The Human Fertility Database:
Aims, Data, Methods, and New Research Opportunities

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

Max Planck Institute for Demographic Research, Rostock
[*]Vienna Institute of Demography, Vienna

Version 6 March 2009
Please do not cite without permission from the authors

Abstract

At present, data on fertility in industrialized countries are fragmented and not easily comparable across countries, time periods, and cohorts. 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. Based on these observations, the Max Planck Institute for Demographic Research and the Vienna Institute of Demography have initiated the Human Fertility Database project, which aims to fill in the gaps in fertility data availability and comparability. In the paper, we describe data processing in the HFD, discuss methodological challenges and provide examples of cohort and period data that it will contain when complete. To show the richness of the HFD and the opportunities that it opens up for research in fertility, we provide some illustrations based on Austrian data.

Contents

1. About 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.

The Human Fertility Database is a joint project of the Max Planck Institute for Demographic Research (MPIDR) and the Vienna Institute of Demography (VID), based at MPIDR, which aims to fill in the gaps in fertility data availability and comparability. The main goal of the database is to document changes and inter-country differences in fertility in the past and in the modern era. We seek to provide open and user-friendly access to detailed and high-quality data on the period and cohort fertility to the international research community and other interested users. Following the example of the Human Mortality Database (HMD, www.mortality.org and www.humanmortality.de) our guiding principles are comparability, flexibility, accessibility, and reproducibility. In this challenging endeavor, we greatly benefit from advice from the HFD Scientific Advisory Board.

The HFD Project includes two principal parts:

1) Human Fertility Database (HFD). The HFD will be entirely based on one and the same type of initial data: the original, officially registered, birth counts by calendar year, age (and/or cohort) of mother and (whenever possible) biological birth order. These data, together with the exposure population estimates taken from the HMD, selected population censuses and register data will be used for further computations by the uniform set of methods. The major HFD output will include detailed fertility rates, cohort and period fertility tables as well as selected aggregate indicators.

2) Human Fertility Collection (HFC) is a companion part of the project, which will hold a range of fertility data that can not be included in the HFD. The HFC will include sets of fertility rates and fertility tables constructed by other researchers, research organizations, and statistical agencies. Produced figures, though reliable and informative, will not be always comparable across countries and time periods due to variability in data origins and estimation methods.

Original data received from national statistical offices, research organizations, and publications will be fully documented. Methods for the processing and calculations of fertility rates will be described in detail in the methods protocol. The HFD will rely on the support of solicited country experts who will help to describe statistical systems and peculiarities of fertility data, to collect these data and to prepare documentation for individual countries. Their contribution will be indispensable for the successful development of the HFD. At present, we are in contact with the experts and have received or expect to receive the data soon from the following countries: Austria, Bulgaria, Canada, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Japan, Lithuania, the Netherlands, Norway, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom and USA.

We believe that the HFD and HFC will become reliable and important data resources for demographers and other scientists as well as students, policy makers, and other interested users.

The purpose of this paper is to present the Human Fertility Database, the major part of the project. We present the scheme of data processing in the HFD, point out some

methodological challenges and provide examples of cohort and period data that it will contain when complete. To show the richness of the HFD and the opportunities that it opens up for research in fertility, we provide some illustrations based on Austrian data.

2. Human Fertility Database (HFD)

The HFD, which is the principal constituent of the project, will feature births and fertility rates specified for periods and cohorts as well as for two different types of order-specific fertility rates – unconditional rates (named also *rates of the second kind* or *incidence rates*) and conditional rates (named also *intensities* or *occurrence-exposure rates*). It will also provide a number of indicators derived from period and cohort fertility tables by birth order.

2.1 Data processing in the HFD (scheme)
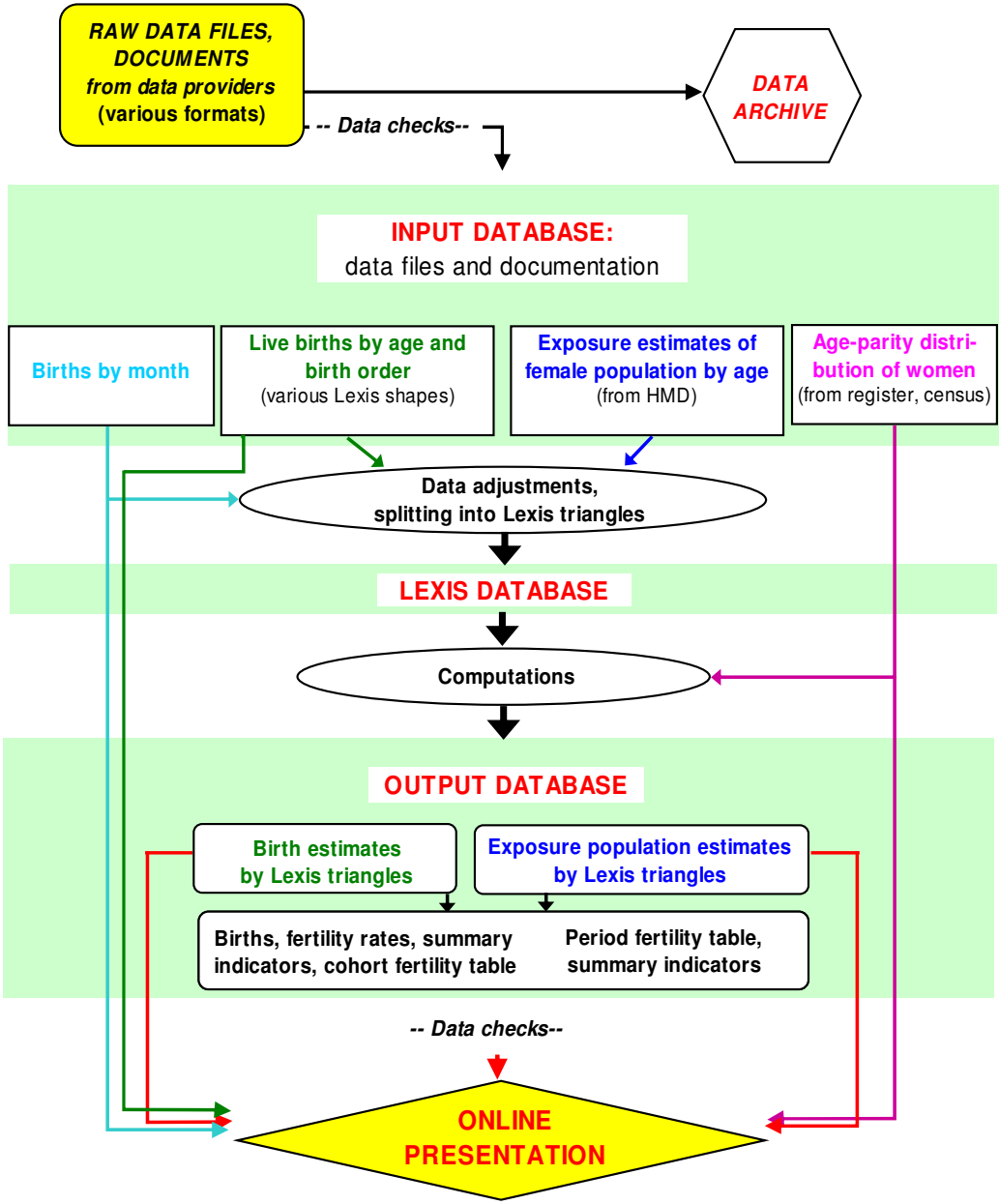
2.1.1 Raw data. Data archive

The data archive stores raw data files and all accompanying documents, supplied by the data providers (i.e., by the country experts, statistical offices, and research institutions). Large part of the data is obtained or purchased directly from the statistical offices. Raw data arrive in a variety of formats (data files, scanned or paper copies of official statistical publications, etc.)

Our standard data request asks for as long time series as possible of the following types of data:

1) Live births by calendar year, age of mother, and mother's year of birth.
2) Live births by calendar year, age of mother, mother's year of birth, and birth order.
3) Distribution of women by age, year of birth, and number of biological children ever born.
    This information usually comes from population registers, population censuses or large-scale surveys representative of the entire population. Estimates of the age-parity composition of the female population are required for the calculation of conditional fertility rates and the construction of fertility tables.
4) Total number of births by month.
    These data are used to estimate the effect of seasonality on the distribution of births during a given year.
5) Female population size on January 1. Population estimates are required for the countries and periods that are not included in the Human Mortality Database and for which the exposure-to-risk is not readily available.

There are four possible data configurations as they appear on the Lexis diagram: *horizontal parallelogram*, *vertical parallelogram*, *square* or *rectangular*, and *triangle* (see Caselli and Vallin (2006)[1] for more extensive explanation).

Horizontal parallelogram (F1) groups together events occurring during two consecutive years t and $t + 1$ within one birth cohort $t - x$ at age x last birthday. It represents events classified by age in completed years and year of birth (birth cohort).
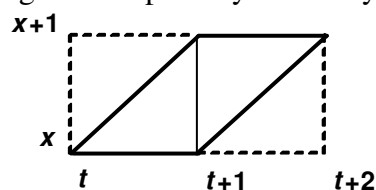
F1: horizontal parallelogram (extending over two units on the x-axis)

Vertical parallelogram (F2) groups together events occurring during the year t to members of the cohort $t - x$ at ages $x - 1$ and x last birthday. In this way, events are classified by calendar year and year of birth (birth cohort) and or by calendar year and age reached during the year.

---

1. [1] Caselli, G. and Vallin, J. (2006). From Situating Events in Time to the Lexis Diagram and the Computing of Rates; in Caselli, G., Vallin, J., and Wusch, G. (Eds.) *Demography and Synthesis*, Vol. 1. Amsterdam et al.: Elsevier, pp. 55-68.
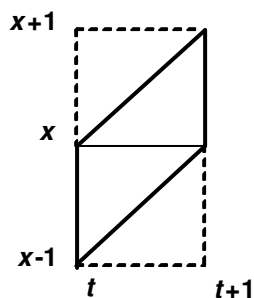
F2: vertical parallelogram (extending over two units on the y-axis)

Square or *rectangular* (F3) groups events occurring during the year t to members of two birth cohorts $t - x$ and $t - x - 1$, who have the same age x last birthday. Events are classified by age in completed years and calendar year.
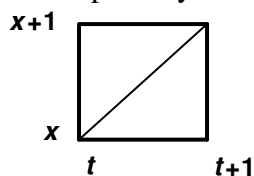
F3: Lexis square (or rectangular)

Lexis triangles allow the reconstruction of all the three formats described above. There are two types of triangles in the Lexis diagram: lower Lexis triangle and upper Lexis triangle. *Lower Lexis triangle* (F4) concerns births (B_L) born in the year t to the birth cohort $t - x$ at age x last birthday. *Upper Lexis triangle* (F5) concerns births (B_U) born in the year t to the birth cohort $t - x - 1$ at age x last birthday. Thus, events are classified by calendar year, age in completed years and year of birth (birth cohort).
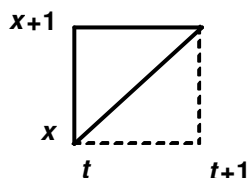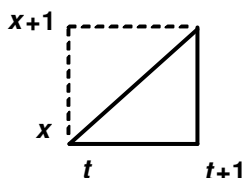
F4: Lower Lexis triangle   *F5: Upper Lexis triangle*

If births by triangle are known, births by squares and the two types of parallelogram can be obtained:

Horizontal parallelogram: $B = B_L(x, t, t-x) + B_U(x, t+1, t-x)$
Vertical parallelogram: $B = B_U(x-1, t, t-x) + B_L(x, t, t-x)$
Square: $B = B_L(x, t, t-x) + B_U(x, t, t-x-1)$

2.1.2 Input Database

The Input Database stores originally collected country raw data, converted into standard input format. For every country included in the HFD, there will be <u>four types of input data files</u>, containing data on:

    1) Births by calendar year, age of mother, mother's year of birth, and birth order,

2) Age-parity distribution of the female population (from population register, population census and large scale survey),
3) Births by month,
4) Female population exposure by age (for countries and periods not available in the HMD).

Additionally, the Input Database includes <u>three text files</u> for each country:
1) *Background and documentation*, providing information about data quality, coverage, territorial changes, historical peculiarities, and other important comments. This document is prepared by the country expert.
2) *Notes*, providing details about specific data points in the data files,
3) *References* to the data sources.

All input files will be available on the HFD website and will be accessible to HFD users unless there are certain restrictions imposed by the data suppliers. For convenience of HMD and HFD users, the HFD files will be named in the same manner as the HMD files. The files available for each country will be as follows:

- XXXbirths.txt (birth data),
- XXXparity.txt (age-parity distribution of women from population register, population census or survey),
- XXXmonthly.txt (births by month of birth),
- XXXpop.txt (female population exposure by age, for countries and periods not available in the HMD),
- XXXcom.pdf (background and documentation),
- XXXnote.pdf (notes for individual data points),
- XXXref.pdf (references).

'XXX' stands for a 3-letter (uppercase) code representing a country name (AUT for Austria, CAN for Canada, etc.). The input data files will be available in ASCII format, and the text files (XXXcom.pdf, XXXnote.pdf, and XXXref.pdf) will be available in Portable Document Format. The input data files will have headings (first line), uniform to all HFD countries. Comma (",") will be used as a field delimiter. Missing values will be coded as a single dot (".").

Structure of the input file for births

<u>Heading</u>: *PopName, Area, Year, YearReg, Age, AgeInt, Cohort1, Cohort2, BirthOrder, OrderInt, PrevBirth, DurInt, Births, Access, Note1, Note2, Note3, RefCode, LDB*

1. *PopName* (the same code 'XXX' as in the file name); population name
2. *Area* (1-digit); territorial/country coverage ('1' means full coverage).
3. *Year* (4-digit); year in which the birth occurred (calendar year).
4. *YearReg* (4-digit); the registration year of the birth. Some countries report births registered in current year, while they actually occurred in previous year. In such cases, *Year*, *Age,* etc. refer to the date of occurrence, and *YearReg* shows the registration year. If the birth was registered in the year of its occurrence, *YearReg* is denoted by '1'and thus is equal to *Year*.
5. *Age* (2-digit; UNK=unkown, TOT=total); age of the mother.

6. *AgeInt* (1- or 2-digit; '–' or '+' for open interval, '.' for unknown or total age); the length of age interval.
7. *Cohort1* (4-digit; '.' for unknown or total age, for open age intervals), year of birth of the mother.
8. *Cohort2* ('1' or 4-digit; '.' for unknown or total age, for open age intervals), year of birth of the mother. If births are classified by Lexis triangles, Cohort2 is equal to Cohort1. If birth counts are provided by Lexis squares, vertical or horizontal parallelograms, Cohort1 shows the birth cohort corresponding to the lower triangle in a given Lexis element and Cohort2 shows the birth cohort in the upper triangle. If the data are sorted by multiple age groups (e.g., five-year age groups), Cohort1 corresponds to the youngest cohort and Cohot2 is related to the oldest cohort.
9. *BirthOrder* (1- or 2-digit; UNK=unknown); birth order (1, 2, …).
10. *OrderInt* (1- or 2-digit; '+' for open interval, '.' for unknown birth order), the length of birth order interval.
11. *PrevBirth* (4-digit; '.' shows that duration-data are not available); year of previous birth.
12. *DurInt* (1- or 2-digit; '–' or '+' for open interval, '.' shows that duration-data are not available). The length of interval between the births.
13. *Vital* (1-digit; 1 = live births, 2 = total births); vitality of births.
14. *Births* (numeric, no fixed length); live births' counts.
15. *Access* (1-char). This field will indicate the confidentiality/accessibility of the data ('C' - confidential, 'O' - publicly accessible, 'U' – unknown).
16. *Note1* (numeric, no fixed length). Each of the three fields ("*Note1*", "*Note2*", "*Note3*") contains a code that links to specific notes in the file "XXXnote.txt". These fields may be empty (denoted by a single dot '.').
17. *Note2*
18. *Note3*
19. *RefCode* (numeric, no fixed length); reference to the source of data. Complete references to the sources are stored in the file XXXref.txt.
20. *LDB* (numeric, 0/1); coded '1' if the data are used to create the Lexis database and coded '0' if not.

Example of the input file for births

Below you can see an example of the input file for births. This is taken from the Austrian data file. Note that the headings of fields normally occupy only one row.

```
PopName,Area,Year,YearReg,Age,AgeInt,Cohort1,Cohort2,BirthOrder,OrderInt,
PrevBirth,DurInt,Vital,Births,Access,Note1,Note2,Note3,RefCode,LDB
AUT,1,1990,1990,12,1,1978,1978,1,1,.,.,1,1,O,3,.,.,.,9,1
AUT,1,1990,1990,12,1,1977,1977,1,1,.,.,1,1,O,3,.,.,.,9,1
AUT,1,1990,1990,13,1,1977,1977,1,1,.,.,1,0,O,3,.,.,.,9,1
AUT,1,1990,1990,13,1,1976,1976,1,1,.,.,1,2,O,3,.,.,.,9,1
AUT,1,1990,1990,14,1,1976,1976,1,1,.,.,1,5,O,3,.,.,.,9,1
AUT,1,1990,1990,14,1,1975,1975,1,1,.,.,1,11,O,3,.,.,.,9,1
AUT,1,1990,1990,15,1,1975,1975,1,1,.,.,1,25,O,3,.,.,.,9,1
AUT,1,1990,1990,15,1,1974,1974,1,1,.,.,1,45,O,3,.,.,.,9,1

…
AUT,1,1990,1990,54,1,1936,1936,1,1,.,.,1,0,O,3,.,.,.,9,1
AUT,1,1990,1990,54,1,1935,1935,1,1,.,.,1,0,O,3,.,.,.,9,1
AUT,1,1990,1990,55,1,1935,1935,1,1,.,.,1,0,O,3,.,.,.,9,1
AUT,1,1990,1990,55,1,1934,1934,1,1,.,.,1,0,O,3,.,.,.,9,1
AUT,1,1990,1990,TOT,.,.,.,1,1,.,.,1,41789,O,,,.,9,1
```

2.1.3 Adjustments to raw data

In the next step, the data on births are processed in a standardized adjustment procedure, which involves the following:
  – redistribution of births of unknown age,
  – redistribution of births of unknown birth order,
  – estimation of the proportion of live births if only information on total births exist, i.e. if the original data include still births,
  – moving births of late registration back to the year of their occurrence (some countries report births registered in current year, while they actually occurred in some earlier year),
  – splitting five-year age groups into one-year age groups,
  – splitting open age intervals into single age groups,
  – making ranges of age and birth order uniform (the standard age range is $12 \geq x \leq 50$ and the standard range of birth orders includes 1, 2, 3, 4, and 5+),
  – splitting bigger Lexis elements into Lexis triangles,
  – etc.

2.1.4 Lexis Database

The birth data that are originally received in other shapes than Lexis triangles are to be split into Lexis triangles in the HFD, i.e. they are sorted both by age and birth cohort of mother. The estimates of female exposure-to-risk (data on the female exposure by age taken from the HMD) are divided into Lexis triangles as well. The obtained numbers in both cases are rounded to 2 decimal places. In this way, we give the user a possibility to distinguish the "real" numbers (i.e., data originally received in Lexis triangles) from the estimates produced by the splitting procedure.

Thus, the Lexis database includes two 'Lexis files': 'Lexis file' for births and 'Lexis file' for exposure population. All further calculations are based on these two files. These two files are displayed as separate output files and are made accessible for the HFD users as well.

Example of the 'Lexis file" for births by year of occurrence, age and year of birth of mother and birth order of child

Year	Age	Cohort	Total	1	2	3	4	5p
1984	12	1972	0.00	0.00	0.00	0.00	0.00	0.00
1984	12	1971	0.00	0.00	0.00	0.00	0.00	0.00
1984	13	1971	2.00	2.00	0.00	0.00	0.00	0.00
1984	13	1970	1.00	1.00	0.00	0.00	0.00	0.00
...								
1984	25	1959	3623.00	1588.00	1484.00	440.00	90.00	21.00
1984	25	1958	3376.00	1470.00	1379.00	421.00	85.00	21.00
1984	26	1958	3283.00	1316.00	1376.00	467.00	104.00	20.00
1984	26	1957	3022.00	1160.00	1323.00	407.00	97.00	35.00
1984	27	1957	3033.00	1135.00	1249.00	505.00	115.00	29.00
1984	27	1956	2757.00	1029.00	1095.00	504.00	100.00	29.00
...								

Example of the 'Lexis file' for female exposure population

```
Austria, Exposure to risk, by Lexis triangles for the year 1984 (1x1)
Last modified: 15-Feb-2009
Year  Age  Cohort    Exposure
1984   12   1972      25024.25
1984   12   1971      26460.17
1984   13   1971      26552.58
1984   13   1970      27322.75
1984   14   1970      27431.25
1984   14   1969      29611.58
1984   15   1969      29717.92
1984   15   1968      30647.42
...
```

2.1.5 Output Database

The major blocks of output data that will become freely available on the website are as follows:
   – Birth counts by Lexis triangles,
   – Female exposure-to-risk by Lexis triangles,
   – Birth counts, fertility rates, and related summary indicators based on larger Lexis elements (Lexis squares, vertical and horizontal parallelograms),
   – Cohort fertility table (parity and age dimension) and table summary indicators,
   – Period fertility table (parity and age dimension) and table summary indicators.

Selected summary indicators, such as total fertility rates, mean ages at childbearing, and parity progression ratios will be displayed in separate files as well. It is envisioned that the HFD will also include the data on births and fertility rates by duration since the last previous birth (birth interval indicators) for the countries, for which this kind of data exist.

A complete list of output data files is provided in Appendix 2. Appendix 1 gives information on the notation.

Examples of some output data files

1) Birth counts. These are period data (Lexis squares), but files containing births by Lexis triangles and by horizontal parallelograms (cohort data) are organized in a very similar way.

```
Austria, Live births by age and birth order (period 1x1)
Last modified: 15-Feb-2009
Year  Age     Total         1         2         3         4        5p
1990   12      2.00      2.00      0.00      0.00      0.00      0.00
1990   13      2.00      2.00      0.00      0.00      0.00      0.00
1990   14     16.00     16.00      0.00      0.00      0.00      0.00
1990   15     74.00     70.00      4.00      0.00      0.00      0.00
1990   16    338.00    320.00     17.00      1.00      0.00      0.00
1990   17    851.00    784.00     62.00      5.00      0.00      0.00
1990   18   1599.00   1433.00    157.00      8.00      1.00      0.00
1990   19   2497.00   2149.00    321.00     24.00      3.00      0.00
1990   20   3722.00   2981.00    669.00     69.00      2.00      1.00
…
```

2) Unconditional fertility rates by age. Again, these are period data, but cohort unconditional fertility rates are analogously displayed.

```
Austria, period age-specific fertility rates for all births (ASFR) & for
specific birth orders (ASFRi)
Last modified: 15-Feb-2009
Year  Age     ASFR      ASFR1      ASFR2      ASFR3      ASFR4     ASFR5p
1990   12   0.00005    0.00005    0.00000    0.00000    0.00000    0.00000
1990   13   0.00005    0.00005    0.00000    0.00000    0.00000    0.00000
1990   14   0.00036    0.00036    0.00000    0.00000    0.00000    0.00000
1990   15   0.00160    0.00152    0.00009    0.00000    0.00000    0.00000
1990   16   0.00713    0.00675    0.00036    0.00002    0.00000    0.00000
1990   17   0.01712    0.01577    0.00125    0.00010    0.00000    0.00000
1990   18   0.03003    0.02692    0.00295    0.00015    0.00002    0.00000
1990   19   0.04472    0.03848    0.00575    0.00043    0.00005    0.00000
1990   20   0.06288    0.05036    0.01130    0.00117    0.00003    0.00002
...
```

3) Total fertility rates (all other files for summary indicators, including mean ages at childbearing and parity progression ratios, are formatted in the same way):

```
Austria, period total fertility rate (TFR) & order-specific total fertility
rates (TFRi)
Last modified: 15-Feb-2009
Year    TFR    TFR1    TFR2    TFR3    TFR4    TFR5p
1978   1.603     .       .       .       .       .
1979   1.598     .       .       .       .       .
1980   1.651     .       .       .       .       .
1981   1.673     .       .       .       .       .
1982   1.661     .       .       .       .       .
1983   1.559     .       .       .       .       .
1984   1.523   0.685   0.506   0.209   0.076   0.047
1985   1.474   0.667   0.495   0.201   0.070   0.042
1986   1.449   0.664   0.485   0.194   0.069   0.038
1987   1.432   0.663   0.480   0.186   0.064   0.038
...
```

2.1.6 Data checks

Rigorous and standardized data checking will be applied during the HFD data processing. A detailed data inspection will be performed once the raw data are received and later again when the output is obtained. The output data files are posted on the website only after a careful investigation.

2.2 Fertility tables specific for parity and age

The data that will be displayed in the HFD are very rich; it is beyond the capacity of one paper to discuss them in detail. By way of illustration, we focus therefore on one specific component — fertility tables and more precisely on period fertility tables, standardized for parity and age. Cohort and period fertility tables will constitute a very important contribution of the HFD in facilitating cross-country comparative analysis on changes in parity-specific fertility patterns and on the issue of fertility postponement and recuperation.

2.2.1 Cohort fertility table

Cohort fertility tables in the HFD comprise the following columns (functions): *Cohort*, x, $b_i(x)$, $l_{i-1}(x)$, $q_i(x)$, $m_i(x)$, $Sb_i(x)$, $chi(x)$.

Notation:

x	Age at childbearing
i	Birth order of child (1-5+); relates to parity i-1 (0-4+)
$b_i(x)$	Table number of births of order i in age interval $[x, x+1)$
$l_{i-1}(x)$	Table population of parity i-1 by age x
$q_i(x)$	Probability of having an ith birth in age interval $[x, x+1)$
$m_i(x)$	Conditional age-specific fertility rates (occurrence/exposure rates) in age interval $[x, x+1)$
$Sb_i(x)$	Cumulative (in respect to age) births of order i by age x
$chi(x)$	Average number of children by age x in the highest parity category i_{max+}

2.2.1 Period fertility table

Period fertility tables describe childbearing histories of synthetic cohorts of women that would live their whole life under the current fertility conditions, assuming the absence of mortality and migration. For studying fertility by parity, a multi-state life table method is applied, where states are defined by current parity. The basic parameters, from which all other estimates are derived, are age and parity specific probabilities and conditional occurrence-exposure rates. In the HFD, conditional fertility rates, $m_i(x)$, which are then converted into probabilities, $q_i(x)$, serve as the major input for the construction of period fertility tables (Formulae 1 and 2). This approach is favored because it allows at least partly to account for the effects of mortality and migration on population exposure $E_{i-1}(x,t)$ over the year (period) t.

$$m_i(x) = \frac{B_i(x)}{E_{i-1}(x)} \tag{1}$$

$$q_i(x) = \frac{m_i(x)}{1 + [1 - a_i(x)] \cdot m_i(x)} \tag{2}$$

where $a_i(x)$ is mean duration time spent at parity i-1 within age interval $[x, x+1)$ by women progressing to parity i during this age interval]. It is assumed that $a_i(x) = ½$.

Period fertility tables in the HFD comprise the following columns (functions): *Year*, x, $w_{i-1}(x)$, $m_i(x)$, $q_i(x)$, $Q_i(x)$, $l_{i-1}(x)$, $b_i(x)$, $L_{i-1}(x)$, $Sb_i(x)$.

Notation:

x	Age at childbearing
i	Birth order of child (1-5+); relates to parity i-1 (0-4+)
$w_{i-1}(x)$	Distribution of female population exposure by age and parity (population weights)
$m_i(x)$	Conditional age-specific fertility rates (occurrence/exposure rates) in age interval $[x, x+1)$
$q_i(x)$	Probability of having an ith birth in age interval $[x, x+1)$
$l_{i-1}(x)$	Table population of parity i-1 by age x
$b_i(x)$	Table number of births of order i in age interval $[x, x+1)$
$L_{i-1}(x)$	Table population exposure of women of parity i-1 within age interval $[x, x+1)$
$Sb_i(x)$	Cumulative (in respect to age) births of order i by age x
$Q_i(x)$	'Lifetime' probability of having an ith birth after age x for women at parity i-1 when reaching age x

Distribution of female population by age and parity. In order to compute conditional rates and to build the period fertility table, the distribution of female population by age and parity, $w_i(x)$, is required. Depending on the data availability, different approaches to obtain these estimates can be applied, and they can be listed in the following hierarchy:

- When reliable population register data provided in the form of annual time series are available, these data will be given the preference.
- The reconstruction of $w_i(x)$ from time series of period data by cumulating fertility of cohorts over their reproductive age span will be employed for the countries that have sufficiently long time series of order-specific period data on births.
- Population census data will be used to derive the initial age-parity distribution (for one starting year, 'base year') for the countries where no high-quality register data exist and where order-specific data on births are available for a short period only.
- Survey data will be used in exceptional cases only, when no other methods of deriving the parity distribution by age can be employed and when the surveys cover a large fraction (>1%) of the female population of reproductive age.

Summary indicators. The HFD will display a number of summary indicators derived on the basis of period fertility tables, including a summary index of period fertility controlling for age and parity (termed by Rallu and Toulemon (1994)[2] PATFR) and its parity components (PATFR$_i$), table period mean ages at all births (TMAB) and at births of specific orders (TMAB$_i$), and table period parity progression ratios (PPR$_{i-1,i}$).

Below we provide an example of the output file for period fertility tables.

Example of the output file for period fertility tables

```
Austria, period fertility tables of births of order 1 to 5p (1x1)
Last modified: 27-Feb-2009
Year   x    w0(x)    m1(x)    q1(x)    l0(x)  b1(x)   L0(x) Sb1(x)    w1(x)    ...
1991  12  1.00000  0.00000  0.00000   10000      0   10000      0  0.00000   ...
1991  13  0.99999  0.00005  0.00005   10000      0   10000      0  0.00001   ...
1991  14  0.99960  0.00047  0.00047   10000      5    9997      5  0.00040   ...
1991  15  0.99850  0.00210  0.00210    9995     21    9984     26  0.00149   ...
```

[2] Rallu, L. and L. Toulemon. 1994. "Period fertility measures. The construction of different indices and their application to France, 1946-89". *Population, An English Selection*, 6: 59-94.

```
1991  16  0.99417  0.00650  0.00648   9974   65   9942     91  0.00548  ...
1991  17  0.98062  0.01737  0.01722   9909  171   9824    261  0.01797  ...
1991  18  0.95776  0.03030  0.02985   9739  291   9593    552  0.03844  ...
1991  19  0.92325  0.04563  0.04461   9448  422   9237    974  0.06853  ...
1991  20  0.87585  0.05946  0.05775   9026  521   8766   1495  0.10697  ...
1991  21  0.81613  0.06798  0.06575   8505  559   8226   2054  0.15045  ...
1991  22  0.75164  0.07686  0.07402   7946  588   7652   2642  0.18929  ...
1991  23  0.68547  0.08682  0.08321   7358  612   7052   3254  0.22377  ...
1991  24  0.61009  0.10194  0.09700   6746  654   6418   3909  0.25689  ...
1991  25  0.53381  0.10904  0.10340   6091  630   5776   4539  0.28121  ...
...
```

An illustration of selected functions and indicators from parity- and age-specific period fertility tables, Austria

Figures 1, 2, and 3 provide a graphical illustration of selected functions and summary indicators from parity- and age-specific period fertility tables computed for Austria. Since order-specific data on births are available in Austria only since 1984, we could not apply the "cumulative approach" to estimate the relative age-parity distribution of women. Instead we used the 1991 population census data, and on the basis of the estimated initial age-parity distribution for 1991 combined with fertility rates by age and birth order in the following years, we computed relative age-parity distributions for all successive years.

Figure 1 depicts conventional summary indicators of fertility (total fertility rates and mean ages at childbearing) as well as summary indicators computed on the basis of the period fertility table.

Figure 2 provides 'life time' probabilities of ever having a first, second, and third child after a certain age and probabilities of having a first, second, and third child by age for 1991 and 2007 (the first and the last year of the period covered in the HFD).

Parity-progression ratios constitute another type of commonly used summary measures of fertility. They are usually computed for cohorts and interpreted as a (summary) probability for women with $i–1$ children to have an ith child during their reproductive lifespan. Parity- and age-specific period fertility tables enable us to compute these indicators for synthetic cohorts as well. In Austria, parity progression ratios have been remarkably stable throughout the period 1991-2007 (see Figure 3).

Figure 1: Indicators of period total fertility and period mean ages at childbearing based on the 'ordinary' computations and on the results from period fertility tables, total births and first births, Austria, 1980-2007
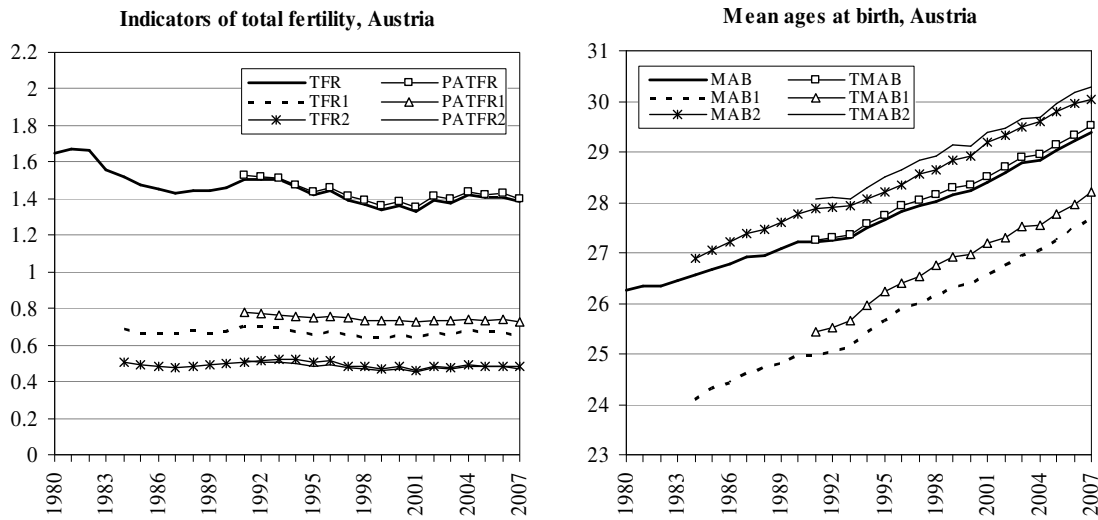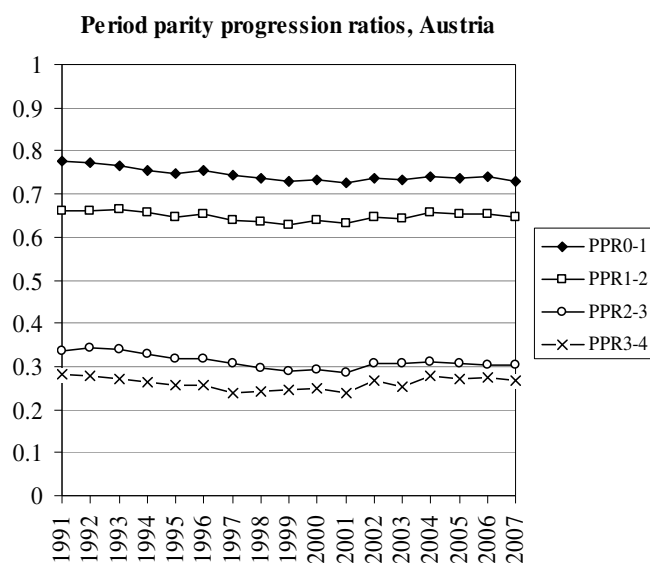

Source: HFD data for Austria.

Figure 2: 'Life time" probabilities of having an *i*th child after age *x* for women at parity *i*-1 when reaching age *x* ($Q_i(x)$) and probabilities of having an *i*th birth by age ($q_i(x)$)
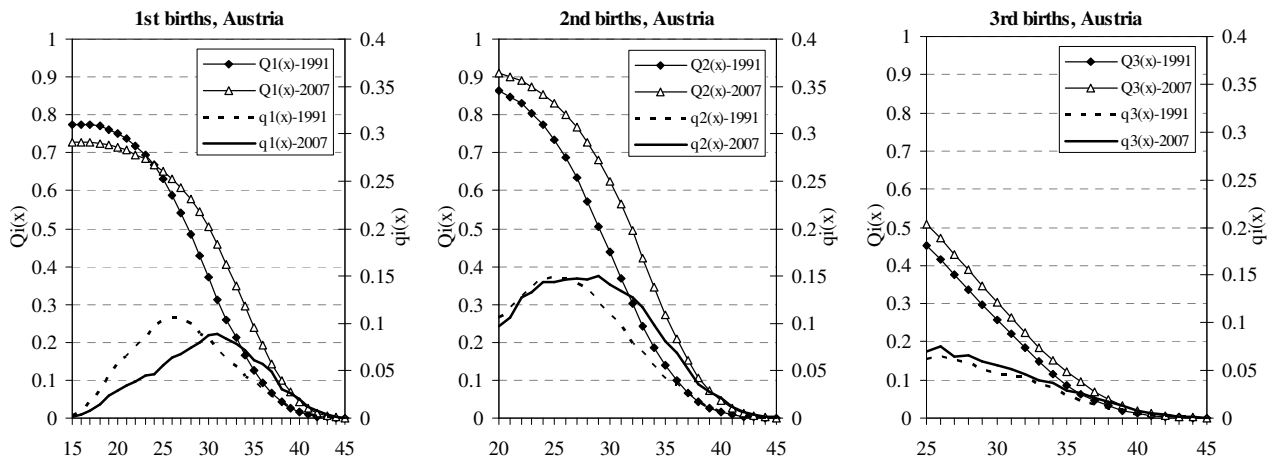
Source: HFD data for Austria.

Figure 3: Period parity progression ratios computed from parity- and age-specific fertility tables, Austria, 1991-2007

Period parity progression ratios, Austria

Source: HFD data for Austria.

Acknowledgements

Appendix

Appendix 1. Notations

	General
x	Age at childbearing
x_{min}	Lowest age at childbearing considered in the analysis
x_{max}	Highest age at childbearing
t	Calendar year
c	Cohort
i	Parity and birth order
i_{max}	Highest parity (birth order) used in the analysis

	Empirical data
$B_i(x,t,c)$	Number of births of order i (only live births are considered)
$P(x,t)$	Population size on January 1
$E(x,t,c)$	Population exposure:
	$E(x,t,*)$ – for rectangular
	$E(*,t,c)$ – for vertical parallelogram
	$E(x,*,c)$ – for horizontal parallelogram
$f_i(x,t,c)$	Unconditional age-specific fertility rates (of the 2nd kind):
	$f_i(x,t,*)$ – for rectangular
	$f_i(*,t,c)$ – for vertical parallelogram
	$f_i(x,*,c)$ – for horizontal parallelogram

	Fertility table
$w_{i-1}(x)$	Relative distribution of female population exposure by parity (population weights); $\sum_i w_i(x) = 1$
$w_{i-1}^{\tau}(x)$	Relative distribution of female population exposure by parity, initially based on census data
$m_i(x)$	Conditional age-specific fertility rates (of the 1st kind, occurrence/exposure rates) in age interval [x, x+1)
$q_i(x)$	Probability of having an ith birth in age interval [x, x+1)
$l_{i-1}(x)$	Table population of parity i-1 at age x
$b_i(x)$	Table number of births of order i in age interval [x, x+1)
$L_{i-1}(x)$	Table table population exposure of women of parity i-1 within age interval [x, x+1)
$chi(x)$	Average number of children by age x in the highest parity category i_{max+}
$Sb_i(x)$	Cumulative (in respect to age) births of order i by age x
$Q_i(x)$	'Lifetime' probability of having an ith birth after age x for women at parity i-1 when reaching age x
$S_{i-1}(x)$	Conditional survivorship: Lifetime probability of not having another child for women at parity i-1 when reaching age x
$a_{i-1}(x)$	Mean duration time spent at parity i-1 within age interval [x, x+1) by women progressing to parity i during this age interval]
	Summary measures
$PPR_{i-1,i}$	Parity progression ratio (lifetime probability of transition from parity i-1 to parity i); also commonly termed a_i
TFR_i	Total fertility rate for birth order i based on unconditional (non-exposure) age-

	specific fertility rates $f_i(x)$
TFR	Total fertility rate based on unconditional age-specific fertility rates $f_i(x)$
CTFR	Cohort total fertility rate
PATFR$_i$	Period fertility index for birth order *i* based on conditional age-specific fertility rates or birth probabilities, derived from the fertility table. Analogue to the period *TFR$_i$*
PATFR	Period fertility index of total fertility based on conditional age-specific fertility rates or birth probabilities; Rallu & Toulemon (1994) call it *the summary index of parity- and age-specific fertility*
TMAB$_i$	Table mean age at childbearing for birth order *i* based on the age distribution of the table number of births of order *i, b$_i$(x)*
MAB$_i$	Mean age at childbearing for birth order *i* based on unconditional age-specific fertility rates $f_i(x)$
MAB	Mean age at childbearing for all birth orders based on the age schedule of unconditional fertility rates $f_i(x)$

Appendix 2. List of output data files[3]

Type of data		Data files
Birth counts	1.	$B_i(x)$ by Lexis triangles
Female population exposure	2.	$E(x)$ by Lexis triangles

Unconditional fertility rates and their derivatives[4]

(1) Period fertility

Lexis squares:

	3.	$B(x)$, $B_i(x)$
	4.	$E(x)$
	5.	$f(x)$, $f_i(x)$
	6.	TFR, TFR_i
	7.	MAB, MAB_i

Vertical parallelograms:

	8.	$B(x)$, $B_i(x)$
	9.	$E(x)$
	10. .	$f(x)$, $f_i(x)$

(2) Cohort fertility

Horizontal parallelograms	11.	$B(x)$, $B_i(x)$
	12.	$f(x)$, $f_i(x)$
	13.	$E(x)$
	14.	TFR, TFR_i
	15.	*Cumulative TFRi by age*
	16.	*PPRi*
	17.	*MAB*, MAB_i, $MAB(40)$, $MAB_i(40)$ [by age 40]

Cohort and period fertility tables

Cohort fertility table:

(Horizontal parallelograms)	18.	*Table: Cohort*, x, $b_i(x)$, $l_{i-1}(x)$, $q_i(x)$, $m_i(x)$, $Sb_i(x)$, $chi(x)$

Period fertility table, $w_{i-1}(x)$:

(Lexis squares)	19.	*Table: Year*, x, $w_{i-1}(x)$, $m_i(x)$, $q_i(x)$, $l_{i-1}(x)$, $b_i(x)$, $L_{i-1}(x)$, $Sb_i(x)$
	20.	*PATFR, PATFR$_i$*
	21.	*TMAB, TMAB$_i$*
	22.	Table $PPR_{i-1,i}$
	23.	$m_i(x)$

---

[3] Consult Appendix 1 for the notation.

[4] This block of data include births and estimates of female population classified by three Lexis elements: Lexis squares, vertical parallelograms, and horizontal parallelograms. Unconditional age- and order-specific fertility rates and related summary indicators are respectively computed on the basis of data of each shape.