The Human Fertility Database: aims, data and methods

A. Jasilioniene, D. A. Jdanov, T. Sobotka*, K. Zeman*, E. Andreev, V. M. Shkolnikov and J. Goldstein

Max Planck Institute for Demographic Research, Rostock, Germany
*Vienna Institute of Demography, Austrian Academy of Sciences

Contact us: info@humanfertility.org

www.humanfertility.org

Prepared for a presentation at Session 11 of the XXVI IUSSP International Population Conference on September 28, 2009

Draft 24 September 2009

Abstract

The Human Fertility Database (HFD) is a joint project of the Max Planck Institute for Demographic Research (MPIDR) and the Vienna Institute of Demography (VID), based at the MPIDR in Rostock, Germany.

It provides open and user-friendly access to detailed, well-documented and high-quality data on period and cohort fertility. Special focus is put on birth order-specific data that are frequently not available from usual data sources.

The HFD aims at facilitating research on changes and inter-country differences in fertility in the past and in the modern era.

The idea of this database was largely inspired by the success story of the Human Mortality Database (HMD; www.mortality.org) directed by John R. Wilmoth. Both for the HMD and the HFD the guiding principles are comparability, flexibility, accessibility, and reproducibility.

1. Introduction. About the Human Fertility Database Project

The Human Fertility Database (HFD) is a joint project of the Max Planck Institute for Demographic Research (MPIDR) and the Vienna Institute of Demography (VID), based at the MPIDR in Rostock, Germany. It provides open and user-friendly access to detailed, well-documented and high-quality data on period and cohort fertility. Special focus is put on birth order-specific data that are frequently not available from usual data sources. The idea of this database was largely inspired by the success story of the Human Mortality Database (HMD; www.mortality.org) directed by John R. Wilmoth. Both for the HMD and the HFD the guiding principles are comparability, flexibility, accessibility, and reproducibility.

The first version of the Database is available as of 28 September 2009 at **www.humanfertility.org**. The website initially features period and cohort fertility data as well as parity-specific fertility tables for six countries with total population around 500 million (United States, Russian Federation, Austria, the Czech Republic, the Netherlands, and Sweden).

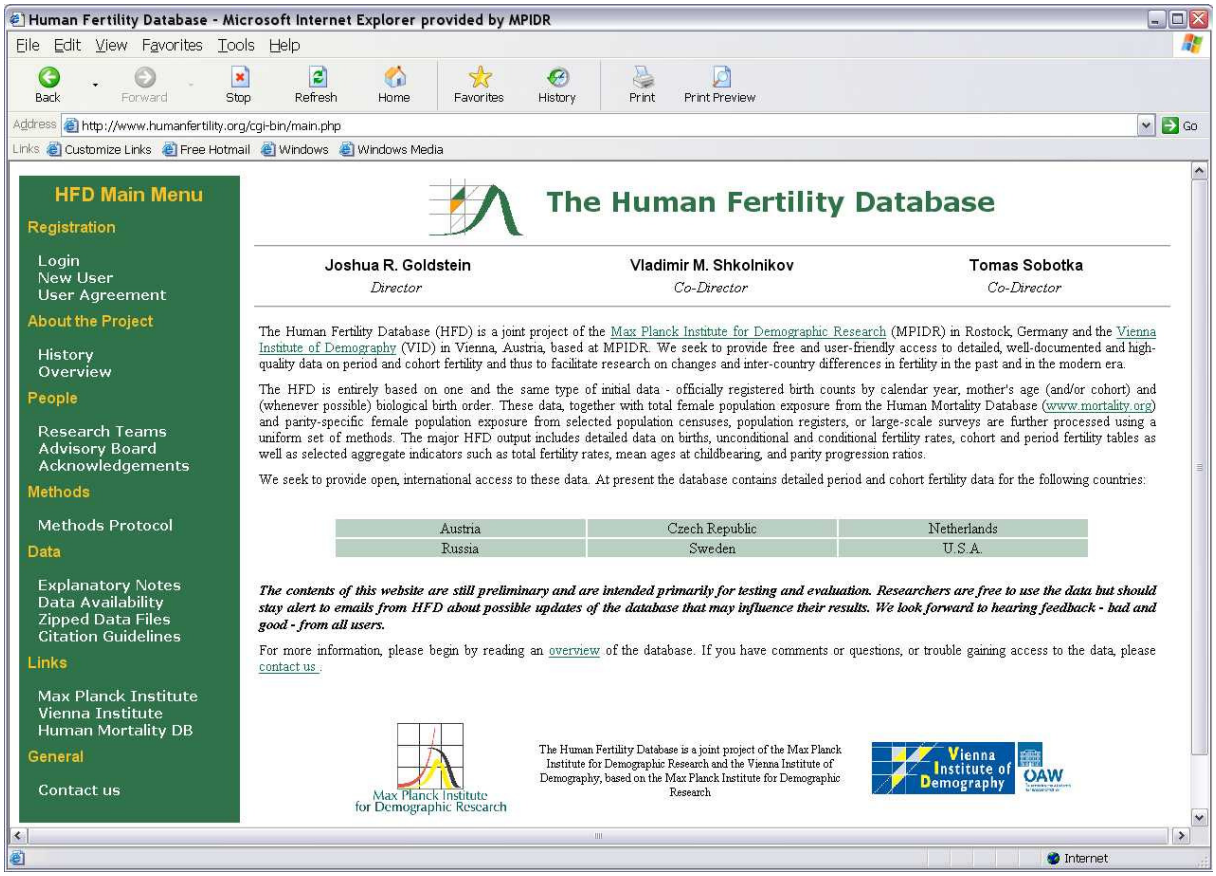
The main goal of the Human Fertility Database is to provide access to detailed high-quality data on cohort and period fertility to a broad audience of users. We are trying to develop the HFD into an important resource for monitoring, analyzing, comparing, and forecasting fertility as well as for studying causes and consequences of fertility change in the industrialized world. The uniform format of HFD data will facilitate comparative analysis across countries and regions, and encourage analysts to move beyond the simplest indicators such as the period Total Fertility Rates.

The Human Fertility Database provides age-, cohort- and (whenever possible) birth-order-specific fertility rates, cumulative and total fertility rates, mean ages at birth and also cohort and period fertility tables for national populations or areas. In addition, the HFD provides input data from which these measures and tables are being computed. The input data consist of detailed birth counts and estimates of female population exposure obtained from officially recognized sources.

The following features should make the HFD particularly attractive to its users:

- High level of detail which will provide the possibility to address different data needs and research questions
- Uniformity of methods and data design, which implies comparability of all the data across countries, cohorts, and periods
- The emphasis on displaying order-specific fertility indicators, which should encourage higher level of sophistication in fertility analyses and forecasts and further innovation in methodological research on fertility
- Free access to all data upon registration

Figure 1: The Human Fertility Database Project website



2. Motivation for starting the Human Fertility Database Project

At present, data on fertility in industrialized countries are fragmented and not easily comparable across countries, time periods, and cohorts. This is particularly the case for parity-specific fertility indicators, which are crucial for understanding fertility behavior. The data limitations represent major obstacle for research and policy making related to ongoing changes of childbearing patterns, decrease and postponement of childbearing and its future prospects. They also become more obvious over time as more attention among the policy-makers, media and wider public is being paid to low fertility in industrialized countries. Such issues can be addressed properly only when high-quality data on both period and cohort fertility become available.

3. History of the Human Fertility Database

Work on the Human Fertility Database began in 2007 as a collaborative project involving research teams at the Max Planck Institute for Demographic Research in Rostock and the Vienna Institute of Demography in Vienna.

The idea of the Human Fertility Database was largely inspired by the success story of the Human Mortality Database that has become an excellent resource of high-quality mortality data for the international research community since its launch in 2002. Vladimir Shkolnikov and Dmitri Jdanov, who are intensively engaged with the HFD, form also a part of the HMD

Team in Rostock and work in a close collaboration with John Wilmoth and his colleagues from the US HMD Team. Their expertise and experience with such a major data collection effort has enabled the rapid take-off of the HFD project.

The last decade saw several important research results closely related to extensive collection of fertility data. In particular, numerous studies on international fertility trends by Gérard Calot, Jean-Paul Sardon and Tomas Frejka, based on an extensive collection of national fertility data, indicated the advantages of analyzing detailed fertility data in a comparative perspective and highlighted the need for more intensive efforts in international data collection.

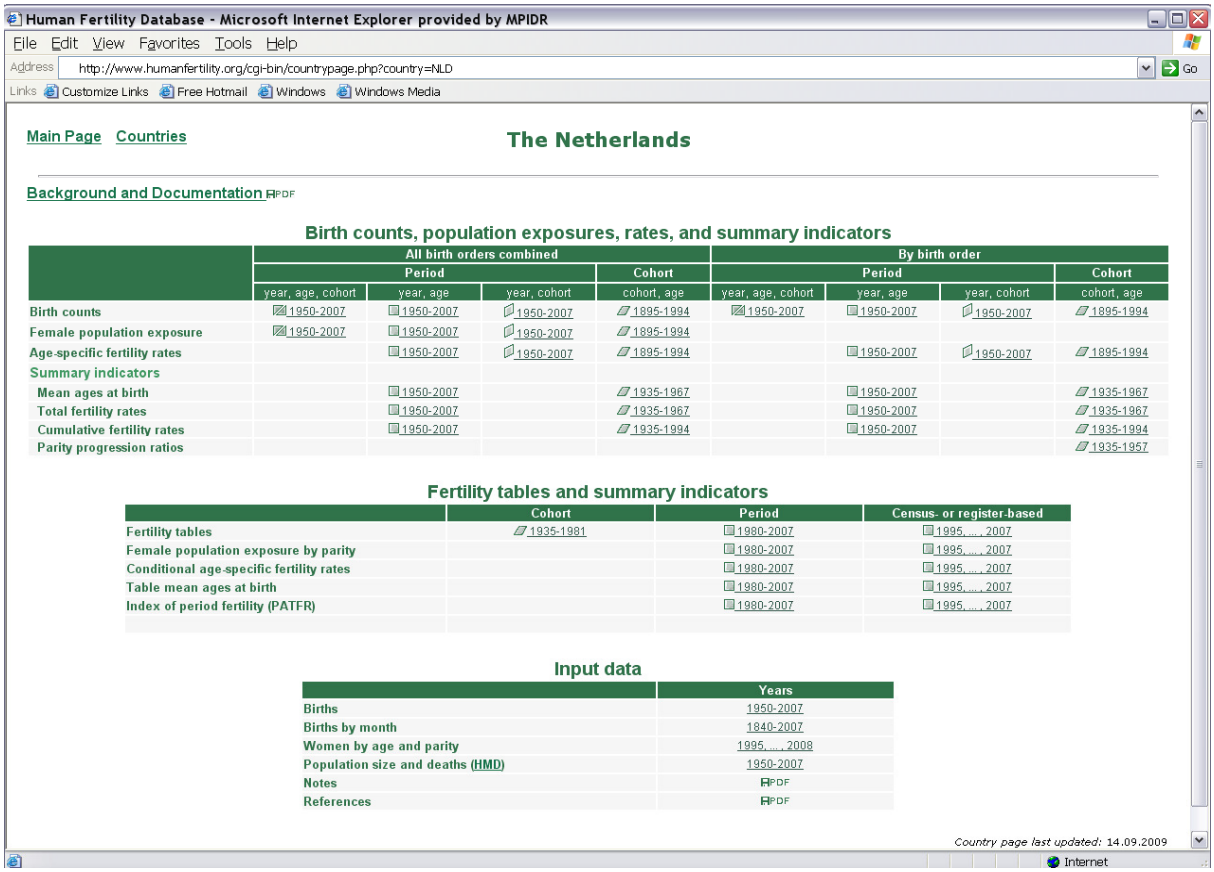
At the First HMD Symposium (Rostock, 2004) John Wilmoth proposed the idea of a fertility database, similar to the HMD. During 2006-2007, there were extensive consultations involving Tomáš Sobotka, Vladimir Shkolnikov and other researchers from both MPIDR and VID. They all shared the view that the low availability of standardized, detailed and comparable fertility data hinders 'state of the art' research on contemporary fertility.

In 2007, Joshua Goldstein, the MPIDR Director, decided to set up the Human Fertility Project at MPIDR and to head this project. The project was included in research plans of the Laboratory of Demographic Data at MPIDR. In parallel, the VID (specifically, Dimiter Philipov and Tomáš Sobotka) and the French National Demographic Institute INED, in particular Laurent Toulemon), included the initiation of fertility database as one of the goals in the eventually successful application for the EU-funded REPRO project in May 2007 and the VID Director, Wolfgang Lutz, included the HFD project in the research programme of VID.

The HFD is designed as a database combining the best of the earlier methodological research and data collection efforts into a single project. It provides detailed data on cohort and period fertility produced by means of scientifically sound and uniform methods, and presented in a user-friendly way that facilitates studies by a broad audience of users.

The first outline of the HFD was presented and discussed in depth with Jean-Paul Sardon, Laurent Toulemon, and John Wilmoth at the Second HMD Symposium held at MPIDR on June 13-14, 2008. The first HFD prototype was shown at the Population Association of America Annual Meeting held on April 30-May 2, 2009 in Detroit. The first fully functioning version of the HFD was demonstrated at the Session 11 at the XXVI Annual IUSSP International Population Conference in Marrakech on September 28.

Figure 2: Example of a country data page



4. Overview of the Human Fertility Database

For each population, one and the same set of methods is being applied for the production of uniform output data. This facilitates comparability of the HFD data and indicators across countries and time.

The HFD is limited to populations where the registration of births by official statistical agencies is virtually complete and where population estimates over the range of reproductive ages are reliable. Methods employed in the HFD for obtaining output data do not include any treatment or adjustment of the input data for completeness and coverage.

A complementary database—Human Fertility Collection (HFC)—will include fertility rates and indicators constructed by other researchers, research organizations and statistical agencies using various methods and data shapes. The HFC data will be, therefore, less consistent and comparable across countries and time than the HFD data. At the same time, the HFC will be more flexible and will contain fertility data for countries and years that cannot be included in the HFD, including estimates of order-specific fertility for countries where birth order registration is restricted to registering birth order within marriage only. It is planned that the HFC will also feature detailed data sets, historical data, and relevant documents.

Similarly to the Human Mortality Database we are following as much as possible four guiding principles: comparability, flexibility, accessibility, and reproducibility.

We provide complete documentation of all data available through this site as well as full descriptions of methods applied and specific features of country data sets. A complete description of the HFD methodology is given in the Methods Protocol. For each country, the description of data sources is given in the References document posted on the respective country page. General country-specific information (completeness, coverage, data quality issues, definitions, etc.) can be found in Background and Documentation files within each country section.

The HFD provides free access to the data. Before gaining full access to the database, you must become a registered user, which requires accepting our user agreement.
If you have comments or questions, or trouble accessing the database, please contact us.

The HFD process for computing output fertility indicators from input data on births and population can be briefly described as a sequence of six steps, which are specified below. In the process, the following data, indicators and outputs are produced:

- Unconditional cohort and period age- and (whenever possible) order-specific fertility rates from birth counts and population exposures,
- Summary indicators of cohort and period fertility from the unconditional age- and order-specific fertility rates,
- Cohort fertility tables from the unconditional age- and order-specific fertility rates,
- Data on the distribution of women by age and parity from censuses or population registers
- Conditional age and parity-specific fertility rates, period fertility tables by age and parity

Detailed descriptions of the HFD methodology are given in the Methods Protocol. The following items provide a concise overview of data processing and methodology.

Births. The HFD collects detailed annual data on live births over the longest possible time periods. Ideally, birth count data are classified by single years of age and year of birth (cohort) of the mother and by birth order of the child (biological birth order). In many cases, however, input birth counts are less detailed (see also Section 2 above). In many countries information about the birth cohort of mothers is not available. In some cases the age of the mother is available by five-year age groups rather than one-year age groups; especially for the period before 1960. To achieve the uniformity of data format with respect to age and birth order, additional splits and adjustments are being performed for the HFD. For many countries or time periods, birth data by birth order are not available. In such cases, order-specific fertility rates and order-specific mean ages at birth as well as cohort and period fertility tables can not be obtained and are not featured in the HFD.

Population denominators. Two types of data on female populations are used. The first (and predominant) type of population data specifies population exposure by age, needed for the computation of unconditional fertility rates. For most countries, female population data by age, presented in the HMD, are used. For countries that are not present in the HMD, the annual population by age is estimated according to the HMD methodology. The second type of population data includes counts of women by age and parity, needed for the computation of conditional fertility rates, which serve as the major input for period fertility tables. These data

are usually available from population censuses or registers and, in rare cases, from large-scale population surveys.

Fertility rates. Fertility rates are ratios of birth counts to corresponding population exposures. Unconditional age-specific fertility rates relate births specified by age of the mother and birth order of the child (when available) to all women of a given age. Conditional age- and order-specific fertility rates measure childbearing intensity among women of specific age and parity (e.g., second births are related to women of parity one only). Cohort fertility rates are computed for every combination of cohort and age. Period fertility rates are computed for every combination of calendar year and age. Furthermore, the HFD includes population exposures and births counts by Lexis triangles, making it possible for an advanced user to compute fertility rates and fertility tables in any configuration desired.

Cumulative and total fertility rates and mean ages at birth. These summary measures are based on fertility rates by age and (when possible) by birth order (incidence rates, central birth rates, rates of the second type). The period Total Fertility Rate (TFR) is computed as a sum of age-specific fertility rates across all ages from ≤12 to 55+. The HFD also displays the total fertility rate by age 40, based on a summation of age-specific fertility rates over all ages under 40. The TFR, which is also specified by birth order, shows the average number of children a woman from the population of interest would have by the end of her reproductive life if she experienced at each age the observed age-specific fertility rates for a given year. Cumulative fertility rates are based on a summation up to the indicated age limit, shown for each single age category. Mean age at birth and mean age at birth by age 40 are computed from the schedule of age-specific fertility rates. They show average ages at birth weighted by age-specific fertility rates over the entire range of reproductive ages or over reproductive ages under age 40, respectively. All the summary measures are computed for all birth orders combined and for specific birth orders by calendar year as well as by birth cohort.

Cohort fertility tables by age and parity. These are increment-decrement life tables, which model the process of childbearing in female cohorts by age and parity. In principle, they describe a two-dimensional cohort progression toward older age and higher parities. Women of the cohort of interest are moving from parity zero (i.e., from being childless) to parity one, from parity one to parity two, and to subsequent parities, by giving births of the corresponding birth orders. For each cohort, the life table functions are computed from the schedule of age- and parity-specific fertility rates as major input data. The distribution of births by age of mother and birth order in the table and the parity distribution of the table population of females correspond to the observed fertility trajectories of cohorts analyzed.

Period fertility tables by age and parity. Many functions in these tables are identical to those in cohort fertility tables and their construction is based on comparable formulas. The period fertility tables describe the fertility progression in a ‘synthetic cohort’ of women on the basis of conditional age- and parity-specific fertility rates observed during one calendar year. In other words, the tables give a period snapshot of fertility of many female birth cohorts and do not correspond to childbearing history of any real cohort. The key input in period tables is the age- and parity-specific distribution of the female population of reproductive age (exposure population, see “Population denominators” above). These distributions are obtained from cohort fertility tables, “golden” censuses that provide the parity distribution in one base year or directly from population censuses or registers. In the latter case, the fertility tables are census or register-based. The main output of the period fertility table is the summary index of period fertility controlling for age and parity, PATFR, and its order-specific

The raw data are collected from official and other validated sources, especially national statistical offices, statistical and demographic yearbooks, special tabulations from national registry data, official statistics websites, and Eurostat. We are collecting relevant documents by data providers and also scientific literature explaining data collection routines, related regulations and practices and other factors affecting the quality of data on births and female population.

Special attention is being paid to the following aspects of 'raw' data:

- completeness and timeliness of birth registration;
- coverage of the whole territory and all population groups by the birth registration;
- actual definitions of live birth and availability of data on live births rather than on all births;
- availability of information about biological birth order rather than information restricted to marital births or births within current marriage;
- age categories of mothers for which birth data are reported;
- reliability of census and registry reporting of the parity of women.

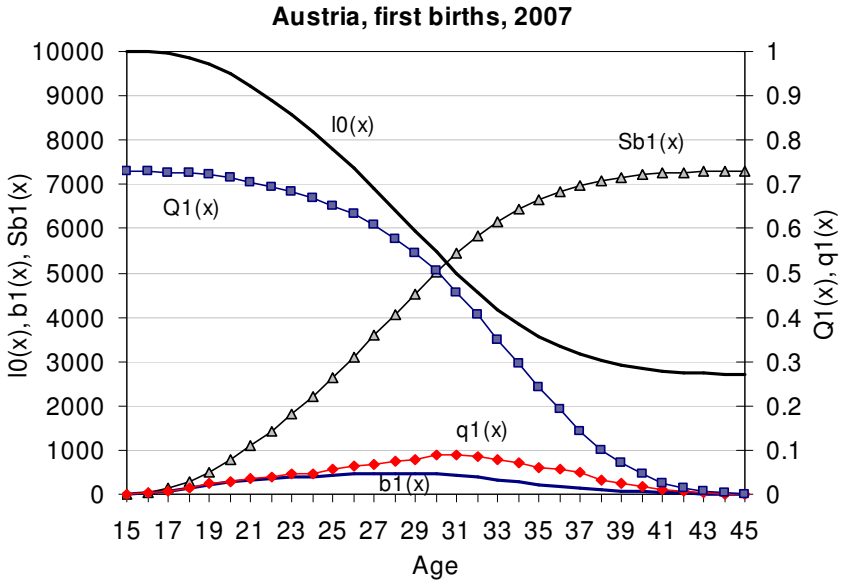
The HFD aims at providing opportunities for comparative studies on fertility in different countries and/or time periods. In this regard, consistency across the data entirety whole data universe is an important priority. That is why we apply a uniform set of procedures to each population.

The desire for uniformity is hindered by the significant variability in the original data formats and the lack of sufficient detail in these 'raw' data. For example, the original birth data can be provided for one-year vs. five-year age groups, they may not always include cohort dimension, they may show broader or narrower ranges of available ages, they may include births with unknown birth order, or they may show total births instead of live births.

The HFD methodology includes procedures for the transformation of any set of raw data into data classified by single years of age ranging from age 12 to 55+, by single-year birth cohorts, and (whenever possible) by birth orders varying from 1 to 5+. Births with unknown age of the mother are distributed proportionally across the range of known ages of the mother. Five-year age groups are additionally split into single-year ages by means of spline interpolation. For each age, births with unknown birth order are distributed proportionally across known birth orders. Birth orders higher than five are aggregated into birth order 5+. If needed, age-specific birth counts are extrapolated toward younger and older ages to cover the range of ages from 12 to 55 years. For each age, births are additionally split by year of birth of the mother (if such information is not present in input data).

At the same time, the data for each country are carefully checked and processed, with a view on their specificities, which are outlined in the country documentation file and, if needed, country experts are consulted. The data processing concerns checks for specific data problems and correction of errors, consultations with local demographers and statisticians, investigations of country documents and literature, comparisons with alternative fertility data and estimates. These procedures help to assure the high data quality standard for each HFD country.

Figure 3: Selected functions and indicators from parity- and age-specific period fertility tables



5. Conclusions and future plans

We hope that the database will expand rapidly in the next year. Our goal is to establish the HFD as a respected and widely used source of high-quality fertility data that will become a key resource for cross-country comparative studies and that will also stimulate innovative research on fertility data, indicators and methods.

By now we have established contacts with research experts and statistical agencies from more than 25 countries. We also aim to improve the website and data presentation, following the feedback from first HFD users. In 2010 we will inaugurate a companion data repository, the Human Fertility Collection. In addition, we plan to expand the HFD by adding a module with data on births, fertility rates and parity progression ratios based on duration since the previous births (birth interval indicators).

6. HFD team

Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany

- Joshua R. Goldstein (Director)
- Vladimir M. Shkolnikov (Co-Director)
- Aiva Jasilioniene
- Dmitri A. Jdanov
- Evgueni M. Andreev
- Jude Mikal
- Sigrid Gellers-Barkmann

Vienna Institute of Demography (VID), Vienna, Austria

- Tomáš Sobotka (Co-Director)
- Kryštof Zeman

7. Acknowledgements

Creating the Human Fertility Database is a very challenging endeavour, and we are very grateful to all our colleagues from various research institutions and statistical agencies for their extensive help.

This project would not have been possible without the encouragement and financial support of the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany and the Vienna Institute of Demography (VID), Vienna, Austria. The European Commission funded the initial work by the VID as part of the project "Reproductive decision-making in a macro-micro perspective" (REPRO) in the Seventh Framework Programme (Grant Agreement: SSH-CT-2008-217173).

Our special gratitude for their valuable comments on the HFD methodology and general advice to the project is due to the members of the HFD Advisory Board:

- John Wilmoth, University of California, Berkeley,
- Laurent Toulemon, INED, Paris,
- Jean-Paul Sardon, INED, Paris,
- Wolfgang Lutz, VID, Vienna

The HFD project greatly relies on the network that includes statistical agencies, other data holders, and country experts that help us in locating, documenting and accessing various data, understanding their origins, quality, and other important details. We thankfully acknowledge the contributions of the following researchers and institutions:

Austria

- Statistics Austria

Bulgaria

- Dimiter Philipov, Vienna Institute for Demography
- Iordan Kaltchev, National Statistical Institute of Bulgaria (NSI)

Canada

- René Houle, Statistics Canada

Czech Republic

- Tomáš Fiala, University of Economics, Prague
- Radek Havel, Czech Statistical Office (CZSO)
- Michaela Němečková, CZSO
- Jiří Vejrych, CZSO

England & Wales

- Julie Jefferies, Office for National Statistics
- Steve Smallwood, Office for National Statistics

Estonia
- Allan Puur, Estonian Interuniversity Population Research Centre

Finland
- Timo Nikander, Statistics Finland

France
- Laurent Toulemon, Institut national d'études démographiques (INED)
- Magali Mazuy, INED
- Jean-Paul Sardon, INED

Germany
- Michaela Kreyenfeld, Max Planck Institute for Demographic Research
- Olga Pötzsch, Federal Statistical Office of Germany

Italy
- Sabrina Prati, National Institute of Statistics (ISTAT)

Lithuania
- Vlada Stankuniene, Institute for Social Research

Netherlands
- Elma van Agtmaal-Wobma, Statistics Netherlands
- Trudy Lisci-Wessels, Statistics Netherlands

Norway
- Trude Lappegård, Statistics Norway
- Svein Inge Holm, Statistics Norway

Poland
- Irena Kotowska, Warsaw School of Economics
- Krzysztof Tymicki, Warsaw School of Economics

Portugal
- Maria Filomena Mendes, University of Évora
- Isabel Tiago de Oliveira, Instituto Superior de Ciências do Trabalho e da Empresa
- Statistics Portugal

Romania
- Aura-Mihaela Alexandrescu, National Institute of Statistics, Romania
- Cornelia Muresan, Babeş-Bolyai University, Cluj-Napoca

Russia
- Evgueni Andreev, Max Planck Institute for Demographic Research

Slovakia
- Michaela Potančoková, Demographic Research Centre, INFOSTAT

Slovenia
- Apolonija Oblak Flander, Statistical Office of the Republic of Slovenia
- Tina Žnidaršič, Statistical Office of the Republic of Slovenia

Spain
- Daniel Devolder, Universitat Autònoma de Barcelona
- Anna Cabre Pla, Universitat Autònoma de Barcelona

Sweden
- Lotta Persson, Statistics Sweden
- Gunnar Andersson, Stockholm University

Switzerland
- Stéphane Cotter, Swiss Federal Statistical Office

USA
- Ward Kingkade, US Census Bureau