently long to justify such a presentation.

Period life tables

Whereas a cohort life table depicts the life history of a specific group of individuals (born in the same year or range of years), a period life table represents the mortality conditions at a specific moment in time. Observed period death rates are only one result of a random process for which other outcomes are possible as well. This inherent randomness is most noticeable at older ages where observed death rates may exhibit large fluctuations. Therefore, we smooth the observed values in order to obtain an improved representation of the underlying mortality conditions.[10] The smoothed death rates are used for subsequent calculations only above some very high ages. The cutoff age is based on a decision rule that involves both age and number of survivors in a particular population. As always, death rates (original at younger ages, smoothed above some age) are then converted to probabilities of death (q_x), and all remaining life table quantities (i.e., l_x, d_x, L_x, T_x, e_x) are derived from the q_x values.[11]

Cohort life tables

Cohort life tables represent the mortality experience of the group of individuals who were born in the same year (or range of years for multi-year cohorts). We compute cohort probabilities

---

[9] See paragraph 7.2, pp. 38-39 of the Methods Protocol for the formulas used to calculate cohort life tables.

[10] For details regarding the smoothing process, see pp. 34-36 of the Methods Protocol.

[11] See pp. 36-38 for the formulas used to derive life table quantities.

of death (q_x) directly from the data and perform no smoothing at older ages.[12] From these q_x values, a complete cohort life table is calculated using the same formulas used for period tables.

Almost-extinct cohorts

The above description assumes that all members of a cohort have died before we compute the life table. Yet, it is often desirable to compute life tables for cohorts that are *almost* extinct (see definition of almost extinct below). In such cases, we must make some assumption about future mortality for cohort members who are still alive at the moment of observation. We assume that their future probabilities of dying are identical to those of a five-year cohort of comparable age observed just before the end of observation. For example, if our observation period ends on December 31, 2000, then the 1900 birth cohort has complete mortality data only up to exact age 99. So, we assume that mortality from exact age 99 to exact age 100 for the 1900 cohort will be the same as mortality rates at the same age among the 1895-1899 birth cohort. We consider a cohort "almost extinct" if the total person-years remaining to be lived (based on our estimate) for that cohort is no more than one percent of the total lifetime person-years lived for that cohort (assuming the life table begins at age 0).[13] In practical terms, this typically means that we treat cohorts that have reached age 90 or older by the end of observation as almost extinct (and therefore estimate their future mortality in order to complete the life table).

Multi-year life tables

Period life tables for multi-year time intervals (e.g., 10-year periods) are based on death rates calculated by pooling deaths and exposures across the time interval before dividing the former by the latter. Similarly, life tables for multi-year cohorts (e.g., 10-year birth cohorts) are derived by pooling data across cohorts before calculating probabilities of death.[14] In both cases, multi-year life tables are then computed as described above. Thus, death rates and life table quantities from multi-year life tables are not simply the average of the respective values across periods or cohorts.

Abridged life tables

For both periods and cohorts, "abridged" life tables (e.g., by five-year age groups) are extracted directly from "complete" life tables (i.e., by single year of age).[15] Deriving abridged tables from complete ones (rather than computing them directly from data in five-year age intervals) ensures that both sets of tables contain identical values of life expectancy and other quantities.

Changes in population coverage

Some countries and areas have experienced changes in their territorial boundaries over the period covered by the HMD. In other cases, there have been changes in the coverage of demographic data (e.g., vital statistics change from covering the *de facto* population to covering the *de jure* population). These changes must be taken into account when computing death rates and life tables. In general, death counts must always refer to the same territory as the exposure-to-risk when calculating death rates. Adjustments to the methods in the event of changes in

---

[12] See pp. 38 for the formulas used to derive cohort q_x values.
[13] For more details regarding almost extinct cohorts, see pp. 40-42 of the Methods Protocol.
[14] For more details on computing multi-cohort life tables, see pp. 39-40 of the Methods Protocol.
[15] For details regarding abridged life tables, see p. 42-43 of the Methods Protocol.

population coverage are described in Appendix D of the Methods Protocol. In cases where there was a change in population coverage, the population estimates available from the country page are given for the population just before (i.e., December 31st of year t-1, indicated as year "$yyyy$-") and just after (i.e., January 1st of year t, denoted as year "$yyyy$+") the change.

Total versus Civilian Mortality

For populations that suffer substantial war losses, mortality estimates (especially for males) based on the civilian population do not accurately reflect period mortality during wartime nor do they represent the true mortality experience for the cohorts that experience heavy war loses. For such cases, we present two sets of estimates whenever possible: one for the civilian population (which excludes military deaths that occur abroad) and a second for the total population (including military deaths that occurred abroad). The latter are generally derived by adjusting death counts and population estimates based on the data available (e.g., deaths and military size reported by the military authority). Period mortality is identical in the two series except during war years. Cohort life tables for the civilian population are provided, but these data are of questionable quality for those populations that experienced substantial war mortality because they do not represent true cohort mortality.