Round Table 1: Data needs and data funding

- How should HFD project develop in the future?
- How can it become more useful to its actual and potential users?
- How can high-quality data repositories “compete” with the “quick and dirty” data covering all the world and easily available online?
- How to secure sustainable long-term financing of HFD and similar data projects?
“Classical” sources of demographic data at population level & their limitations

- Publications by national statistical offices / other official institutes
  - Limited to one country
  - More detailed data or documentation often unavailable in English
  - Mostly focused on recent data, without historical series
  - No universal format of data, vastly different levels of detail

- Collections of official data (Eurostat, WHO mortality database)
  - Data published as they are provided
  - Mixture of “good” and “bad” data
  - Limited historical coverage
  - Data not readily comparable across time and countries
  - No documentation about data collection, estimations, reliability, data manipulation, etc.

- Official reports or data collections which must cover specific range of countries (UN, World Bank, CIA,…)
  - Mixture of real data, sample-based estimates, and modeled data
  - Data are not comparable across time and countries
  - Relatively short data series
  - Fragmentary documentation
HFD: a new standard for demographic databases

• Why to spend resources on production of demographic data in times of easy data availability on the Internet?
Substantial “added value” compared to other data sources: very detailed, well documented, uniform, high quality data, comparable across time and space.

• HFD main principles:
free & open access to all raw data, thorough documentation of data sources and computational methods, data in user-friendly format, data comparability (across countries and time), reproducibility, flexibility, responsibility for the data => data are ready for analysis

• Target users:
researchers, university faculty members, students.

Principal data flows

Search for original data and data descriptions from official sources. Preliminary checks. → Input Database of raw data. Various data shapes. → Uniform data: event counts and pop exposures by age, calendar year, and cohort. → Output Database of resulting tables and aggregate measures.
An Example of Data Adjustment: SLOVAKIA

• Since 2012 the SÚSR do not include births abroad to the official statistics of births

• It concerns ~ 10% births: TFR is underestimated (2013: 1.34 X 1.47) and also MAB1 is different (26.9 X 27.7)

• For 2012-2014 we got data including births abroad unofficially from our colleague
Prospects

HFD
- New countries (Australia, Belgium, South Korea, New Zealand, Romania,…)
- New methods (calculation of population exposures, splitting aggregated data)
- HFD “lite” for developing countries? China, India, Latin America,…

HFC
- New section on male fertility in HFC (17 countries)
- Regional data (44 countries)
- Cohort fertility
Challenges

Current:
- Low quality of population and fertility data
- Series breaks in the birth data
- Regular updates

Future:
- Personal data protection
- HFD/HFC expansion: which regions, data, factors to focus on?

Permanent:
- Funding of the HFD project
- Promotion of the project
Challenges

- Funding of the HFD project
- Promotion of the project

GBD 2017 Population and Fertility Collaborators

Summary

Background Population estimates underpin demographic and epidemiological research and are used to track progress on numerous international indicators of health and development. To date, internationally available estimates of population and fertility, although useful, have not been produced with transparent and replicable methods and do not use standardised estimates of mortality. We present single-calendar year and single-year of age estimates of fertility and population by sex with standardised and replicable methods.

Methods We estimated population in 195 locations by single year of age and single calendar year from 1950 to 2017 with standardised and replicable methods. We based the estimates on the demographic balancing equation, with inputs of fertility, mortality, population, and migration data. Fertility data came from 7817 location-years of vital registration data, 429 surveys reporting complete birth histories, and 977 surveys and censuses reporting summary birth histories. We estimated age-specific fertility rates (ASFRs; the annual number of livebirths to women of a specified age group per 1000 women in that age group) by use of spatiotemporal Gaussian process regression and used the ASFRs to estimate total fertility rates (TFRs; the average number of children a woman would bear if she survived through the end of the reproductive age span [age 10–54 years] and experienced at each age a particular set of ASFRs observed in the year of interest). Because of sparse data, fertility at ages 10–14 years and 50–54 years was estimated from data on fertility in women aged 15–19 years and 45–49 years, through use of linear regression. Age-specific mortality data came from the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2017 estimates. Data on population came from 1257 censuses and 761 population registry location-years and were adjusted for underenumeration and age misreporting with standard demographic methods. Migration was estimated with the GBD Bayesian demographic balancing model, after incorporating information about refugee migration into the model prior. Final population estimates used the cohort-component method of population projection, with inputs of fertility, mortality, and migration data. Population uncertainty was estimated by use of out-of-sample predictive validity testing. With these data, we estimated the trends in population by age and sex and in fertility by age between 1950 and 2017 in 195 countries and territories.
Open Data and Funding

Open Data:
**Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.

**Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.

**Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups.

Financing HFD and similar projects:
funding schemes do not support existing data infrastructure
**Open Data and Funding**

**Open Science and Open Data:** availability and access; re-use and redistribution; universal participation.

*Who should pay for open data?*

Financing HFD and similar projects: funding schemes do not support existing data infrastructure

- Grants for research (might include generation of *new* data sets)
- HFD requires big resources which can’t be allocated as “additional costs” in research grant
HUMAN FERTILITY DATA PROJECT
A joint venture of Max Planck Institute for Demographic Research and Vienna Institute of Demography