



Data Resource Profile

Data Resource Profile: The Human Mortality Database (HMD)

**Magali Barbieri,^{1,2*} John R Wilmoth,^{1,3} Vladimir M Shkolnikov,^{4,5}
Dana Gleit,⁶ Domantas Jasilionis,⁴ Dmitri Jdanov,^{4,5} Carl Boe,¹
Timothy Riffe,¹ Pavel Grigoriev⁴ and Celeste Winant¹**

¹Department of Demography, University of California, Berkeley, CA, USA, ²French Institute for Demographic Studies, Paris, France, ³Population Division, United Nations, Department of Economic and Social Affairs, New York, NY, USA, ⁴Max Planck Institute for Demographic Research, Rostock, Germany, ⁵New Economic School, Moscow, Russia and ⁶Georgetown University, Washington, DC, USA

*Corresponding author. Department of Demography, University of California, 2232 Piedmont Avenue, Berkeley CA 94720-2120, USA. E-mail: magali@demog.berkeley.edu

Accepted 19 May 2015

Data resource basics

Over the past 100 years, human longevity has increased more than during all preceding human history combined. This remarkable change was caused by reductions in mortality; first generally at younger ages, then later progressing to older ages. The precise causes as well as the social and economic impact of this change are still unclear. What are the major patterns, driving forces and future prospects of this process? To what extent are these changes determined by period factors such as advances in medicine and economic development or by cohort effects such as early life experiences and accumulation of behavioural risks over the life course? How much further can human longevity extend?

These are central questions for scholars and policy makers. The Human Mortality Database (HMD), as a collection of detailed, consistent and high quality human mortality data, is an important resource for addressing these fundamental questions. Although there are other international databases of mortality indicators, some of which cover many more countries (eg those compiled by the United Nations Population Division, the World Health Organization, Eurostat and the Institute for Health

Metrics and Evaluation), they all have limitations mostly related to quality and detail of the data. For example, most lack information on advanced ages, are restricted to short time periods or focus on periods but not cohorts; these disadvantages are addressed by the HMD.

Data resource area and population coverage

The HMD strives to include all populations for which death registration and census data are virtually complete. As a result, countries and areas in the HMD are relatively wealthy and for the most part highly industrialized. At present the database contains detailed population and mortality data for 37 developed countries and 46 populations (including sub-national groups), with series starting as early as 1751 in Sweden and covering more than a 100 years for 16 populations. The selection criteria for inclusion of a country in the HMD rely mainly on the quality of its data. In order to be included in the HMD, a country's death registration system must be nearly complete (close to 99%). For such countries, original statistics are then collected extending as far back in time as there are age-classified census data and annual death counts in a

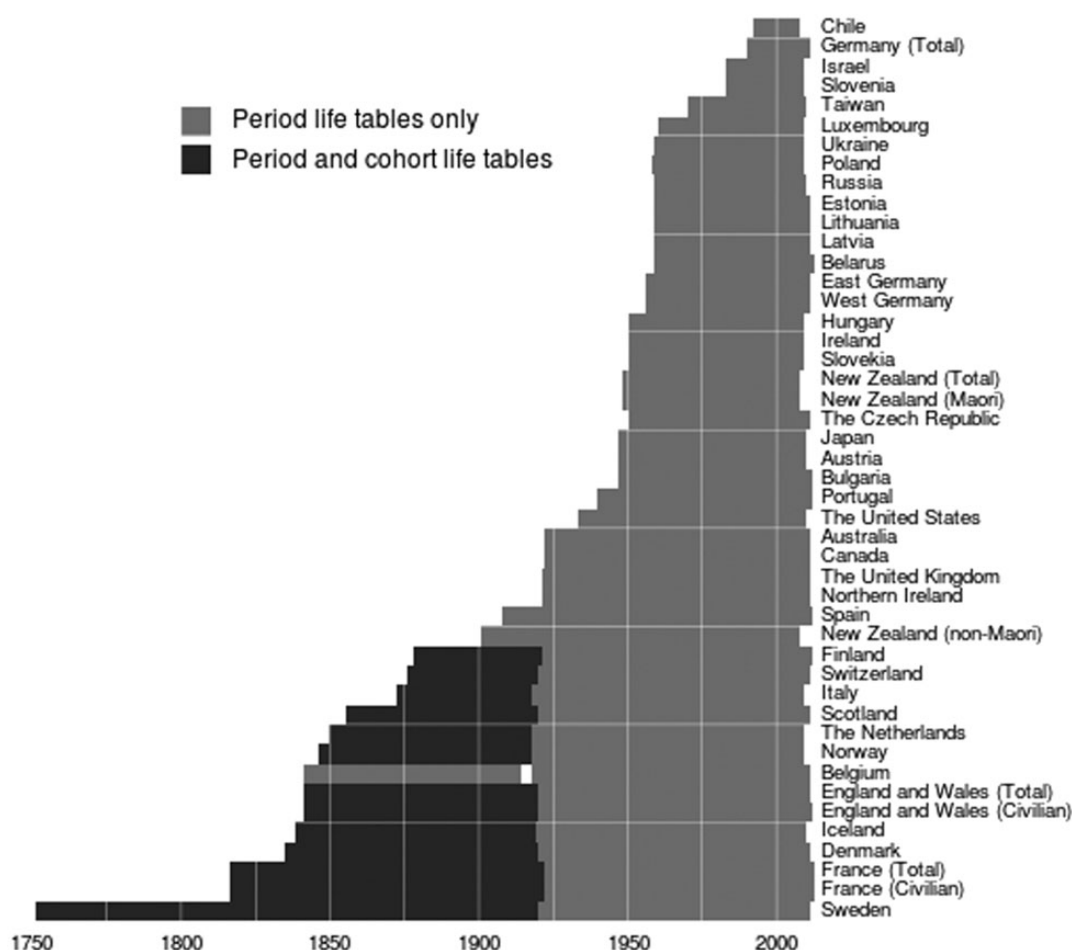


Figure 1. HMD period and cohort mortality series by population and year (as of January 2015). Cohort life tables are indexed by the year of birth. Cohort life tables are unavailable in Belgium because of missing input data for the period 1914–18. Based on an earlier design by Adrien Remund (2015)¹.

sufficiently detailed format (at a minimum, there should be 5-year age groups with a separate category for infants and an open interval at age 80 or higher). Note that for historical data (especially for years preceding 1870–80), the quality is often lower, especially at ages 80 and above. In such situations, the HMD has well-defined and transparent methods that are applied in a consistent manner, accompanied by prominent warnings that alert users to potential problems and issues. Figure 1 shows the years for which data series are available for each population in the collection.

For some data series, additional details are provided on specified sub-populations: (i) for populations that have suffered substantial war losses and where available data allow computation of population and mortality indicators for the total (including the military) and civilian populations separately, such as is the case for England and Wales and for France; and (ii) whenever original data series are available for important population sub-groups as collected by the official statistical agency for a country. These include separate series for West and East Germany; for England and Wales,

Scotland and Northern Ireland in the United Kingdom; and for the Maori and non-Maori of New Zealand.

Figure 2 shows life expectancy trends in all HMD countries and averages for selected groups of countries since 1970 for men and women. The figure highlights the progress in most of the populations along with a broadening inter-country variation, in particular the growing disparity between the countries of the former Soviet Union and other advanced countries.

Measures

- The HMD contains uniform death rates and complete (by single year of age up to 110+) and abridged (ages 0, 1–4, 5–9, 10–14... 110+) period life tables for extended periods of time, with regular updates at 2- to 3-year intervals in general, though more frequently for some of the largest countries. All standard life table quantities are provided, including age-specific death rates, probabilities of dying, numbers of survivors,

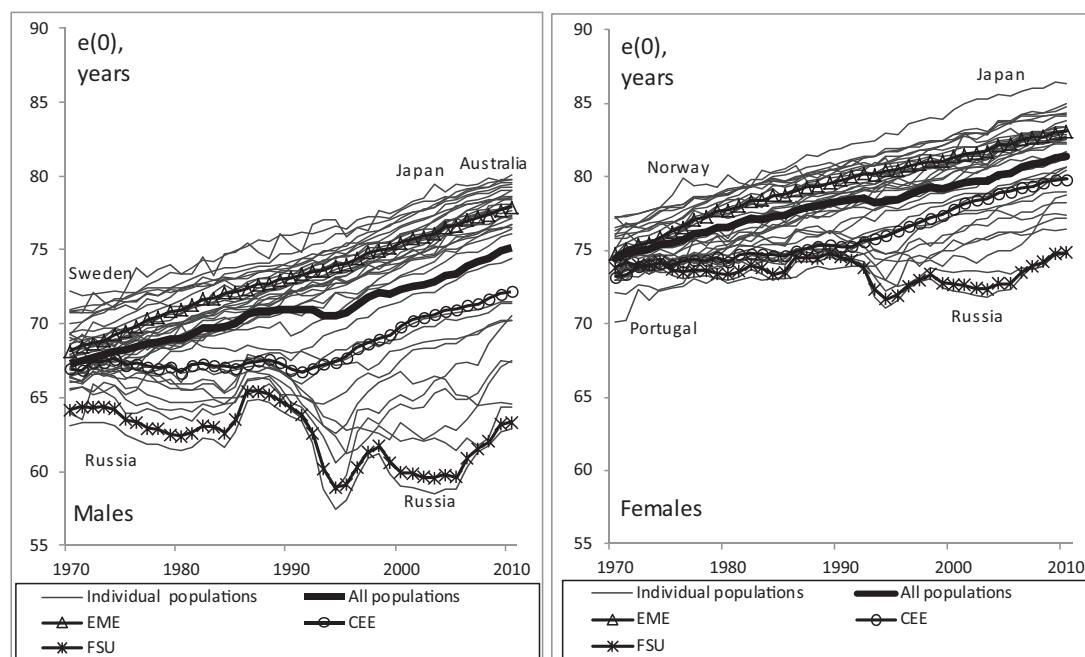


Figure 2. Trends in life expectancy at birth by sex from 1970 to 2010 for HMD populations and population groups. Individual populations: 36 selected HMD national or sub-national populations. Established market economies (EME): Australia, Austria, Belgium, Canada, Denmark, England and Wales, Finland, France, Iceland, Ireland, Italy, Japan, Luxembourg, The Netherlands, New Zealand, Northern Ireland, Norway, Portugal, Scotland, Spain, Sweden, Switzerland, USA, West Germany. Central and Eastern Europe (CEE): Bulgaria, Czech Republic, East Germany, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia. Former Soviet Union (FSU): Belarus, Russia, Ukraine.

person-years lived, person-years remaining and remaining life expectancy.

- In addition to complete series of period life tables, cohort life tables are provided where the observation period is sufficiently long to include at least one cohort observed from birth until extinction (around age 110). In that case, life tables are presented for all extinct cohorts as well as for almost-extinct cohorts (i.e. those that have reached 90 years of age or older).
- The HMD also includes the original raw data collected from reliable sources (mostly as published or distributed by national statistical offices). These raw or 'input' data include the birth, death and census counts as well as official population estimates, from which all mortality estimates and life tables are derived. The data are accompanied by detailed documentation for the database as a whole (with both a complete methods protocol and a summary protocol) and for each particular country (see Figure 3 for a snapshot of a typical HMD country page).

Methods

The HMD was launched online in 2002. It is a resource developed and maintained jointly by the Department of Demography at the University of California, Berkeley

(USA) and the Laboratory of Demographic Data at the Max Planck Institute for Demographic Research (MPIDR) in Rostock (Germany). The French Institute for Demographic Studies (INED) has been involved more recently. Information about the background, history and methodology of the HMD can be found at [www.mortality.org].

Within the HMD team, country specialists are responsible for a specific country or set of countries. Country specialists are in charge of establishing contact with in-country experts (usually at the national statistical office or in the academic community), carrying out all systematic steps involved in assessing the reliability of input data, computing mortality rates and life tables and updating existing data series. The process of constructing the mortality series for each country is described in detail in the HMD methods protocol and in a brief summary available on the website [www.mortality.org].² Major steps include the following.

- Processing birth counts: annual counts of live births by sex are collected for each population over the longest time period available. These counts are used mainly for estimating the size of individual cohorts from their birth.
- Processing death counts: death counts are collected by sex, completed age, year of death and year of birth (whenever available). Deaths of unknown age are

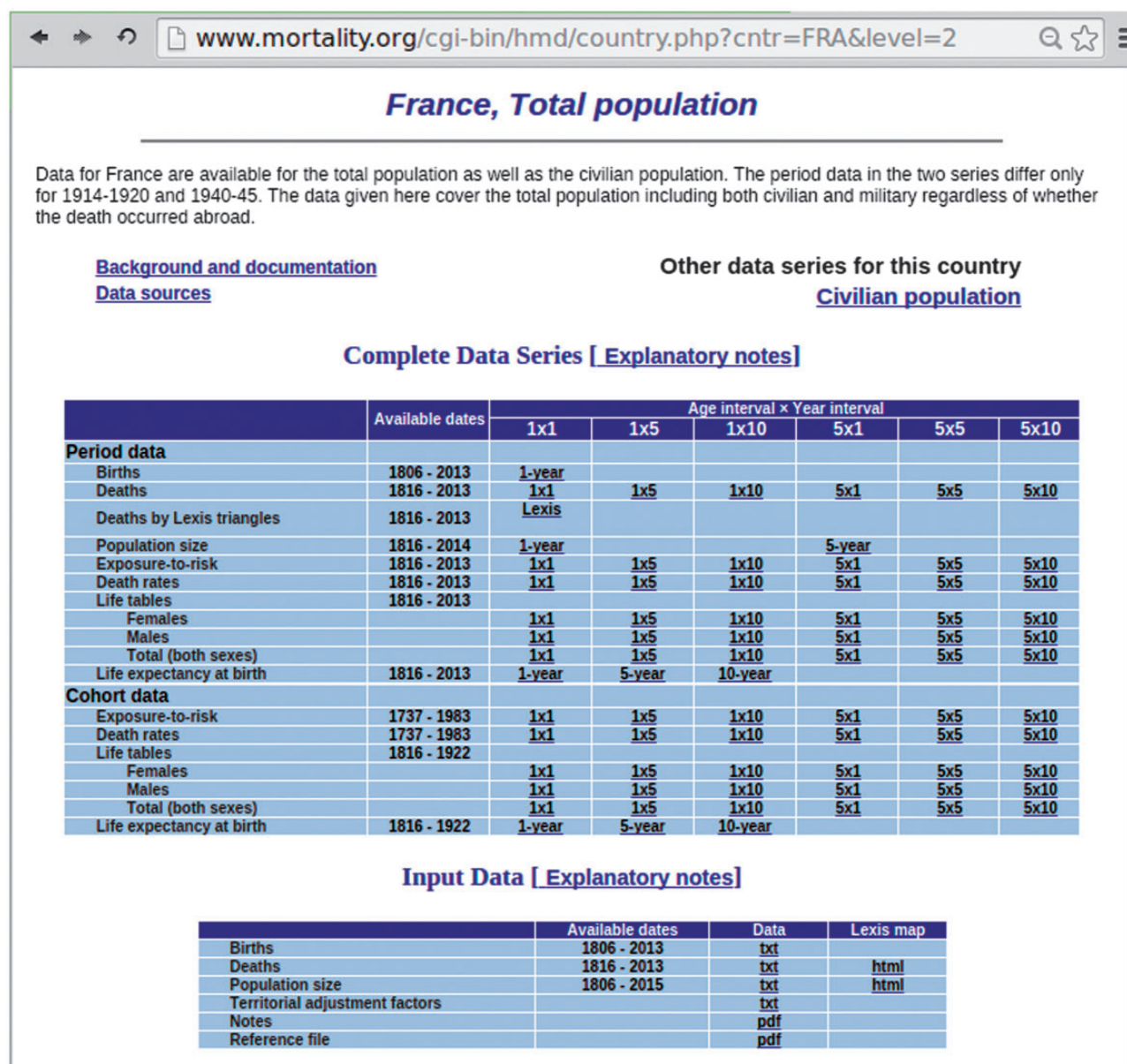


Figure 3. Screen shot of a typical HMD country page.

- distributed proportionately across the age range and aggregated deaths are split into finer age classes.
- iii. Estimating population size: census data and, where available, annual population estimates by sex and age are collected from statistical offices. If reliable annual population estimates are not available, estimates of population size are derived using an intercensal survival method. To adjust for lower data quality at ages above 80, population estimates are derived by the extinct generation method and the survivor ratio method.
- iv. Death rate calculations: For both periods and cohorts, death rates are computed as the ratio of deaths to exposure-to-risk in matched intervals of age. Some adjustments for changes in territorial boundaries are

- introduced where necessary to ensure that death counts refer to the population exposed to risk.
- v. Construction of life tables by single calendar year and single year of age: Period death rates are converted to probabilities of dying by a standard method. Cohort probabilities of dying are computed directly from estimates of deaths and exposure and are related to cohort death rates in a consistent way.
- vi. Construction of multi-year and abridged life tables: life tables are systematically provided by single calendar year and single year of age, no matter the format of the original data. They are also provided by 5-year and 10-year periods and for 5-year age groups by pooling detailed data over years, ages and cohorts.

In addition to these general methods, customized procedures are developed to address country-specific issues, such as gaps in the input data for certain years (for example, 1914–18 in Belgium). These special methods are described in a separate document available as an appendix to the methods protocol (available at www.mortality.org).

Before we include data for an additional country or extend the data series for an existing country, extensive checks are performed to assess the reliability of input data. Country experts are invited to share their understanding of the local statistical and demographic situation and they are consulted whenever specific issues arise. In addition to country-specific knowledge and verifications, a set of standard data quality indicators is computed (including: indices of attraction to particular ages in the reporting of deaths and population; plausibility of mortality patterns; and sex ratios at birth). Diagnostic charts are produced automatically to guide country specialists in their analyses. An internal report is written by the country specialist and discussed with the HMD team before a decision is reached about whether or not to include an additional country or to extend an existing series to an earlier period. The experience accumulated over the past 12 years suggests that there are few remaining countries that can be added to the database in its current form, due to the strict requirements in terms of data availability and quality. However, the existing HMD data series continue to be updated regularly as new data become available for each country.

Data resource use

Over 30 000 unique users have accessed the HMD at various times since 2002. As of December 2014, a non-exhaustive list of known publications that cite the HMD includes over 1000 journal articles (many published in this journal), 140 books or book chapters, 58 dissertations and theses, 51 official reports and 293 technical reports or working papers: see [<http://www.mortality.org/Public/HMD-Publist.pdf>] for the detailed list. The HMD has been and continues to be a major reference.

Epidemiologists and public health specialists have used the HMD to evaluate the representativeness of population samples with respect to mortality and for assessment of healthy life expectancy, years of life (or working life) lost and other composite indicators. For example, Popham *et al.*³ examined cross-national variations in 37 high-income countries in relation to their welfare systems and found that Nordic countries exhibit smaller inequalities in male mortality but not in female mortality. Berry *et al.*⁴ assessed the cost-effectiveness of certain medical interventions based on a combination of age-specific death rates for all causes from the HMD, supplemented with

information about breast cancer mortality, prevention and treatment from other sources; they concluded that screening and therapy jointly contributed to breast cancer mortality trends in the USA from 1975 to 2000. In addition, several epidemiological studies have used the HMD as a baseline against which to compare survivorship of longitudinal study respondents or clinical trial participants.^{5–8}

The reliability of HMD indicators at advanced ages has allowed gerontologists and population biologists to monitor survival and mortality trends up to the very end of the life cycle. Together with other life and social science researchers, they have relied on the HMD collection to study ageing, population dynamics and mortality change. For example, Oeppen and Vaupel⁹ demonstrated the near-linear increase in record life expectancy at birth (+ 0.24 years per calendar year for women, or 6 h a day) from 1840 to the present. Across cohorts born from 1870 to 1920, progress was even steeper: + 0.43 life-years per year of birth, or 10 h a day.^{10,11} The rise of the maximum age at death has been slower: for Sweden from 1969 until 1999, following an acceleration from the slow increase of the previous century, the maximum age at death rose at around 0.11 years per calendar year for both men and women, as shown by Wilmoth *et al.*¹² using HMD data.

Investigating whether increased life expectancy translated into a parallel increase in years spent in disability, Christensen *et al.*¹³ concluded that, overall, healthy life expectancies increased faster than overall life expectancies in high-income countries, thus resulting in morbidity compression. This result was confirmed by a later study, also using HMD data, that contrasted survival rates and health status (measured by survey data) in two Danish cohorts, which determined that cognitive function and the ability to perform various activities of daily living were significantly better in the 1915 than in the 1905 cohort.¹⁴

Because of the standardized methods implemented to produce life tables, HMD data have also been extensively used by demographers and statisticians to model mortality and survival and to forecast population and longevity. Such studies have identified the geographical areas and age patterns responsible for diverging mortality trends in Europe, particularly Eastern Europe. Comparing Russia, Lithuania and Belarus, Grigoriev *et al.*¹⁵ found that although all three countries experienced a decline in mortality over the 1990s, the speed of recovery was strongly and inversely related to the severity of the transition from a centralized to a market economy. Diverging trends within the broad group of industrial nations have also been investigated using HMD data. Such analyses have improved our understanding of the degradation in the US international ranking on life expectancy since the 1980s. In particular, Wilmoth *et al.*¹⁶ showed that this degradation is only

partly attributable to increasing geographical inequalities within the USA over the past three decades.

HMD users also include: university teachers and students, who experiment with newly learned demographic and statistical techniques; policy makers, who simulate the economic impact of changes in retirement systems or of immigration reforms given various demographic scenarios; international organizations such as the United Nations or the World Health Organization, which use the HMD as a complement to their own databases; as well as private-sector organizations (insurance companies and actuaries in particular) who need reliable and harmonized life table estimates.

Strengths and weaknesses

The main strengths of the HMD result from its four guiding principles:

- comparability of the reconstructed series through the use of standardized methods to format and process input data;
- flexibility in the format provided to users, with six alternative output data series for each population;
- accessibility through the internet to all registered users; and
- reproducibility via extensive documentation and provision of original input data, which in many cases are unpublished and difficult to access by other means (especially for historical information).

In addition to regular updates, HMD data series are extended backward in time whenever additional data are identified.

Users are encouraged to contact the HMD staff when they feel the need, whether because of difficulties in accessing the data or because of particular questions or concerns about specific country series or data points. The Frequently Asked Questions section contains information on matters of general interest.

One of the main weaknesses of the HMD is a consequence of its strict data requirements in terms of both format and quality. These requirements have limited the number of national populations that could be included in the database.

Another weakness of the database is the lack of more detailed information. In particular, the database does not include any information about the deceased's characteristics except for age and sex. Thus, HMD mortality estimates are not available by socioeconomic status, because the necessary data are not available in a consistent manner over time and across countries. Nor is cause-of-death information available at this time. For various reasons, data series in the HMD are not generally available at a finer

geographical level than national. However, as indicated below, we and our collaborators are in the process of adding sub-national data series for selected countries.

Future developments

- A number of countries show improvements in the quality of their vital statistics and could eventually be added to the HMD.
- HMD mortality series are provided at the national level only, with a few exceptions as mentioned above. The Canadian Human Mortality Database [<http://www.bdlc.umontreal.ca/chmd/>] and the Japanese Mortality Database [<http://www.ipss.go.jp/p-toukei/JMD/index-en.asp>], two collections that are independent from the HMD but follow the same methods protocol, provide examples of sub-national HMDs. The HMD teams at the University of California, Berkeley and at the Max Planck Institute are developing a similar database for the USA and Germany with separate mortality data series for the American states and German federal states.
- Efforts are also currently under way to expand the HMD series by adding cause-of-death data for a selected group of countries. This work has been challenged by unexpected costs and restrictions on data access, as offices of national statistics have become increasingly attentive to confidentiality issues.
- In order to encourage their use by other researchers, many of the most useful computer scripts are currently being translated into open-access software (R) and documented extensively; when that process is complete, they will be published on the HMD website.

Data resource access

Registration, which is quick and free of charge, is required before access to the data is provided. New users are guided through the registration process by accessing the following URL: [<http://www.mortality.org/registration>].

Related Databases

The Human Life Table Database (HLD), developed by many of the same researchers who helped built the Human Mortality Database, is another source of mortality data that can be found at: [<http://www.lifetable.de/>]. The HLD is a collection of life tables for many national and sub-national populations from both developed and developing countries. The HLD life tables originate from official and non-official publications and are uneven with respect to data quality.

Following the success of the Human Mortality Database, the Max Planck Institute for Demographic Research collaborated with the Vienna Institute of Demography to develop the Human Fertility Database (HFD). The HFD is available at: [<http://www.humanfertility.org/>]. Based on the same guiding principles as the HMD, it provides harmonized, detailed and high quality period and cohort fertility data series for 26 developed countries.

HMD in a nutshell

- The HMD was established to document the exceptional mortality decline of the modern era and to facilitate research into its causes and consequences.
- First published in 2002, the HMD is an open-access database with uniformly constructed series of mortality rates, life tables, death counts and population exposures for 37 countries and 46 populations, with data classified by age (up to age 110 +), sex, year of death and year of birth.
- Concentrating on countries with the highest quality data, the HMD contains extended time series with over 100 years of observation for one-third of the included countries.
- A rigorous methods protocol is systematically implemented to process the input data and construct all HMD series, accounting for various kinds of data idiosyncrasies and disruptions; the uniform methods ensure high consistency of the output data across time and countries.
- Due to transparent implementation, extensive data quality checks, uniform methods and high quality and comparability of the data, the HMD has become a common reference for mortality research; as of December 2014, over 1500 publications had cited the database.
- The data series are updated regularly and can be downloaded upon registration at: [<http://www.mortality.org/registration>].

Funding

The HMD receives financial and logistical support from its two sponsoring institutions, the University of California, Berkeley (UCB), and the Max Planck Institute for Demographic Research. The UCB team has been supported by awards (R01 AG011552 and R01 AG040245) from the U.S. National Institute on Aging (NIA); it also benefits from the supportive infrastructures of the NIA-funded Center on the Economics and Demography of Aging (CEDA) at UCB and of the French government-funded Institute for Demographic Studies (INED).

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or other funders, nor those of the institutions to which the authors are affiliated (the United Nations, the University of California, the Max Planck Institute for Demographic Research and the French Institute for Demographic Studies).

Acknowledgments

We are deeply indebted to scientific collaborators and members of national statistical offices from around the world who have supplied data or expert advice for particular countries or areas. We also wish to acknowledge the support received over the years from colleagues at the Department of Demography, the University of California, Berkeley and at the Max Planck Institute for Demographic Research. We are particularly grateful to the following people who have directly contributed at some point in the past or who continue to contribute to the construction of HMD data series: Evgeny Andreev, Kirill Andreev, Ludmila Andreeva, Michael Bubenheim, Vladimir Canudas-Romo, Eva Kibele, Rembrandt Scholz, Nadine Ouellette and Dimitre Philipov.

Conflict of interest: None declared.

References

1. Remund A. Jeunesses vulnérables? Mesures, composantes et causes de la surmortalité des jeunes adultes. PhD [dissertation]. Geneva, Switzerland: University of Geneva; 2015.
2. Wilmoth JR, Andreev K, Jdanov D *et al.* *Methods Protocol for the Human Mortality Database*. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. 2007. <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf> (12 May 2015, date last accessed).
3. Popham F, Dibben C, Bambra C. Are health inequalities really not the smallest in the Nordic welfare states? A comparison of mortality inequality in 37 countries. *J Epidemiol Community Health* 2013;**67**:412–18.
4. Berry DA, Cronin KA, Plevritis SK *et al.* Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;**353**:1784–92.
5. Antiretroviral Therapy Cohort Collaboration. Mortality of HIV-infected patients starting potent antiretroviral therapy: comparison with the general population in nine industrialized countries. *Int J Epidemiol* 2009;**38**:1624–33.
6. Cervantes F, Dupriez B, Pereira A *et al.* New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood* 2009;**113**:2895–901.
7. Koopman JE, Rozing PM, Kramer A *et al.* Senescence rates in patients with end-stage renal disease: a critical appraisal of the Gompertz model. *Aging Cell* 2011;**10**:233–38.
8. Ludwig H, Bolejack V, Crowley J *et al.* Survival and years of life lost in different age cohorts of patients with multiple myeloma. *J Clin Oncol* 2010;**28**:1599–605.
9. Oeppen JE, Vaupel JW. Broken limits to life expectancy. *Science* 2002;**296**:1029–31.
10. Shkolnikov VM, Jdanov DA, Andreev EM, Vaupel JW. Steep increase in best-practice cohort life expectancy. *Popul Dev Rev* 2011;**37**:419–34.

11. Goldstein JR, Wachter KW. Relationships between period and cohort life expectancy: Gaps and lags. *Popul Stud (Camb)* 2006;**3**:257–69.
12. Wilmoth JR, Deegan LJ, Lundström H, Horiuchi S. Increase of maximum life-span in Sweden, 1861-1999. *Science* 2000;**289**: 2366–68.
13. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. *Lancet* 2009;**374**:1196–208.
14. Christensen K, Thinggaard M, Oksuzyan A *et al*. Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. *Lancet* 2013;**382**:1507–13.
15. Grigoriev P, Shkolnikov V, Andreev E *et al*. Mortality in Belarus, Lithuania, and Russia: divergence in recent trends and possible explanations. *Eur J Popul* 2010;**26**:245–74.
16. Wilmoth JR, Boe C, Barbieri M. Geographic differences in life expectancy at age 50 in the United States compared with other high-income countries. In: Crimmins EM, Preston SH, Cohen B (eds). *International Differences in Mortality at Older Ages*. Washington, DC: National Research Council of the National Academies, 2010.