

Poster Session for Demographic methods and data (P17)

The Human Fertility Collection: an emerging source of demographic data

O. Grigorieva¹, D. Jdanov¹, T. Sobotka², K. Zeman², A. Jasilioniene¹,
K. Kubisch¹, V. Shkolnikov¹, and J. Goldstein¹

¹Max Planck Institute for Demographic Research, Rostock

²Vienna Institute of Demography, Vienna

ABSTRACT

The Human Fertility Collection (HFC) is a joint project of the Max Planck Institute for Demographic Research (MPIDR) and the Vienna Institute of Demography (VID). It is designed to supplement the Human Fertility Database (www.humanfertility.org) and provide a variety of fertility data collected from various external sources. In contrast to the HFD, the collection comprises data not only from official statistics but also published by individual researchers and research organizations and may involve additional dimensions (regions, ethnic groups, etc.). The relevant documentation is also offered to the users. The HFC aims at including a wide range of countries and covering the longest possible time periods. The data in the HFC are provided in the standardized format and can be easily and openly accessed. The HFC website is expected to be launched in August 2013 (www.fertilitydata.org). The database will contain unconditional fertility rates and selected aggregate indicators for a number of national and subnational populations. Whenever possible, order-specific data will be also made accessible to the users. This paper describes the HFC, computational procedures used for data harmonization as well as its structure.

INTRODUCTION

The Human Fertility Collection (HFC, www.fertilitydata.org) is designed as a complementary data source for the Human Fertility Database (www.humanfertility.org). It intends at providing free and open access to a variety of fertility data that due to high quality restrictions can not be included in the principal database. Similarly to the HFD, the data in the HFC can be obtained in standardized format. In the contrast, however, the HFC offers data obtained not only from official statistics but also published by individual researchers and research organizations. In general, the HFC data are less consistent and comparable across countries and time than the HFD data but the collection is more flexible and contains fertility data involving additional dimensions (e.g., regions, ethnicity, etc.). The raw data and the relevant documentation are also made accessible to the users.

The HFC website will be launched in August 2013 and will contain unconditional fertility rates and selected aggregate indicators for a number of countries. Whenever possible, birth-order-specific data will be also provided.

HFC DATA SOURCES

At the initial stage, the major source of the HFC data is the countries' statistical agencies. In addition, the data from the European Demographic Observatory (ODE collection, kindly passed to the MPIDR by one of its creators, Jean-Paul Sardon) are incorporated in the HFC. Further expansion of the HFC includes adding more data published by individual researchers, for instance by R. Heuser (1976) and L. Toulemon.

HFC INDICATORS AND SELECTION PRINCIPLE

Unconditional age-specific fertility rates¹ (ASFR) are input data obtained from the original sources. The rates are used for further computation of cumulative period fertility rates (CPFR), total fertility rates (TFR) and mean ages at birth (MAB).

In general, to be included in the HFC data should meet following minimal requirements:

1. Data fit to any of the following age scales:

- a) for single ages: 14 (or 14 and younger), 16, 17,..., 49, 50+;
- b) for 5-year age groups: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49.

If original data have wider age classification, i.e. age groups before or after the youngest/oldest age group of above mentioned age scales (e.g., 12, 13, 14, ..., 49, 50, ..., 55+), these data are also considered in the HFC but additional computations are performed for harmonizing the data.

2. Data by birth order (when available) should be at least up to birth order 3+.

DATA FILES AND FORMATS

The following types of files are available in the HFC:

- ✓ Pooled data files consisting of data for a single country or the whole HFC in one file. There are four kinds of such files:
 - comprising of ASFR by aggregated (5-year) age groups with the country's original age scales;

¹ Unconditional age-specific fertility rates are obtained by dividing the births to women at age x in a given year t by person-years lived in that year by all women of this age, irrespective of their parity status.

- containing of ASFR by single age groups with the country's original age scales;
- consisting of ASFR and CPFR by single age groups with the standard age scale for all countries: 14 and below, 16, ..., 50+;
- containing of TFR and MAB estimated by the HFC team.
- ✓ Single files consisting of country's ASFR provided by the original age scale/birth order and obtained from a particular source;
- ✓ Original raw data in form of electronic data files or scanned copies of published data;
- ✓ Complete lists of references and data sources;
- ✓ List of notes.

All data files are provided for all birth orders combined, and when available, by birth orders. Below is an example of the structure of the pooled and single data files.

Table 1. Names and description of the fields in the pooled data file consisting of ASFR and CPFR

Name	Short description
Country	Country codes. Standardized international country codes maintained by International Organization for Standardization (ISO 3166-1 alpha-3 codes).
Region	A numeric code to specify the regional coverage (when necessary), otherwise “.”
Residence	A code to define the residence (1 - total, 2 - urban, 3 – rural)
Ethnicity	A code to dentine the ethnic group the data are given for (when available), otherwise “.”
Year1	Beginning of the calendar period
Year2	End of the calendar period
Age	Lower limit of age interval (age x)
AgeInt	Length of age interval. Usually, has a value of 1, or 5. Values “-99” and “99” are used for the lower and the upper open age intervals, respectively.
AgeDef	Age definition: age in completed years (ACY) or age reached during the year (ARDY), or “.”
Vital	Vitality of births (1- live births, 2 - total, or “.”)
ASFR	Age specific fertility rate for all birth orders combined
ASFR1 to ASFR5p	Age specific fertility rates for corresponding birth orders
CPFR	Cumulative period fertility rate by age x for all birth orders combined
CPFR1 to CPFR5p	Cumulative period fertility rate by age x for corresponding birth orders
Source	Abbreviation for the data source that differentiates between the origins of the data (e.g., statistical office, special collection, etc.)
RefCode	The systematic data set number offering to the user the country-source specific information.
Note	A code of the note which is defined in the Note file
Split	A numeric code to specify if splitting is applied to data (0 – original data; 1 – calibrated spline estimator, 2 – Hermite spline).

DATA PREPARATION AND SELECTED COMPUTATIONS

In general, data preparation includes the following steps:

1. Splitting of ASFR by 5-year age groups into ASFR by single ages (if original data are not available by single ages). The procedure is done using the Calibrated spline estimator²;
2. Splitting of open age intervals (e.g., 15- into 14- and 15), which is based on the Hermite spline;
3. Aggregation of data by birth order (for birth orders 3+ and 4+);
4. Calculation of cumulative fertility rates including total fertility rate;
5. Estimation of the mean age at birth.

Cumulative fertility rate is a hypothetical construct that can be interpreted as the average number of children that would be born to a woman by age x if she experienced at all ages below x the set of age-specific fertility rates observed in a given year. The cumulative rate is computed as follows:

$$CPFR(x, t) = \sum_{z=x_{\min}}^{x-1} n_z f(z, t) \quad (1)$$

where, $f(z, t)$ is the ASFR for a specified age interval; x and z refer to the age in completed years (ACY) or the age reached during the year (ARDY); $x_{\min}$ corresponds to the lowest age at childbearing considered in the analysis; n_z refers to the length of age interval.

The **Total fertility rate** represents the mean number of children a woman would have by the end of her reproductive life if she experienced at each age the age-specific fertility rates observed in a given year:

$$TFR(t) = \sum_{z=x_{\min}}^{x_{\max}} n_z f(z, t) \quad (2)$$

Mean age at birth refers to the average age of the mother at childbearing. In the HFC, the mean age at birth is calculated by weighting the ages by the ASFRs:

$$MAB(t) = \frac{\sum_{z=x_{\min}}^{x_{\max}} \bar{x} \cdot f(z, t)}{\sum_{z=x_{\min}}^{x_{\max}} f(z, t)} \quad (3)$$

Where, value $\bar{x}$ is the mean age at birth within the elementary age interval $[x, x+1)$:

$$\bar{x} = x + a(x) \quad (4)$$

with $a(x)$ being the average share of the age interval $[x, x+1)$ lived before giving birth to a child (all $a(x)$ values are assumed to be equal to 0.5 for ACY and zero for ARDY).

² The method was proposed by C.P.Schmertmann (2012). Calibrated spline estimation of detailed fertility schedules from abridged data. MPIDR working paper, 2012-022.