

Formats of HLD Data Files

Contents

Introduction	1
1. Single source data files.....	2
2. Pooled, or multiple source, data files	2
3. Structure of the data files	2
3.1 Life tables.....	2
3.2 Files for life expectancy at birth.....	4
4. Original life tables	4
APPENDIX 1: Names of files.....	6

Introduction

The Human Life-Table Database (HLD) provides two types of data: original life tables “as they were published” and recalculated life tables. The calculation of life tables is described in detail in the HLD Methods Protocol¹. Here we only describe the format of the HLD data files. The HLD is a collection of published life tables or tables which are directly submitted to the database by researchers. Thus, we consider as a single data source a publication or a single submission. All tables in the HLD are organized in a country-specific manner, i.e. all data files are grouped by country. If a data source contains life tables for more than one country, only the table with data for the respective country will appear at country page. For every data source the HLD provides the following country-specific data and information:

- standardized complete life table in text format (if the original publication contains a complete table);
- standardized abridged life table in text format;
- reference to the original statistical publication or to another data source;
- scanned copy of the original life table as published in PDF format.

In addition, we provide for each country two pooled data files one of which contains all life tables available for the respective country and the second one the life expectancy at birth extracted from these life tables. Finally, we also provide the pooled file for the whole HLD. This file contains all life tables available in the HLD. To facilitate rapid downloads, all pooled HLD data files are zipped.

All data files are formatted comma-delimited text files that are easily readable in Excel and statistical packages. Detailed composition of the file names is given in Appendix 1.

¹ Available at <http://www.lifetable.de/docs/methods.pdf>

The first row in the data files contains the header. Missing values are coded as dots ('.'). All life tables are calculated with radix 100,000 ($l_0 = 100,000$). Life expectancy is recalculated with two decimals, even if the original publication provides it with less precision. In addition to standard life table functions recalculated (in most cases) from the original l_x (see Methods Protocol for details), the data file includes the original value of life expectancy as it was published.

1. Single source data files

The single source data file presents the life tables by each data source separately. This is a comma-delimited formatted text file. The description of all data fields is given in section 3.1. The single source data files may contain more than one life table if the original publication contains several life tables for the respective country. If the original life table is available as a complete one, the HLD data file will include both complete and abridged recalculated version of this life tables. The abridged life table is calculated from the recalculated complete table. If the publication does not provide a complete life table and only an abridged table is available, the HLD data file includes only the recalculated (from the original frequency function) abridged life table. In both cases, we keep the original age scale. In particular, the open age interval remains unchanged.

2. Pooled, or multiple source, data files

There are two types of pooled data files provided for each country, which contain the following data:

1. recalculated complete and abridged life tables for all data sources;
2. life expectancy at birth extracted from all available tables.

The structure of the pooled data files for the life tables is same as the structure of the single source data files. The description of data fields for the life expectancy at birth data file is given in section 3.2.

The HLD provides the possibility to download all life tables for all countries in one file. This HLD pooled file has the same structure as the country-specific pooled files.

3. Structure of the data files

3.1 Life tables

Heading of the pooled and single source data files:

Country, Region, Residence, Ethnicity, SocDem, Version, Ref-ID, Year1, Year2, TypeLT, Sex, Age, AgeInt, $m(x)$, $q(x)$, $l(x)$, $d(x)$, $L(x)$, $T(x)$, $e(x)$, $e(x)Orig$

The description of the fields:

1. *Country* (three-character alphabetic): ISO 3166-1 alpha-3 (or similar) codes for countries or areas. A description of all codes is given on the HLD web site in section “Codes used in data files”
 2. *Region* (alphanumeric): region codes for identifying the principal subdivisions (e.g., provinces or states) of countries. The field is equal to zero (0) for the whole country data. A description of all codes is given on the HLD web site in section “Codes used in data files”
 3. *Residence* (alphanumeric): subpopulation (e.g. urban or rural) code. The field is equal to zero (0) for the total population.
 4. *Ethnicity* (alphanumeric): code for ethnicity, religion or race, zero (0) for the total population.
 5. *SocDem* (alphanumeric). Coded for socio-demographic characteristics (e.g. low educated). The field is equal to zero (0) for the total population.
 6. *Version* (numeric). Version of the life table. This field is used if several version of a life table for one and the same year are available from one data source (alternative estimates, life tables revised after the census, etc.).
 7. *Ref-ID* (alphanumeric): data source code. A description of all codes is given on the HLD web site in “Data sources” section. This code has the format NNNN.PP, where N and P denote any character. The first part of the source code (NNNN) provides information about principal data source (e.g. book or demographic yearbook), while the second part (PP) refers to exact place where this table(s) appears (page or section in the book, issue of the yearbook, etc.).
 8. *Year1* (four-digit numeric): first year of the period the data pertain to.
 9. *Year2* (four-digit numeric): the last year of the period the data pertain to. If data belong to a single calendar year, the fields “Year1” and “Year2” are identical.
 10. *TypeLT* (numeric): type of life table. The HLD provides three types of life table: recalculated complete life table (*TypeLT*=1), abridged tables calculated from the recalculated complete table (*TypeLT*=2), and the abridged table recalculated from a published abridged table (*TypeLT*=4).
 11. *Sex* (numeric): sex, 1 for males and 2 for females.
 12. *Age* (numeric): age, the value of *Age* is the lower age limit.
 13. *AgeInt* (numeric; “99” for the upper open age intervals): the length of the age interval.
- The next 7 fields contain standard life table functions:
14. $m(x)$: central death rate for the age interval [*Age*, *Age*+*AgeInterval*).
 15. $q(x)$: probability of death within age interval [*Age*, *Age*+*AgeInterval*).

- 16. $l(x)$: probability of survival from birth to exact age Age
- 17. $d(x)$: number of deaths within the age interval [$Age, Age+AgeInterval$).
- 18. $L(x)$: number of person-years lived within the age interval [$Age, Age+AgeInterval$]
- 19. $T(x)$: number of person-years lived after the exact age Age
- 20. $e(x)$: life expectancy at exact age Age . This life expectancy is recalculated from one of the frequency functions (l_x, d_x or q_x) and may differ from the original (published) life expectancy.
- 21. $e(x)_{Orig}$: life expectancy extracted from the original (published) life table. In life tables included in the HLD before May 2017 the original life expectancy was calculated from the original published T_x (which is not presented in the recalculated life table).

3.2 Files for life expectancy at birth

Heading of the pooled data file for life expectancy at birth:

Country, Region, Residence, Ethnicity, SocDem, Version, Ref-ID, Year1, Year2, TypeLT, Sex, $e(x)$, $e(x)_{Orig}$

All data fields have the same format as the respective data fields in the life table files.

4. Original life tables

Images of published life tables are presented in portable document format (.pdf) as shown in Figure 1.

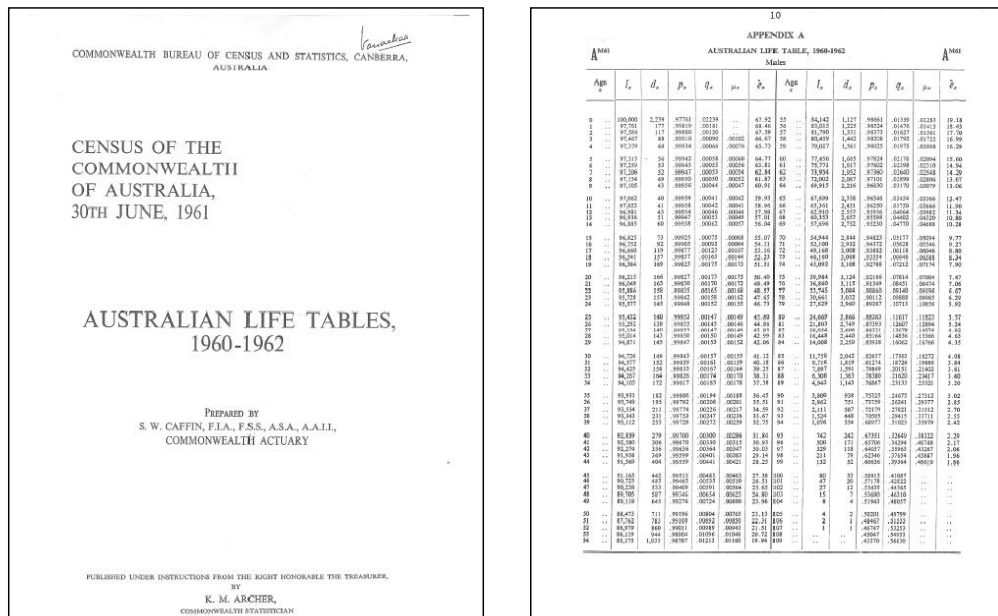


Figure 1. Presentation of the published Australian life table for the period 1960-62 in PDF format.

The first page of such PDF contains bibliographic information on the data source. The following pages present the table itself.

APPENDIX 1: Names of files

For single source data (.txt) files and respective PDFs we use the following rules to construct the file name. The first three characters of the file name stand for a standard (uppercase) ISO code, representing a country or area name (AUT for Austria, CAN for Canada, etc.). For countries, the ISO 3166-1 alpha-3 codes are used (see, for example, Wikipedia: http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3). For sub-populations or areas not corresponding to an ISO country, the ISO 3166-2 region codes are used (where available) as an extension of the country code (see http://en.wikipedia.org/wiki/ISO_3166-2). The next four characters are used to indicate a subpopulation (ethnicity, region, residence place, or socio-demographic status). For the total population we use “0000” code. Followed by 8 characters which represent the period covered by the data (e.g. 20002005 for the period from 2000 to 2005). Finally we add 2 characters indicating if we are either dealing with a complete or an abridged original table (CU or AU) as well as one character to identify the version of the table (1 or 2).