Explanatory Notes

Complete Data Series

Description

The complete data series includes the following types of data:

- Live birth counts
- Death counts
- Population size on January 1st
- Population exposed to risk of death (period & cohort)
- Death rates (period & cohort)
- Life tables (period & cohort)

Period data are indexed by calendar year, whereas cohort data (if available) are indexed by year of birth.

Cohort Death Rates (and Exposure) are provided if there are at least 30 consecutive calendar years of data for that cohort. For example, the mortality series for Sweden begins in 1751, therefore we can show death rates for the 1675 birth cohort for ages 76 and older. The cohort death rates at younger ages are shown as missing (denoted by "."). Similarly, if the mortality series ends in 2002, we can show death rates for the 1972 cohort up to age 29 because by December 31, 2002 everyone in that cohort has reached age 30. Yet, mortality data for age 30 will remain incomplete until December 31, 2003.

Cohort Life Tables are presented for a population if there is at least one cohort observed from birth until extinction (i.e., the date by which all cohort members are assumed to have died). In that case, life tables are provided for all extinct cohorts and for some almost-extinct cohorts as well (see the Methods Protocol, pp. 42-44).

Format of Data Files

- Data files are tab-delimited text (ASCII) files.
- Files are organized by sex, age, and time.
- Population size is given for one-year[1] and five-year[2] age groups.[3]
- Deaths,[3] exposure-to-risk, death rates, and life tables are given in similar formats of age and time:
    - 1x1 (by age[1] and year)
    - 1x5 (by age[1] and 5-year time interval)
    - 1x10 (by age[1] and 10-year time interval)
    - 5x1 (by 5-year[2] age group and year)
    - 5x5 (by 5-year[2] age group and 5-year time interval)
    - 5x10 (by 5-year[2] age group and 10-year time interval)
- Deaths are also given by Lexis triangles (i.e., by age,[1] birth cohort, and calendar year).[3]

[1]One-year age groups (or "by age") means 0, 1, 2,..., 109, 110+.

[2]Five-year age groups means 0, 1-4, 5-9, 10-14,..., 105-109, 110+. Age groups are defined in terms of completed age, so "5-9" extends from exact age 5 to just before the 10th birthday (sometimes written elsewhere as "5-10").

[3]Some of these numbers are estimates (of population size or numbers of deaths), not actual counts, and therefore may be expressed as non-integers.

Life Tables include the following columns:

Year	Year or range of years (for both period & cohort data)
Age	Age group for n-year interval from exact age x to just before exact age $x+n$, where n=1, 4, 5, or ∞ (open age interval)
m(x)	Central death rate between ages x and $x+n$
q(x)	Probability of death between ages x and $x+n$
a(x)	Average length of survival between ages x and $x+n$ for persons dying in the interval
l(x)	Number of survivors at exact age x, assuming l(0) = 100,000
d(x)	Number of deaths between ages x and $x+n$
L(x)	Number of person-years lived between ages x and $x+n$
T(x)	Number of person-years remaining after exact age x
e(x)	Life expectancy at exact age x (in years)

See the Methods Protocol (pp. 34-44) for more details about life table calculations.

Important Notes

- Deaths, population estimates, death rates, and life tables are provided by single years of age up to 109, with an open age interval for 110+. However, these data are sometimes the product of aggregate raw data (e.g., 5-year age groups, open age intervals), which have been split into single years of age using the methods described in the Methods Protocol. The original raw data that were extracted from published or unpublished sources are available from the HMD Input Database.
- For populations with territorial changes, two sets of population estimates are given for years in which a territorial change occurred. The first set of estimates (identified as year "19xx-") refers to the population just before the territorial change, whereas the second set (identified as year "19xx+") refers to the population just after the change. For example, in France, the data for "1914-" cover the previous territory (i.e., as of December 31, 1913), whereas the data for "1914+" reflect the territorial boundaries as of January 1, 1914.
- For period life tables, the central death rate m(x) is used to compute probabilities of death q(x). The values of m(x) below age 80 are by definition equal to the observed population death rate M(x) shown on each country page. At older ages, however, the number of deaths and the exposure-to-risk eventually become quite small, and thus observed death rates display considerable random variation. Therefore, we smooth the M(x) values for ages 80 and older and use these smoothed values to compute q(x) above a certain age (based on the number of observed deaths). For details, see the Methods Protocol (pp. 35-37). This procedure helps to avoid certain difficulties in period life table calculations at older ages that may be caused by: 1) extremely high death rates resulting from exposure being smaller than the number of deaths, 2) death rates of zero resulting from no deaths at an age where exposure is non-zero, and 3) undefined death rates at all ages where exposure is zero. For cohort life table calculations, such difficulties are not present.

Input Data

Description

The Input Database houses the raw data that are the basis for all HMD calculations. Input data files for each population are accessible from the country page.

Every country/area has at least three input data files: births, deaths, and population size. For countries that have experienced border changes during the period covered by the HMD, the Input Data also include a file of territorial adjustment factors.

In addition, there are four PDF files for each population. The *Background and Documentation* file contains basic information regarding the data for that particular population. The *Notes* file contains specific notes pertaining to individual data points. The *Data sources* and *Reference* files include a complete list of data sources for every number in the raw data files, with the latter file also containing codes that

link these sources to individual data points in these files.

Format of Data Files

- Data files are stored in comma-delimited text (ASCII) files.
- See Structure of the Input Database for detailed information about data types and file formatting.

Important Notes

- The Input Data are manipulated according to procedures specified by the Methods Protocol to create the Complete Data Series .
- An important goal of the HMD is to provide users with open access to all the Input Data. However, we must limit access in a small number of cases because of restrictions imposed by the organizations or individuals who have contributed the data.

© 2022 MPIDR // Contact // Imprint // Privacy Policy // Sitemap