Title

Data Resource Profile: Human Life table Database (HLD)

Author List

Ainhoa Alustiza[1*], Magali Barbieri [2,3], France Meslé[3], and Dmitri Jdanov[1]

Data resource basics

Life tables are the oldest and most important tool in demography. Life tables are also probably the most renowned and recognised tool within non-demographers, either they belong to the scientific community or the general public.

They has become progressively essential for more professionals from different fields: sociologists, epidemiologists, public health researchers, biologists, mathematics and actuaries, geographers, geologists, medical doctors, palaeontologists, historians, and many others. In social science the study of mortality trends, human longevity and life span limits have been very prolific fields in the past decades, as a result of the dramatic decrease of mortality in the last century. Understanding these changes, the driving factors and the future consequences for the population, both from a collective and an individual perspective, are still central questions that raise the interest of not just researchers, but also policy makers and the general public.

To document longevity revolution and provide sold basis for research at national level, the international team of researchers from the Department of Demography at the University of California Berkeley and the Max Planck Institute for Demographic Research created the Human Mortality Database (HMD). The HMD provides high quality comparable -across time and space- continuous mortality data series. During the early stage of work, the HMD team collected a lot of published life tables. Many of these tables were provided by Väinö Kannisto. This collection by its design could not be incorporated in the HMD, and therefore, a sister database – the Human Life-Table Database, HLD, was established. In contrast to the HMD, the HLD includes a variety of life tables of different quality. It contains unique historical life tables and most recent data, officially published data and alternative estimates. The HLD team makes basic quality check, relying on the data providers and the original publications authors the main responsibility.

Data resource area and population coverage

The HLD is a compilation of period life tables that aims to cover all the population areas and time periods for which exist official publications from national statistical offices or significant researchers. Therefore, the HLD is not a static collection, but a database in continuous growth. By the end of 2019 the HLD contained around 9000 period life tables, both complete and abridged, for 133 countries or population areas worldwide, covering a maximum 268 year time range that varies between 1751 and 2019.

The HLD is a project developed in collaboration between three institutions: the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany, the Department of Demography at the University of California at Berkeley (UCB), USA, and the Institut d'Études Démographiques (INED) in Paris, France. The work is being carried out at the MPIDR.

The HLD includes national population life tables and various subpopulation ones whenever are available: life tables by regions for 45 countries, by settlements –urban and rural residents- for 14 countries, by national-specific ethnicities for 9 countries and by socio-economic status –farmers, low or very low income- for 2 countries.

The selection criteria to include life tables in HLD relies on the availability of published life tables that are not based on purely model assumptions. Therefore, HLD life tables include [1] official life tables published by the correspondent national statistical office, [2] non-official

tables from scientific –peer reviewed or not- publications, [3] data collections produced by research centres, [4] personal contributions from significant researchers in demography or attached to official institutions, [5] other organizations and institutions official publications that do not apply high modelling assumptions on the development of their tables. As a result, HLD includes a large and diverse set of life tables covering most of the world countries, 133 out of the 195 countries in the world (around 70%). This heterogeneity requires special caution from the users, as the wide selection of life tables included in HLD cannot guarantee comparability across time and populations.

Measures

HLD publishes identical and consistent indicators in all the life tables. Currently, only period life tables are available in the website. HLD gathers *complete* or single age interval life tables whenever they are available. For the rest of the cases, *abridged* or five year age interval tables are offered. For the *complete* life tables, the HLD includes in the text files an additional recalculated *abridged* table.

The HLD tables contain all the functions –columns- of the standard life tables, in addition to the basic individual life-identifying codes or indicators. The identification variables are:

- the country code –based on the ISO 3166-1 alpha-3-,
- the four subpopulation codes, when applicable:
    i. region code
    ii. ethnicity code
    iii. settlement code
    iv. socio-demographic status code
- a code to identify the version of the table. It may happen that different sources publish a life table for the same geographical area for the same period of time. In these cases, we include both tables with a different version code that distinguishes them. These cases may be particularly interesting for researchers, since they provide a great opportunity to compare disparities due to different decisions: revision of the data, methodological differences, etc.
- the code that identifies the source
- the beginning year of the life table
- the ending year of the life table
- the code that distinguishes whether the table is complete or abridge

Subsequent to these table identification codes HLD life tables include the standard life table columns or functions themselves, recalculated from original q_x or l_x values, depending on their availability (see for details the HLD Methods protocol):

- sex
- age
- age interval
- age-specific death rates between ages x and $x+n$ ($_nm_x$)
- probabilities of death between ages x and $x+n$ ($_nq_x$)
- number of survivors at the exact age x (l_x)
- number of deaths between exact ages x and $x+n$ ($_nd_x$)
- number of person-years lived within ages x and $x+n$ ($_nL_x$)
- number of person-years lived after the exact age x (T_x)
- –re-estimated- life expectancy at age x (e_x)
- In addition, HLD publishes the life expectancy at age x disclosed in the original source or publication ($e_{(x)Orig}$), which may vary slightly from the re-estimated one.

Besides these measures that HLD publishes in text format files for individual life tables and in zipped files for aggregated data, scanned copies of the original tables and publications and the appropriate citation of the source are also available for each table in *pdf* format.

Methods

The HLD was created and launched online in June 2002 within the HMD project. Originally, it included life tables for less than 50 countries and it has been constantly increasing since then, achieving the current amount of 9000 tables and 133 countries. The procedure behind the HLD include the following steps:

i. **Search of life tables**: The HLD team search historical and contemporary life tables through all the available channels: different types of paper publications –journals, reports, books, etc.-, online publications, national statistical offices websites, personal communication with researchers and research centres, etc. There is no strong inclusion criteria in the HLD, but obviously wrong data and life tables based exclusively on model data are not included.

ii. **Re-estimation of the life tables**: Once the life tables are gathered by the HLD team (input data), basic quality checks and a standardize process are carried out, so all the resulting life tables (output data) are based on a standard methodology. The goal of this process is to address the methodological differences in life table constructions associated with the conversion of the death probabilities or frequency functions (l_x, d_x or q_x) into the duration functions (T_x, L_x or e_x). The HLD utilises only one of of the original life table frequency function, usually the l_x, and recalculates the rest of the duration functions, in the same unvarying way for all the life tables included in the database (all countries, all years). The resulting tables enclose all the basic functions of the standard life tables -m_x, d_x, q_x l_x, T_x, L_x or e_x-, that some of the original publications may not have. Since the resulting values of the own estimated e_x may present small disparities with the original publication ones, the HLD life tables include an additional column containing the life expectancy values of the original publication - $e_x(orig)$. Nevertheless, life tables in the HLD cannot be considered entirely comparable, since the original initial data might be absolutely distinct, and consequently incomparable. Thus, users should take into account the possibility of such problems.

iii. **Preparation of the publication files**: Different downloadable files are available in order to facilitate the access of the users to all the HLD tables. On the one hand, for each individual life table there is a text file including the life table functions and a *pdf* file containing the appropriate reference of the source and a scanned copy of the original life table. On the other hand, pooled files have been created in order to simplify and facilitate data access. For each particular country, two aggregated files in *csv* format are available: the first one includes all the available life tables for that geographical area and the second one all the available values of the life expectancy at birth for that specific country. Besides these pooled files, there is accessible an aggregated "*master*" file –"*All HLD data*"- that includes all the available life tables in HLD, that is, all countries and all years.

The HLD keeps continuously growing as new data become accessible: latest and newest tables of existing countries in HLD, introduction of tables that were not incorporated yet to our database –new findings of past years or subpopulations, input of the less prioritized tables due to the limitations of time and human resources working on the HLD, or even the availability of data for new countries.

Data resource use

The number of users and more importantly, scientific work based on HLD has been continuously increasing since it was launched for the first time in 2002. Currently, around 150-200 unique users access monthly the HLD. At present a non-exhaustive list of publications based on data provided by HLD is being developed and it is revealing the diversity of disciplines, besides demography itself, that rely on HLD data to develop their research or achieve their goals. This list include publications that since 2003 have been

published mainly as peer review articles, but also as reports, book chapters and other similar scientific works.

HLD users belong to diverse scientific disciplines: demographers, sociologists, epidemiologists, geographers, economists, historians, actuaries, mathematics, medical doctors, public health researchers and even biochemists and biophysics. Non-scientific areas, like insurance sector, are known users of the HLD. Institutions like the United Nations, that develop and publish their own life tables are important users of this database. Such a large and diverse spectrum of professionals gives already an idea of the numerous potential research areas that the HLD can be applied in.

A more detailed look into the publications shows the enormous quantity of research topics developed with the HLD as a data source: classical demographic topics as the improvements in mortality and life expectancy, differences in mortality around the world or within specific populations, trends in lifetime inequalities, convergence and divergence patterns in life expectancies, etc. More formal, methodological and mathematical demographic works have been also developed with HLD: multi stage survival analysis, lifetime and mortality modelling, life expectancy reconstructions, life equations related to senescence processes, etc. Epidemiology and public health are important areas where the HLD has been adopted to study trends in diseases as dementia, cancer or more general and classic questions involving human ageing and genetics.

Strengths and weaknesses

The main advantages that the HLD provides to its potential users are:

- **Accessibility**. The HLD offers through a unique website location most of the life tables published by national statistical institutes around the world and the ones published or provided by the most significant scientists and research centres. The user-friendly design and the freely downloadable aggregated files, without any registration requirement, constitute an added value for users.
- **Permanent access to the original files**. All the original sources are well documented for each life table, providing not just the appropriate citation, but a scanned copy of the original table. With the passage of time some of the life tables, after being published, have been eliminated from the providers websites or the original Internet links are not active anymore. These tables, nevertheless, can be accessed through the HLD.
- **Comparability between countries and periods**. HLD standardization process makes comparable life tables from different geographical areas and time periods.
- **File formats**. Due to the recalculation process, all the life tables included in the HLD can be access either individually, by country or through a single master file in text or csv files. Although most of the recent tables are available through the website in computer readable formats, many of the ones included in HLD had been manually input from books or pdf files. AS a consequence, HLD tables are presented in identical machine readable files and therefore users can convert them easily into the most appropriate format.
- **Entire life tables**. All standard columns of life tables are provided, including tables - m_x, d_x, q_x l_x, T_x, L_x, e_x and $e_x(orig)$, so researchers can choose the most appropriate indicators for their studies.
- **Coverage.** HLD covers a very large number of countries and geographical areas worldwide -133-. Time coverage is also significant, including data that varies from 1751 to 2019. In addition to this immense coverage, HLD include subpopulation life tables. 2705 of the life tables refer to national populations. This means that the vast majority of the life tables included in the HLD refer to some subpopulation: currently, tables corresponding to different regions, type of residence (urban-rural, cities), ethnicities and socio-economic status are included. This particularity expands significantly the research possibilities that the HLD offers.

- **Periodically updates**. New life table publications are regularly incorporated to HLD. Since the HLD aims to cover the greatest number of populations (countries, years, subpopulations), HLD maintenance and update does not limit to the entry of the most recent tables, but to the input of any available table that meets the inclusion criteria. Therefore, HLD welcomes and encourages any unpublished contribution that could be added to the database.

HLD fundamental weaknesses or disadvantages are related to:

- **Data quality –disparities-**. HLD does not develop life tables from death and population counts, so does not control or distinguish different data quality life tables. In this sense, HLD practises as a repository of life tables, being the original source or provider the one that defines and takes responsibility of the data quality of the tables. The large diversity of countries and periods covered and the large amount of providers turn this data quality inconsistency inherent to the database. This disadvantage can be used though by some researchers as a possibility to improve related methodological issues.
- **Not continuous series.** Although there are continuous life table series for some countries, even some subnational populations, that is not the case for many –small, less developed- countries in HLD.
- **No single calendar year**. Due to the nature of some of the populations, life tables cannot be calculated by single calendar year and single year of age. Therefore, many tables refer to two, three, four or even five calendar years and some of them are abridged life tables.

Data resource access

Data is freely available through the URL [https://www.lifetable.de/]. No registration is required in order to access the data. These are the downloadable files:

- Aggregated file including all the HLD life tables (csv format)
- For each specific country, aggregated file containing all the available life tables (csv format)
- For each specific country, aggregated file containing the available time series of the life expectancy at birth (csv format)
- Individual life tables (text format), and the correspondent original scanned life tables (pdf format), including the source reference.

Related databases

The HMD developed in collaboration between the MPIDR and the UCB is the most significant source of mortality data. The HLD was actually conceived and developed within the HMD project, becoming progressively a data resource itself. Currently, the HMD offers high quality mortality data for 41 -developed- countries, including life tables. The longest time series of life tables published in the HMD is for Sweden, covering 266 years from 1751 to 2017.

Author affiliations

[1] Max Planck Institute for Demographic Research, Rostock (Germany)

[2] Department of Demography at the University of California Berkeley (USA)

[3] Institut National D'Études Dëmographiques, INED, Paris (France)