

# Methods Protocol for the Human Fertility Database

A. Jasilioniene, D. A. Jdanov, T. Sobotka, E. M. Andreev, K. Zeman, and  
V. M. Shkolnikov; with contributions from J. Goldstein, E. J. Nash, D. Philipov, and G. Rodriguez

<b>INTRODUCTION.....</b>	<b>3</b>
<b>1. GENERAL PRINCIPLES.....</b>	<b>4</b>
1.1 TIME, AGE, AND COHORT.....	4
1.2 BIRTH EVENT, BIRTH ORDER, AND PARITY OF A WOMAN .....	5
1.3 LEXIS DIAGRAM .....	5
1.4 STANDARD CONFIGURATIONS OF AGE, TIME, AND COHORT .....	8
1.5 PERIOD AND COHORT DIMENSIONS .....	9
1.6 ADJUSTMENTS TO RAW DATA.....	10
1.7 INPUT AND OUTPUT DATA FILES AND FORMATS .....	11
<b>2. DATA PROCESSING IN THE HFD .....</b>	<b>12</b>
2.1 DATA FLOWS AND DATA TRANSFORMATIONS.....	12
2.2 RAW DATA. DATA ARCHIVE .....	14
2.3 INPUT DATA.....	14
2.4 LEXIS DATA.....	15
2.5 OUTPUT DATA .....	15
<b>3. OVERVIEW OF THE COMPUTATIONS .....</b>	<b>17</b>
3.1 SEQUENCE OF COMPUTATIONS .....	17
3.2 LEXIS REGIONS FOR UNCONDITIONAL PERIOD AND COHORT FERTILITY RATES .....	20
3.3 LEXIS REGIONS FOR COHORT AND PERIOD FERTILITY TABLES .....	23
<b>4. COMMON ADJUSTMENTS TO INPUT DATA ON BIRTHS .....</b>	<b>26</b>
4.1 DISTRIBUTING BIRTHS WITH UNKNOWN PARAMETERS.....	27
4.2 SPLITTING BIRTHS CLASSIFIED BY ONE-YEAR AGE OR COHORT GROUPS INTO LEXIS TRIANGLES .....	28
4.3 SPLITTING AGGREGATED AGE GROUPS INTO ONE-YEAR AGE GROUPS .....	30
4.4 ITERATIVE PROPORTIONAL FITTING (IPF).....	32
<b>5. ADJUSTMENTS OF CENSUS PARITY DISTRIBUTIONS .....</b>	<b>33</b>
5.1 ESTIMATING THE AGE-PARITY DISTRIBUTIONS OF WOMEN FOR 1 JANUARY OF THE CENSUS YEAR.....	33
<b>6. UNCONDITIONAL FERTILITY RATES AND SUMMARY INDICATORS.....</b>	<b>36</b>
6.1 CALCULATION OF POPULATION EXPOSURE .....	36
6.2 UNCONDITIONAL AGE-SPECIFIC FERTILITY RATES .....	38
6.3 CUMULATIVE AND TOTAL FERTILITY RATES.....	39
6.4 TEMPO-ADJUSTED TOTAL FERTILITY RATE .....	40
6.5 MEAN AGES AT BIRTH .....	41
6.6 STANDARD DEVIATION IN THE MEAN AGE AT BIRTH.....	42
6.7 COHORT PARITY PROGRESSION RATIOS .....	43
6.8 CRUDE BIRTH RATES .....	44
<b>7. COHORT FERTILITY TABLES .....</b>	<b>44</b>
7.1 CONSTRUCTION OF THE COHORT TABLE .....	45
<b>8. PERIOD FERTILITY TABLES.....</b>	<b>46</b>
8.1 PARITY-SPECIFIC POPULATION EXPOSURE .....	47

8.1.1 *Cumulating cohort fertility rates over long periods of time* ..... 47

8.1.2 *Use of a “golden” census* ..... 48

8.1.3 *Direct use of census or register data* ..... 50

8.2 **CONDITIONAL AGE-SPECIFIC FERTILITY RATES**..... 50

8.3 **CONSTRUCTION OF THE PERIOD TABLE**..... 51

8.4 **SUMMARY INDICATORS BASED ON THE PERIOD FERTILITY TABLE** ..... 52

**ACKNOWLEDGMENTS** ..... **53**

**REFERENCES**..... **54**

## Introduction

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. The database is intended to help 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 the international research community and other interested users with free and user-friendly access to detailed and high-quality data on period and cohort fertility. Following the example of the Human Mortality Database (HMD, [www.mortality.org](http://www.mortality.org) and [www.humanmortality.de](http://www.humanmortality.de)), our guiding principles are comparability, flexibility, accessibility, and reproducibility. When complete, the database will contain original data for around 35-40 countries or areas, as well as input data used in computations.<sup>1</sup>

The HFD is based on one and the same type of initial data: the original, officially registered, live birth counts by calendar year, the age of the mother and/or the mother's year of birth (i.e., birth cohort), and, whenever possible, the biological birth order. These data, together with the exposure population estimates generated using data on population size and deaths (mostly) from the HMD, selected population censuses, and register data; are used for further computations with a uniform set of methods. More detailed information – for example, about sources of raw data<sup>2</sup>, specific adjustments to raw data, and comments about data quality – is provided separately in the documentation for each country, region, or population. The major HFD output (for all birth orders combined and, when available, by birth order) includes detailed age-specific fertility rates; total fertility rates; tempo-adjusted total fertility rates; mean ages at birth; the standard deviation in the mean age at birth; cohort and period fertility tables; as well as selected aggregate indicators from these fertility tables.

This document first describes the general principles that are used in constructing and presenting the database. Next, it outlines the data processing within the HFD and provides an overview of the

---

<sup>1</sup> The Human Fertility Database works with the official data provided or published by national statistical agencies. The HFD methodology is based on the assumption that the input data are restricted to countries and years for which data cover the entire population, and for which the registration of demographic events is complete or nearly complete. We have not established precise criteria for the inclusion, since we are still learning about the statistical systems of many countries. We plan that the HFD will cover almost all countries included in the HMD and several other countries with high-quality data. In order to provide fertility data for countries and periods for which the birth data are incomplete or do not come from the official sources, or do not correspond to the HFD data quality standards, we are also assembling a collection of data compiled by other organizations or individuals—the Human Fertility Collection (HFC).

<sup>2</sup> The term “raw data” refers in this document to the original data files obtained from individual countries (birth counts, population exposure data, parity distribution data from population registers and censuses) that were not yet modified or standardised for the purposes of the HFD.

steps followed for converting raw data into fertility rates and fertility tables. The remaining sections contain detailed descriptions of all of the calculations that are performed in the HFD.

## 1. General principles

The notation and principles that are used in this document and in the HFD are mostly consistent with the HMD and with conventional demographic literature.

### 1.1 Time, age, and cohort

Time, age, and birth cohort can be considered as either discrete or continuous variables. Time  $t$  is measured in calendar years. When saying that a demographic event (e.g., birth) occurs in calendar year  $t$  (or simply in year  $t$ ), we mean that the event takes place at some exact time during the time interval  $[t, t+1)$ <sup>3</sup>. At any given time, the length of time that has passed since the individual's birth is the individual's age. An individual "of age  $x$ " (or "aged  $x$ ") has the exact age within the interval  $[x, x+1)$ . Age defined in this way is simply termed "age", "age at last birthday", or "age in completed years" (ACY).

It should always be possible to distinguish between discrete and continuous notions of time and age by usage and context. For example, a population aged  $x$  at time  $t$  refers to all individuals, with their ages being in the interval  $[x, x+1)$  at the exact time  $t$ ; for instance, on 1 January of year  $t$ . The population exposure-to-risk at age  $x$  in year  $t$  refers to the total number of person-years lived in the age interval  $[x, x+1)$  in year  $t$ .

The age of each individual increases during a calendar year. A woman aged  $x-1$  at the beginning of the year is aged  $x-1$  between the beginning of the year and her birthday, and is aged  $x$  between her birthday and the end of the year. Thus, age  $x$  is the age reached during the year (ARDY). In this document, simple references to "age  $x$ " or to "age in completed years " $x$  (ACY)" always mean the age at the last birthday. Age  $x$  reached during the year is denoted " $x$  (ARDY)".

Birth cohort  $c$  (or simply cohort  $c$ ) is a set of individuals born during a calendar year  $c$ . Individuals aged  $x$  in year  $t$  belong to two consecutive cohorts,  $t-x$  and  $t-x-1$ . Individuals reaching age  $x$  during year  $t$  (ARDY) all belong to one cohort,  $t-x$ .

---

<sup>3</sup> Interval  $[a,b)$ . The brackets designate the interval containing values  $k$ , such that  $a \leq k < b$ . A squared bracket,  $[$ , denotes a closed interval that includes its limit point (in this case  $a$ ), whereas a round bracket denotes an open interval that excludes its endpoint.

## 1.2 Birth event, birth order, and parity of a woman

Birth is a repeatable demographic event since women can give birth to one, two, or more children. Unlike death or out-migration, birth does not cause the withdrawal of mothers from the population under study; i.e., women of reproductive age. Therefore, the collection of retrospective information about the total number of live births ever born to women from different cohorts or age categories provides particularly useful information. The number of children ever born alive to a woman is referred to as parity, and is expressed in integer values beginning from zero.<sup>4</sup>

Births can also be distinguished by the order (e.g., first, second, third, etc.) in which they occur over an individual woman's life span. Contemporary fertility analyses have mainly focused on the true (biological) birth order, counting all live-born children born to a woman, irrespective of her past or current marital status.<sup>5</sup> Order-specific births are non-repeatable events. A woman of parity 0 can give birth to a first child only once. A woman of parity 1 can give birth to a second child only once, etc. With respect to parity-specific populations, order-specific births are exclusive demographic events: births of order  $i$  to women of parity  $i-1$  lead to their transition to the next parity category  $i$ .<sup>6</sup>

In most countries, statistical systems report the numbers of live births that do not include stillbirths. The majority of countries also collect data on births by biological (or true) birth order. In some countries, however, information is still collected only about birth order within a marriage or within a current marriage.

In the HFD, when order-specific birth counts are available, the birth order is specified for categories 1 to 5+. Correspondingly, the parity of a woman is distinguished from 0 to 4+.

## 1.3 Lexis diagram

The Lexis diagram is a device for depicting the stock and flow of a population and the occurrence of demographic events over age and time. The Lexis diagram is useful for describing both the format of raw data and various computational procedures. Figure 1.1 shows a small section of a Lexis diagram that has been divided into 1x1 square cells (i.e., one year of age by one year of time).

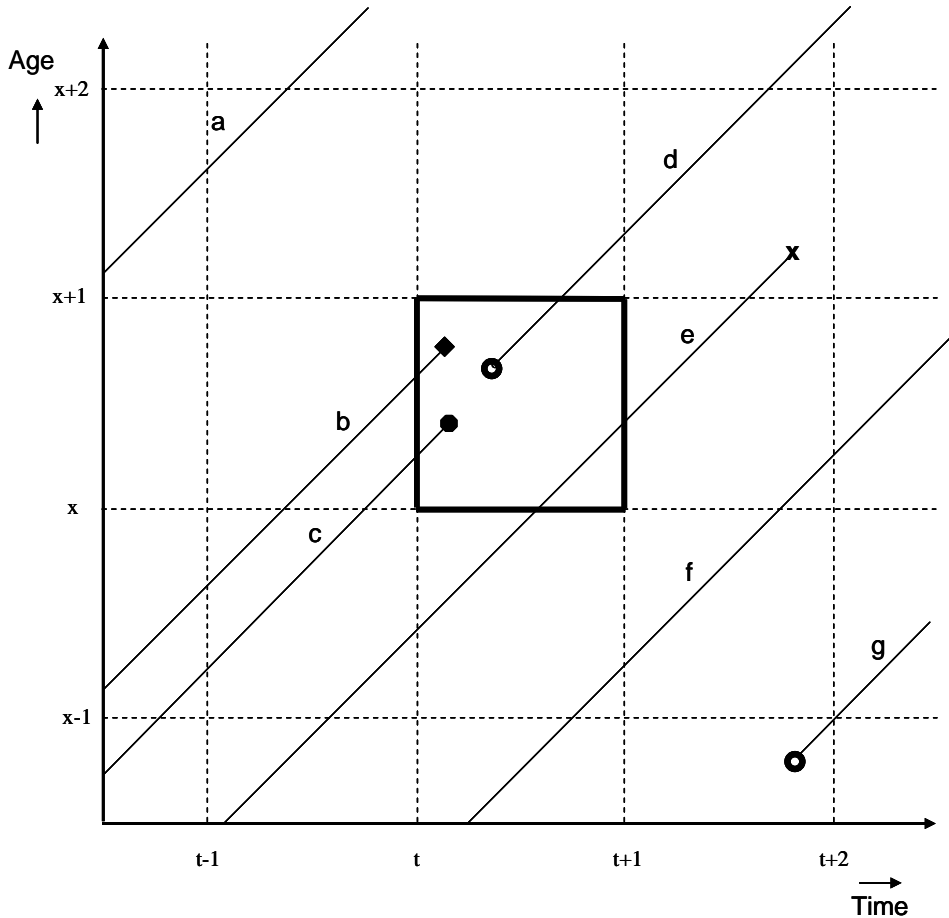
---

<sup>4</sup> Following conventional practice, the term "birth order" refers to the birth order of the child, whereas the term "parity" refers to the number of children a woman has at the time of observation.

<sup>5</sup> In the past, when marriage was assumed to be the only "legitimate" status for childbearing, marital birth order was a cornerstone of fertility analysis and for birth-order statistics provided by many statistical agencies. The de-coupling of marriage and childbearing in recent decades, with many countries registering 50% or more births outside marriage, renders data on marital birth order useless for reconstructing women's fertility histories. Therefore, the HFD collects only data by biological birth order, although many statistical agencies still publish data by birth order within marriage.

<sup>6</sup> Multiple births potentially complicate the definition of the parity of a woman and the birth order of children. Conventionally, most statistical agencies count children born in multiple deliveries as children with separate birth orders. For instance, a woman at parity 1 who gives birth to twins makes an instant transition to parity 3, and her twin births are assigned birth orders 2 and 3 (usually, only live-born children are counted).

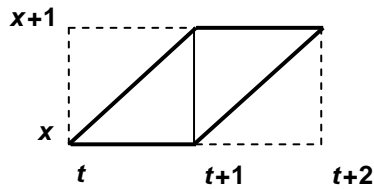
Each 45-degree line represents individual life trajectories. These trajectories may end in death, which is denoted by a cross (x, line e); in out-migration, which is denoted by a square (line b); or in a move to a higher parity by giving birth, which is denoted by a solid circle (line c). An individual may also migrate into the population, which is denoted by an open circle (lines d and g). Other life-lines indicate that individuals pass through the section of the Lexis diagram under consideration without experiencing any event (lines a and f).



**Figure 1.1. Example of a Lexis diagram with life trajectories and demographic events**

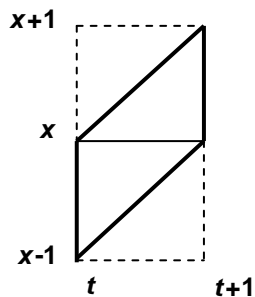
There are four conventional data configurations (ways to classify data) as they appear on the Lexis diagram: *horizontal parallelogram*, *vertical parallelogram*, *square (or rectangle)*, and *triangle* (see Caselli and Vallin (2006) for a more detailed explanation).

The *horizontal parallelogram* (F1) groups together events occurring during two consecutive years  $t$  and  $t + 1$  within one birth cohort  $t - x$  at age  $x$ . In this Lexis shape, events are classified by age in completed years (ACY) and year of birth (birth cohort).



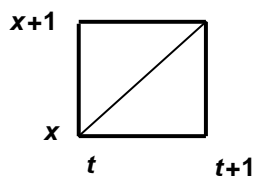
F1: *Horizontal parallelogram (extending over two calendar years on the t-axis)*

The *vertical parallelogram* (F2) groups together events occurring during the year  $t$  to members of the cohort  $t - x$  at ages  $x - 1$  and  $x$ . In this Lexis shape, events are defined by calendar year and year of birth (cohort) or age reached during the year (ARDY).



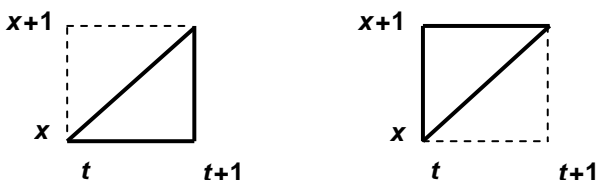
F2: *Vertical parallelogram (extending over two years of age on the x-axis)*

The *square or rectangle* (F3) groups together events occurring during the year  $t$  to members of two birth cohorts  $t - x$  and  $t - x - 1$ , who have the same age  $x$ . In this Lexis shape, events are classified by age in completed years (ACY) and calendar year.



F3: *Square (or rectangle)*

*Lexis triangles* are the basic Lexis elements that allow for the reconstruction of any of the three Lexis shapes described above. In Lexis triangles, events are classified by all of the three possible dimensions: calendar year, age, and year of birth (birth cohort). There are two types of triangles on the Lexis diagram: the lower Lexis triangle and the upper Lexis triangle. The *lower Lexis triangle* (F4) contains events such as births or deaths that take place in the year  $t$  to the birth cohort  $t - x$  at age  $x$ . The *upper Lexis triangle* (F5) contains events that occur in the year  $t$  to the birth cohort  $t - x - 1$  at age  $x$ .



F4: *Lower Lexis triangle*      F5: *Upper Lexis triangle*

If births are classified by Lexis triangles, the births in squares and in the two types of parallelogram shapes can obviously be obtained by summing up the corresponding triangles.

### 1.4 Standard configurations of age, time, and cohort

On each HFD country page, detailed original data on births, population exposures, fertility rates, and fertility tables are presented in three data blocks “Summary Indicators”, “Age-Specific Data”, and “Fertility Tables”. These data are classified by single calendar years, single years of age, and single-year birth cohorts. The age ranges from age 12 or younger ( $\leq 12$ ) to age 55 and older (55+).

The birth data are presented by four types of arrays, with data configurations corresponding to the four Lexis shapes: Lexis triangles, Lexis squares, horizontal parallelograms, and vertical parallelograms (see Section 3). Data by Lexis triangles are aimed at advanced users, making it possible for them to compute fertility rates and other fertility indicators in any configuration desired.

Period age-specific fertility rates are computed for every single age from  $\leq 12$  to 55+. Summary indicators of period fertility are computed over the whole range of age-specific rates (using data by Lexis squares). The total fertility rates, the mean ages at birth, and the standard deviation in the mean age at birth are also provided by age 40 to facilitate comparisons with the cohort data for women who are close to completing their reproductive period. Period fertility tables specified by age and parity are based on the age-parity distribution of the female population obtained by two approaches: i.e., either solely by reconstructing the fertility of cohorts over their reproductive age spans, or by combining the latter approach with the use of data from a “golden” census<sup>7</sup>. The first type of period fertility tables and the resulting indicators are computed when the parity distribution covers ages from  $\leq 12$  to 44 or higher; the highest age category is determined by data availability and can range from 44 to 55+. The age range

<sup>7</sup> The standard method for obtaining the age-parity distribution of women, which is necessary for the computation of period fertility tables, is the reconstruction of lifetime fertility of cohorts from the time series of fertility rates by age and birth order. In some cases, however, especially when the age- and order-specific birth data are available for a short period only, the age-parity distribution from a population census or register can be used to build left-censored cohort fertility histories. This method makes it possible to extend the time series of data on the period age-parity distribution of women, and thus of the period fertility tables. The population census or register used for this purpose is called the “golden” census. The use of the golden census is described in more detail in Section 3.3 (see Figure 3.5) and Section 8.1.2.



$\leq 12$  to 55+ is displayed when parity distribution data are available for women of all reproductive ages. Period fertility tables built using a “golden” census,<sup>8</sup> as well as census- or register-based fertility tables, are also calculated over this entire range of ages (see Section 3).

For the cohort fertility measures, the age ranges vary depending on the indicator and the length of observation. Cohort age-specific fertility rates (horizontal Lexis parallelograms) are computed for all of the observed single ages for which data are available. For the youngest and the oldest cohorts included, only one open-ended age is observed: age  $\leq 12$  for the youngest cohort and age 55+ for the oldest cohort. More ages are observed for all of the intermediate cohorts. If the duration of the observation period is sufficiently long, the fertility rates for many of the cohorts cover their entire reproductive age range.

Cohort cumulative fertility rates are computed from age 15 or younger<sup>9</sup> and until the last observed age, up to age 55+. Completed cohort fertility and mean ages at birth are computed over ages ranging from 15 or younger through 50 or older. Because many of the cohorts approaching the end of their reproductive span have practically completed their fertility histories, the HFD also displays the completed cohort fertility and the mean ages at childbearing achieved by age 40; these indicators should be useful to researchers in estimating or projecting the completed fertility rates of those cohorts. Cohort life tables are built for ages ranging from 15 or younger to at least 25 or older (up to a maximum of 55+), depending on the age of the cohort in the latest year for which birth data are available in the HFD.

On each country page, the original input data (raw data organised in a uniform format) used in the calculations are shown in the data block named “Input Data”.

### **1.5 Period and cohort dimensions**

In contemporary industrialised countries, birth is mostly a result of rational choice. Women and couples often plan future childbearing or postpone births until conditions are more favourable (Kohler et al. 2002, Sobotka 2004b, Frejka and Sardon 2006). Such changes in the timing of period fertility may strongly affect period fertility indicators, which consequently become poor predictors of final

---

<sup>8</sup> The standard method for obtaining the age-parity distribution of women, which is necessary for the computation of period fertility tables, is the reconstruction of lifetime fertility of cohorts from the time series of fertility rates by age and birth order. In some cases, however, especially when the age- and order-specific birth data are available for a short period only, the age-parity distribution from a population census or register can be used to build left-censored cohort fertility histories. This method makes it possible to extend the time series of data on the period age-parity distribution of women, and thus of the period fertility tables. The population census or register used for this purpose is called the golden census. The use of the golden census is described in more detail in Section 3.3 (see Figure 3.5) and Section 8.1.2.

<sup>9</sup> Depending on data availability, the starting age is between 12 and 15. Births below age 12 are extremely rare and are grouped together with births at age 12.

family size among particular cohorts of women. The fact that the period fertility rates that are usually used are frequently “distorted” by shifts in fertility timing (“tempo effects”) underlines the need to include in the HFD a broad array of cohort fertility indicators as well as tempo-adjusted total fertility rates.

It is important to bear in mind that the cohort data and indicators of fertility displayed in the HFD, including the completed cohort fertility and parity distributions, are based on statistical models that may not fully correspond with the fertility behaviour of real birth cohorts. In particular, the method of reconstructing cohort fertility history by cumulating the fertility rates of given cohorts over long periods of time is based on a rather strong assumption that migration and mortality are not selective with respect to fertility; i.e., that those who die or migrate have at any given reproductive age the same parity distribution and completed fertility as those who survive and stay in a country. This assumption is relatively harmless in the case of mortality, which is very low at reproductive ages for women in most of the developed countries; but it is more problematic in the case of migration. Many developed countries have experienced an influx of immigrants in recent decades, and female migrants tend to have different parity distributions and fertility behaviour than native-born women, which violates the statistical assumption that migration has no effect on fertility. In addition, the permanent or temporary out-migration of younger women—e.g., for work-related and family-related reasons—which has become quite common across Central and Eastern Europe, may violate the model assumptions behind the estimation of cohort fertility rates. Therefore, the cohort data in the HFD should be understood as statistical approximations of “real” cohort behaviour computed for a set of people who lived their lives according to the observed cohort rates in a country, which are themselves subject to effects from in- and out-migration. Consequently, the HFD may occasionally feature implausible estimates of cohort parity distribution, and especially of childlessness, for some cohorts in countries with intensive migration. When possible, the country documentation file and specific notes will warn the user about such data.

### ***1.6 Adjustments to raw data***

The HFD is designed to provide opportunities for conducting comparative studies on fertility in different countries and time periods. Thus, consistency across the whole data universe is an important priority. The desire for uniformity is hindered by significant variability in data formats and a lack of sufficient detail in the raw data. The raw data on births are often classified only by calendar year and the age of the mother (either of the two categories of age, age in completed years or age reached during the year, may be shown) or by calendar year and the birth cohort of the mother. For some countries and

calendar years, the birth data are available by five-year age intervals only. They may show broader or narrower ranges of available ages, they may include births with an unknown age of the mother or an 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  $\leq 12$  to 55+, by single-year birth cohorts, and (when possible) by birth orders varying from 1 to 5+. Births with an unknown age of the mother are distributed proportionally according to the birth data in which the age of the mother is specified. Within each age, births with an unknown birth order are distributed proportionally across the known birth orders. Aggregated age groups are additionally split into single-year ages by means of spline interpolation. Birth orders higher than five are combined into birth order 5+. Within each age, births are additionally split by the year of birth of the mother (if such information is not present in the input data).

It is important to understand that data classified in the most detailed way and provided on the HFD country pages are in many cases obtained from less detailed raw data. Before being displayed on the HFD country pages, the raw data are adjusted and are additionally split into finer cells by methods specified in Section 4. Although there are some obvious advantages in maintaining a uniform format in the presentation of fertility rates and fertility tables, it is important not to interpret these somewhat artificial data literally. Users should also be aware that the HFD displays estimated birth counts with a precision of two decimal places. In all cases, the user must take responsibility for understanding the sources and limitations of all data provided in the HFD, which are documented in detail in the country Background and Documentation files and in specific explanatory notes in the input data files.

### ***1.7 Input and output data files and formats***

On each HFD country page, output data resulting from calculations using the HFD methodology are placed in three data blocks labelled “Summary Indicators”, “Age-Specific Data”, and “Fertility Tables”. The output data are given in easily readable ASCII formats. More detailed information about specific parts of the output data can be found in the next section of this document, and also in the [Explanatory Notes](#).

Input data used for the computation are provided in the fourth block, “Input Data”. Users should note that, although the input data are presented in standardised format with the same set of dimensions for each country, there is significant inter-country variability in data formats with respect to available age ranges, open-ended age intervals, Lexis shapes, specification of data for unknown ages and birth orders, etc. (for more detail, see the next section of this document and the [Explanatory Notes](#) and [Data](#)

Formats documents). The input data are provided in a comma-delimited format—where each column is separated by a comma—readable in Excel and statistical packages.

The input data files can also be valuable resources for more advanced HFD users, as these files often contain additional information that is not further processed in the standard HFD computations, and is therefore also not displayed in the output database. The additional information provided in the input data can, for example, cover periods not shown in the output database due to the absence or low quality of population exposure data, show births by individual birth orders beyond the highest birth order considered in the HFD output (e.g., fifth and higher), or show live births by age categories that could not be processed using standard HFD computations. Particularly useful are census- and register-based data on the parity distribution of the female population by age and cohort that allow users to compute a range of alternative cohort fertility indicators and parity distributions for historical cohorts that are often not included in the output files. Nevertheless, researchers using HFD input data should always bear in mind that these data come in a variety of non-standardised formats, and that the HFD assumes no responsibility for their format, quality, or documentation; especially for the segments not further used in the computations for the output database. In many cases, the data included in the input data files but not used for further calculations have some problems, and should be used with caution. Users are strongly advised to read the country’s Background and Documentation file before using such data.

## 2. Data processing in the HFD

### 2.1 Data flows and data transformations

Figure 2.1 illustrates the stages of data processing in the HFD, as well as the major types of data involved. The raw data on births, as well as all of the relevant documents and exchanges, are stored in their original formats in the Data Archive. These data are then transformed into a standard format and placed in the Input Database. For each country, this database includes the following data files: live births by month; live births by the age of the mother and/or the mother’s year of birth and (when available) birth order; data files used for the estimation of population exposures by age<sup>10</sup>; and (when available) the age and parity distributions of women from censuses, population registers, or large-scale surveys. These files are posted in the data block “Input Data” on the HFD country pages.

---

<sup>10</sup> Exposure population in the HFD is usually estimated using data on population size and deaths from the Human Mortality Database, available at <http://www.mortality.org> or <http://www.humanmortality.de>. For countries not included in the HMD, data on population size and deaths come from other sources; they are usually provided by statistical offices.

The input data files are further adjusted and refined to obtain for each country annual estimates of births and population exposures by Lexis triangles (calendar year-age-cohort configuration). For countries and years for which census- or register-based data on the distribution of female population by age and parity are available, age- and parity-specific population weights are derived. The whole universe of these detailed estimates across all available years and countries constitutes the Lexis Database.

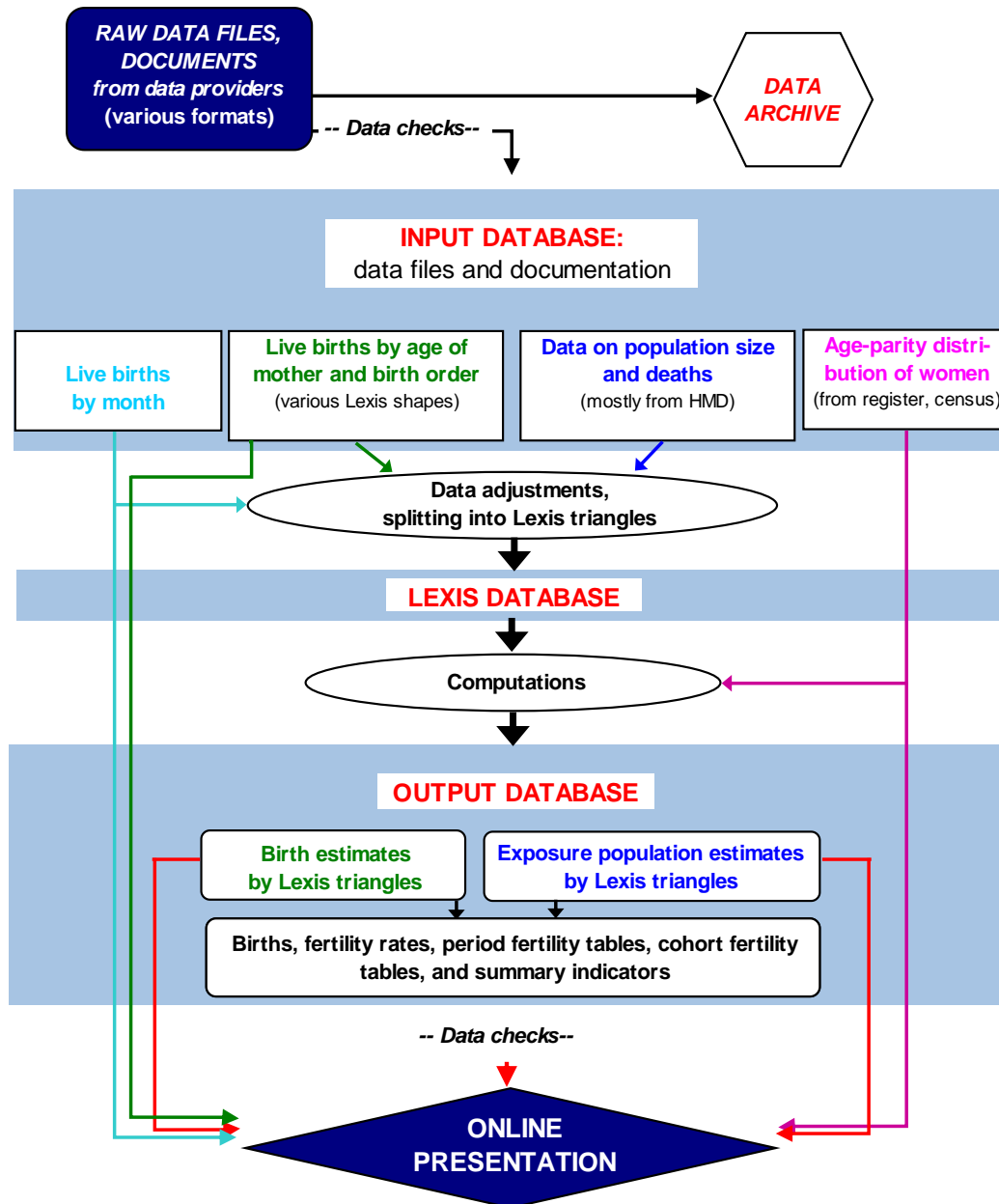


Figure 2.1. Scheme of the HFD data processing

The Lexis Database already provides some parts of the HFD Output; namely, the birth and population exposure estimates by Lexis triangles. From these data, births by Lexis squares and by vertical and horizontal parallelograms are computed. Other calculated outputs include unconditional period and cohort age-specific fertility rates, cohort and period cumulative fertility rates, total fertility rates, tempo-adjusted total fertility rates, mean ages at birth, standard deviation in mean ages at birth. If birth order-specific fertility data are available, cohort parity progression ratios and cohort fertility tables are computed. The Output Database also includes period fertility tables and census- or register-based period fertility tables.

Finally, for each country, all of the files of the Output Database are displayed on the HFD website.

## **2.2 Raw data. Data archive**

The Data Archive, which is not accessible for HFD users, stores raw data files and all of the accompanying documents supplied by the data providers (i.e., by the country experts, statistical offices, and research institutions). A large part of the data is obtained or purchased directly from statistical offices. Raw data arrive in a variety of formats (Excel or text data files, scanned or paper copies of official yearbooks, and other statistical publications).

Our standard data request to a country data provider is for time series that are as long as possible of the following types of data:

- Live births by calendar year, the age of the mother, and the mother’s year of birth.
- Live births by calendar year, the age of the mother, the mother’s year of birth, and birth order.
- Distribution of women by the age and/or the year of birth, and the parity (number of biological live-born children). This information usually comes from population registers, population censuses, or large-scale surveys representative of the entire population.
- Live births by calendar year and month.
- Female and male population by age as of 1 January, death counts by age and sex at the most detailed level available, and live birth counts by sex. These data are requested only from countries that are not present in the Human Mortality Database.

## **2.3 Input data**

All input data used for the computation of output fertility indicators are provided on the country pages. These data include the raw data originally collected for each country, which have been

converted into a standardised format (described in the [Data Formats](#)). For each country included in the HFD, there are the following files, which contain data on:

- Births by calendar year, the age of the mother, (when available) the mother’s year of birth, and (if available) birth order.
- The age- and parity-specific distribution of the female population (when available: from population censuses, population registers, or large-scale surveys).
- Births by calendar year and month.
- Female and male population sizes and death counts by age and sex, and birth counts by sex (for countries and periods not available in the HMD).

In addition to the data files, the Input Database includes two documents in PDF format. The first one contains *Notes* regarding specific points or parts of the input data. The second one contains *References* to sources of the input data.

## **2.4 Lexis data**

The HFD methodology makes it possible to turn the input data, which vary in many respects across countries and time, into uniform Lexis data. In these data, births and population exposures are classified by Lexis triangles (by calendar year-age-birth cohort cells). Categories of age, birth order, and parity are predefined as follows. Age varies from 12 and younger to 55+. The birth order of children varies from one to 5+. The parity of mothers varies from zero to 4+. Births are classified by Lexis triangles. The age-parity distributions of women are given by one-year age groups as of 1 January.

The Lexis data are of paramount importance, since they form the basis for all further calculations.

## **2.5 Output data**

This section lists the fertility indicators that are provided in the three output data blocks: “Summary Indicators”, “Age-Specific Data”, and “Fertility Tables”.

The first block, “Summary Indicators”, consists of two sub-blocks: “Period summary indicators” and “Cohort summary indicators”.

Period summary indicators include the following data:

- Total number of live births.
- Crude birth rates.

- Total fertility rates (including total fertility rates by age 40).
- Tempo-adjusted total fertility rates, using Bongaarts’ and Feeney’s method.
- Mean ages at birth (including mean ages at birth by age 40).
- Standard deviation in the mean age at birth (including the standard deviation in the mean age at birth by age 40).

Cohort summary indicators, included in the second sub-block, are as follows:

- Completed cohort fertility (including the completed cohort fertility by age 40).
- Parity progression ratios.
- Mean ages at birth (including mean ages at birth by age 40).
- Standard deviation in the mean age at birth (including the standard deviation in the mean age at birth by age 40).

The summary indicators of period and cohort fertility are, where appropriate, calculated on the basis of fertility rates by age; and, when available, by birth order. In the case of period summary indicators, these rates are sorted by Lexis squares, and for producing cohort summary indicators, horizontal parallelograms are used.

The second block, “Age-Specific Data”, is also made up of two sub-blocks, separating period and cohort data. Both sub-blocks cover the same array of indicators: birth counts, population exposures, and rates. Period data are provided by Lexis triangles, squares, and vertical parallelograms; while cohort data are organised by horizontal parallelograms.

The fertility indicators provided in the block “Age-Specific Data” are as follows:

- Live birth counts for all birth orders combined and (when available) by birth order by all Lexis shapes (Lexis triangles, squares, vertical and horizontal parallelograms).
- Female population exposures by all types of Lexis shapes<sup>11</sup> (Lexis triangles, squares, vertical and horizontal parallelograms).
- Unconditional age-specific fertility rates for all birth orders combined and (when possible) by birth order (by Lexis triangles, squares, vertical and horizontal parallelograms).
- Cumulative fertility rates (by Lexis squares, vertical and horizontal parallelograms).

The third block, “Fertility Tables”, contains two sub-blocks: “Period fertility tables” and “Cohort fertility tables”.

---

<sup>11</sup> We do not provide here female population exposures by parity. The estimates of exposures by parity are presented in the data block “Fertility tables” because of their nature.



The sub-block “Period fertility tables” consists of period fertility tables as well census- or register-based fertility tables, specific for age and parity, and provides the following data:

- Fertility tables.
- Female population exposures by age and parity.
- Conditional period fertility rates, controlling for age and parity. These data are extracted from period fertility tables and/or census- or register-based fertility tables featured in this sub-block.
- Period table summary indicators, which include parity- and age-adjusted total fertility rates (PATFR)<sup>12</sup> and table mean ages at birth.

The cohort fertility tables are displayed in the sub-block “Cohort fertility tables”.

The period fertility tables are built on the basis of data by Lexis squares, while the construction of the cohort fertility tables involves data by horizontal parallelograms. Period data are indexed by calendar year, whereas cohort data are indexed by year of birth.

The detailed description of the computational procedures is provided in the following sections.

### 3. Overview of the computations

This section is a summary of the HFD methodology for obtaining the output data from the Lexis data on births and population exposures and from census or population register data on the parity distributions of women. Detailed descriptions of the computations are provided in the following sections of this document. First, we provide a brief description of *how* various calculations are made. Then we show examples of Lexis diagrams for locating birth and population data used in different kinds of calculations in the year-age-cohort coordinates.

#### 3.1 Sequence of computations

The HFD process for computing output indicators of fertility from input data on births and population can be briefly described as a sequence of six steps:

1. **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 the year of birth (cohort) of the mother, and by the 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

---

<sup>12</sup> Rallu and Toulemon (1994) termed it “the summary index of period fertility controlling for age and parity”; the acronym *PATFR* is based on the original name in French.

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 a uniformity of the 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, cannot be obtained, and are therefore not featured in the HFD.

2. **Population denominators.** In the HFD, data on the female population as well as on the total population (i.e., men and women together) are used. There are two types of data for the female population. The first type specifies population exposure by age, which is needed for the computation of unconditional fertility rates. For most countries, the female population exposure is estimated using data on population size and deaths from the HMD. For countries that are not included in the HMD, these types of data are collected together with the data on births. The second type of population data contain counts of women by age and parity, which are needed for the computation of conditional fertility rates, and 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. Data on total population are used only for the computation of crude birth rates.
3. **Fertility rates.** Fertility rates are ratios of birth counts to corresponding population exposures. Unconditional age-specific fertility rates relate births specified by the age of the mother and the 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 ages and parities (e.g., second births are related to women of parity one only). Cumulative fertility rates are based on a summation of fertility rates up to the indicated age limit, shown for each single age category. Cohort fertility rates are computed for every combination of cohort and age observed in the available data. Period fertility rates are computed for every combination of calendar year and age observed in the available data. Furthermore, the HFD provides 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.

4. **Summary measures.** The summary measures are crude birth rates, total fertility rates (including the completed cohort fertility)<sup>13</sup>, tempo-adjusted total fertility rates, mean ages at birth, cohort parity progression ratios, and the standard deviation in the mean age at birth. The simplest summary measure is the crude birth rate, which is the ratio of total live births to total population in a given year. It is expressed as the number of live births per 1,000 of the population. The other above-listed summary measures are based on fertility rates by age and (when possible) by birth order. The period total fertility rate (TFR) is computed as a sum of age-specific fertility rates for a certain calendar year across all ages from  $\leq 12$  to 55+. The completed cohort fertility (CCF) is computed as a sum of age-specific fertility rates in a certain cohort. The HFD also displays the TFR and the CCF by age 40, based on a summation of age-specific fertility rates over all ages under 40. The TFR 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. It is a hypothetical (synthetic) indicator, which can also be specified by birth order. The HFD also provides the period TFR adjusted for tempo effects using the Bongaarts and Feeney (1998) method. The completed cohort fertility shows the average number of children (or children of a specific birth order) born to women belonging to certain cohort over their whole reproductive lives. The cohort parity progression ratio expresses the probability of giving birth of birth order  $i+1$ , conditional on reaching parity  $i$ . The mean age at birth and the mean age at birth by age 40 are computed from the vector of period and cohort 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. The period and the cohort standard deviation in the mean age at birth is computed on the basis of age-specific fertility rates and mean ages at birth; this measure shows the extent of variability from the period or the cohort mean age at birth, computed from the entire schedule of age-specific rates and over the range of ages under age 40. All of the summary measures are computed for all of the birth orders combined and for specific birth orders.
5. **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. For each cohort, the life table functions are computed from

---

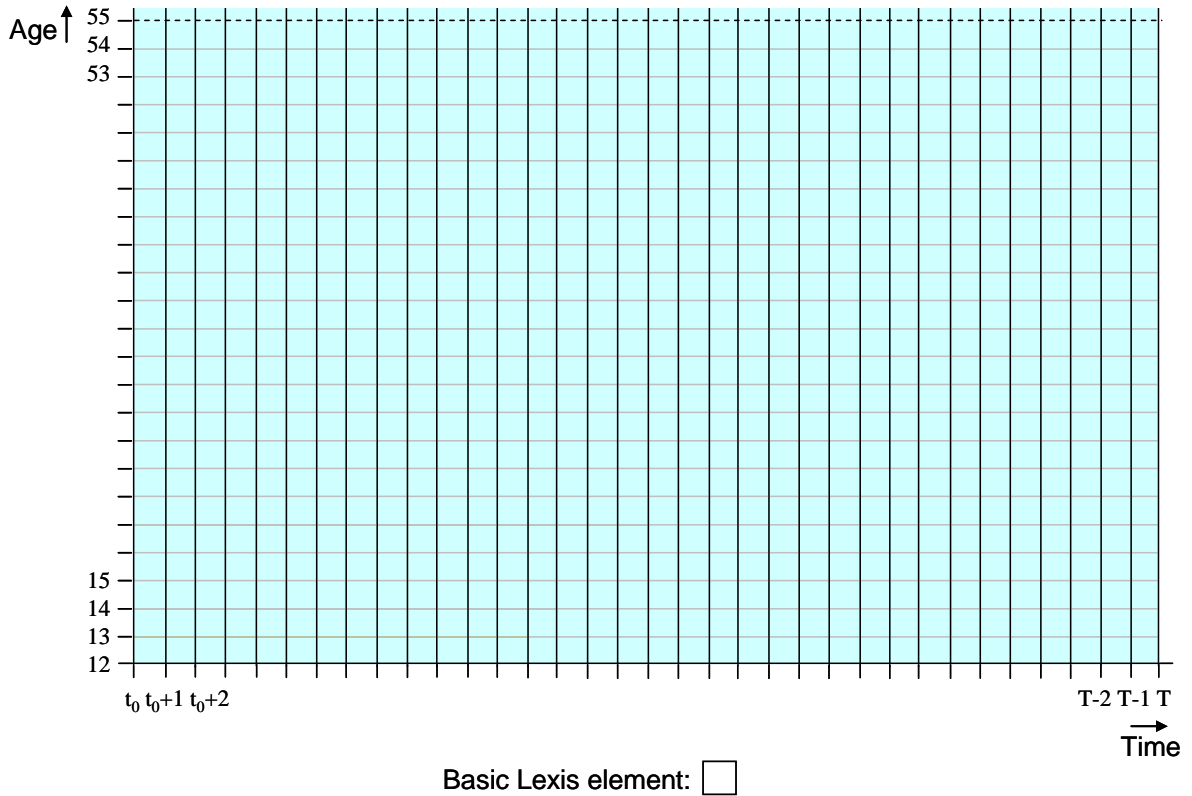
<sup>13</sup> When we refer to both period and cohort dimensions, the term "total fertility rate" denotes both the period total fertility rate and the completed cohort fertility. When cohort fertility is discussed separately, only the "completed cohort fertility" is used.

the array of age- and parity-specific fertility rates as major input data. The distribution of births by the age of the mother and the birth order in the fertility table and the parity distribution of the table population of females correspond to the observed fertility trajectories of the cohorts analysed.

6. **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 the 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 and from “golden” censuses that provide the initial 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. One of the main outputs of the period fertility table is the parity- and age-adjusted total fertility rate (PATFR) and its order-specific components.

### ***3.2 Lexis regions for unconditional period and cohort fertility rates***

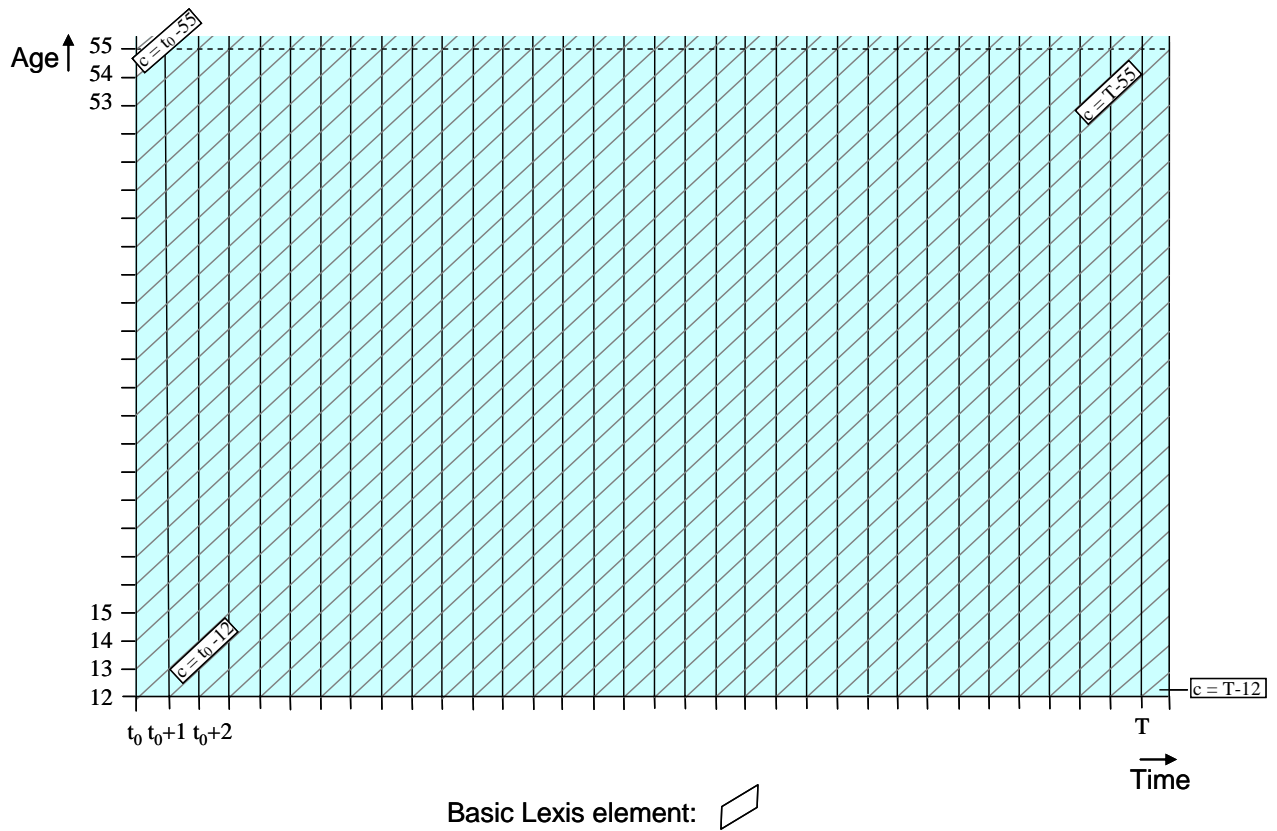
Let us assume that, for a given country, the Lexis database is already estimated and the calendar year-, age-, and cohort-specific births and population exposures are available. Let us then consider a rectangular region of such data, which lasts from 1 January of the year  $t_0$  to 1 January of the year  $T+1$  (horizontal axis), and which extends over the standard range of ages from  $\leq 12$  to  $55+$  (vertical axis). With this array, the *unconditional period age-specific fertility rates* are being computed for all Lexis squares (age-year cells) within the entire region covering all observed years from  $t_0$  to  $T$  and all observed ages from  $\leq 12$  to  $55+$  (coloured area in Figure 3.1). As indicated earlier, births and population exposures in each Lexis square can be obtained by summing up the two elementary triangles building up the square.



**Figure 3.1. Lexis region for unconditional period fertility rates by Lexis squares (year-age cells), cumulative period fertility rates, period total fertility rates, and period mean ages at birth**

From the age-specific fertility rates, the summary indicators of fertility, such as total fertility rates and mean ages at birth, are computed for every year from  $t_0$  to  $T$ .

For a fertility analysis that looks simultaneously at period and cohort patterns of fertility, it is useful to have *fertility rates by vertical parallelograms* (by year-birth cohort cells). Using the same range of initial data on births and population exposures, these rates are being computed within the region, covering horizontally the years from  $t_0$  to  $T$ ; and covering vertically the ages (ARDY) across the entire range, from  $\leq 12$  to  $55+$  (Figure 3.2). While in the first year  $t_0$ , fertility rates cover cohorts from  $t_0-55$  to  $t_0-12$ ; in the last year  $T$ , the cohorts from  $T-55$  to  $T-12$  are covered.

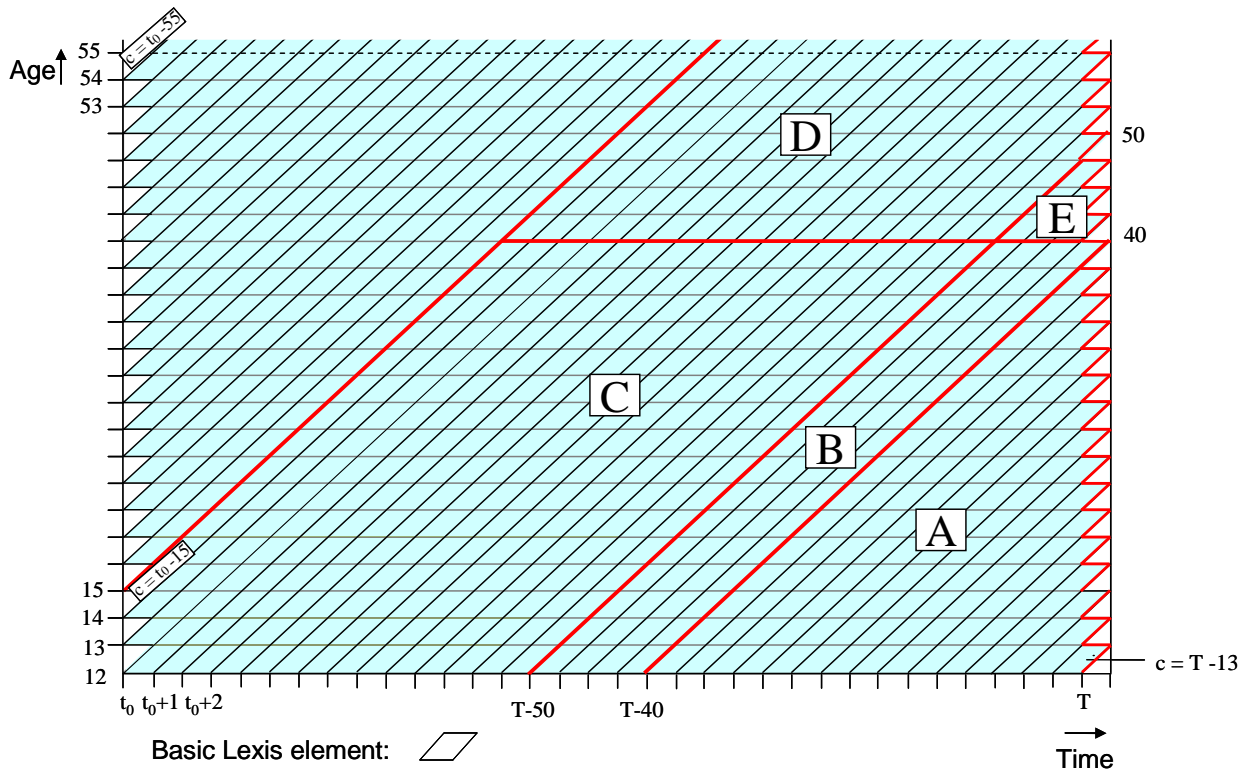


**Figure 3.2. Lexis region for unconditional fertility rates by vertical parallelograms (year-cohort cells)**

With the same region of Lexis data on births and population exposures, the *cohort fertility rates by horizontal parallelograms* (cohort-age cells) are computed for each cohort for which at least one data cell is available. Figure 3.3 clearly shows that the oldest observable cohort is the one born in the year  $t_0-55$ , and that the youngest cohort is the one born in the year  $T-13$ .

The *cohort cumulative fertility rates* are computed for cohorts that are observed from age 15 or younger. The fulfillment of this condition corresponds to the Lexis region  $A+B+C+D+E$  in Figure 3.3. The *completed cohort fertility*, the *cohort mean ages at birth*, and the *standard deviation in the cohort mean age at birth* are computed only for the cohorts who are observed from age 15 or younger until age 50 or an older age, and are considered as the cohort that have completed their childbearing. In Figure 3.3, corresponding cohorts belong to the Lexis regions  $C+D$ , respectively. The number of cohorts for which the above summary indicators can be computed is equal to  $(T-50)-(t_0-15)+1$ . This

obviously implies that a cohort analysis of completed or nearly completed childbearing requires long observation periods. However, the completed cohort fertility, the cohort mean ages at birth, as well as the standard deviation in the cohort mean age at birth, are additionally calculated over the range of reproductive ages under age 40 (see the Lexis regions B+C in Figure 3.3).

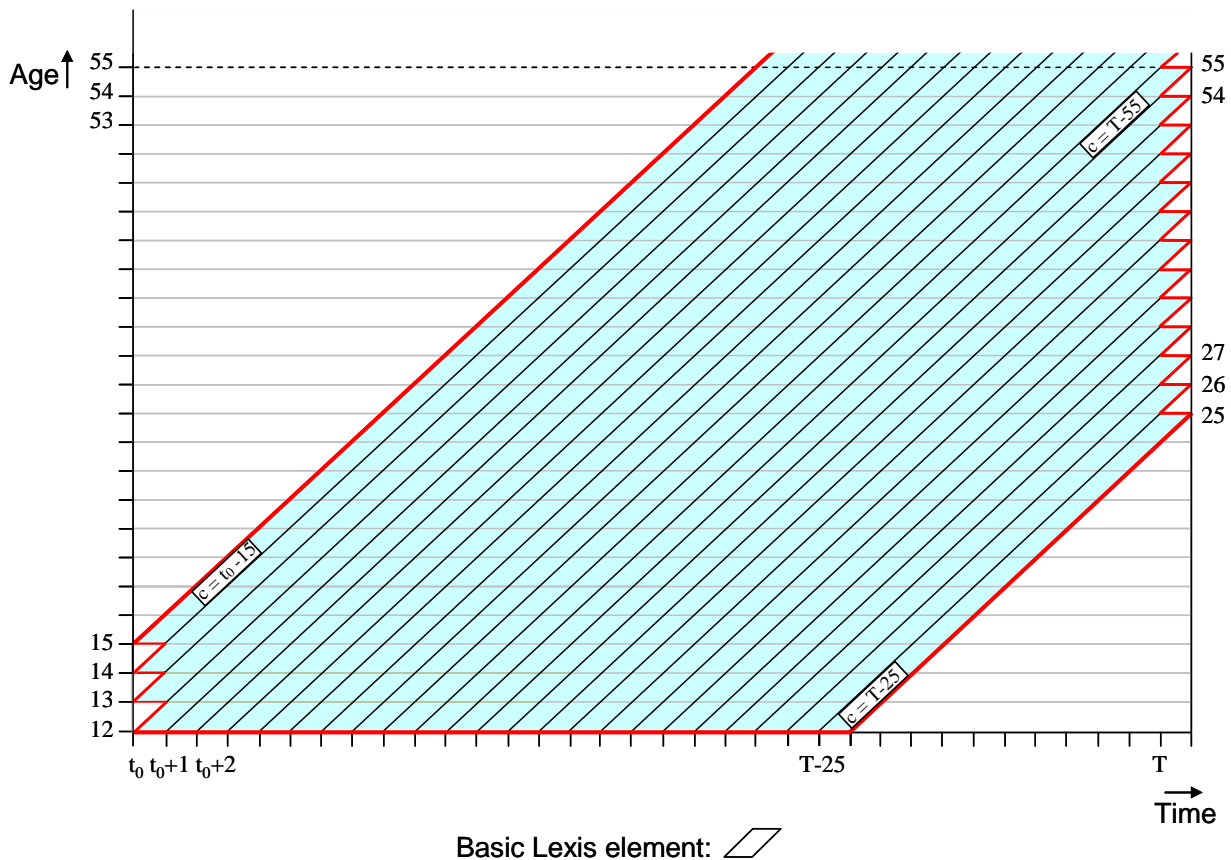


**Figure 3.3. Lexis regions for unconditional cohort fertility rates by horizontal parallelograms (cohort-age cells), cumulative cohort fertility rates, completed cohort fertility, and cohort mean ages at birth**

### 3.3 Lexis regions for cohort and period fertility tables

The cohort fertility tables are based on the cohort unconditional fertility rates by horizontal parallelograms (cohort-age cells). These tables are constructed for cohorts who are observed from age 15 or a younger age until age 25 or an older age. Figure 3.4 shows that the cohorts satisfying both conditions belong to the coloured diagonal region<sup>14</sup>. The youngest and the oldest cohorts for whom the cohort fertility tables can be constructed are  $T-25$  and  $t_0-15$ , respectively.

<sup>14</sup> It should be noted that only complete horizontal parallelograms are used in the computation of the cohort fertility tables.

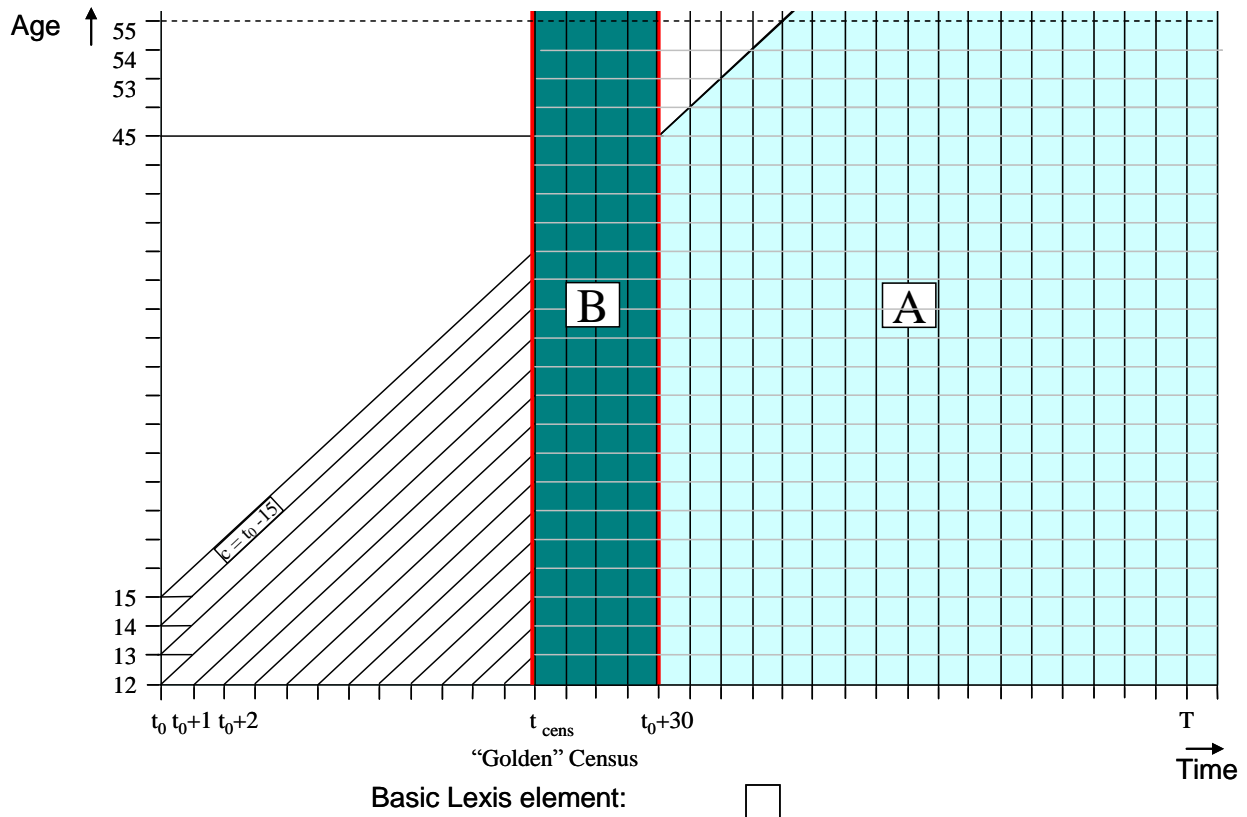


**Figure 3.4. Lexis region for the cohort fertility tables based on horizontal parallelograms (cohort-age cells).**

The period fertility tables are based on the conditional age and parity-specific fertility rates by Lexis squares (year-age cells). In order to compute these rates, female parity-specific population exposures must be estimated (see Sections 3.1 and 5).

We assume that age 45 is an age by which the fertility of cohorts is nearly completed and women’s parities are very close to their final values. Therefore, period fertility tables are constructed in the HFD for all of the years when cohort parity distribution can be observed for ages 15 or younger through 45 or older. In Figure 3.5, the first observed cohort  $t_0-15$  reaches age 45 in the year  $t_0+30$ . Beginning from this year, it is possible to compute the period fertility tables using parity distributions of women obtained by cumulating the cohort fertility as the population denominator. Accordingly, these tables are being computed for every year from  $t_0+30$  to  $T$  (region A in Figure 3.5).



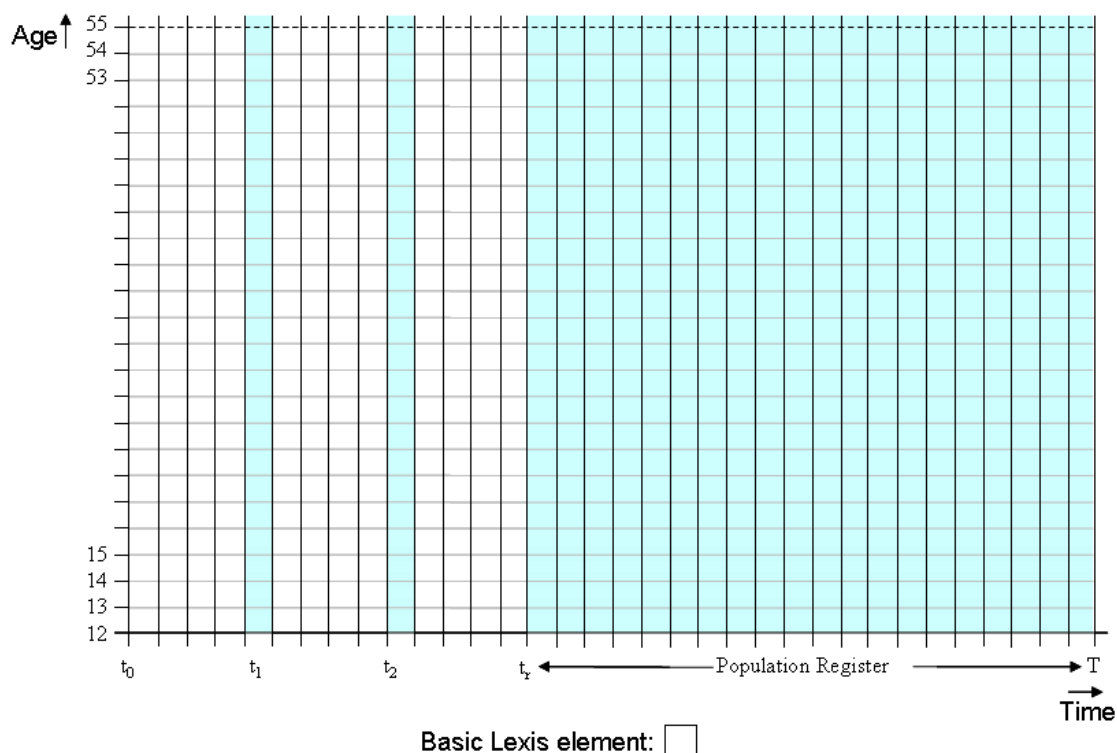


**Figure 3.5. Lexis regions for the period fertility tables based on Lexis squares (year-age cells).**

In some cases, the region for the computation of the period fertility tables can be extended by using additional information taken from a population census. If the census takes place in the year  $t_{\text{cens}}$ , then the period fertility tables can additionally be constructed for a region B in Figure 3.5. In order to obtain a continuous series of census-based parity distributions covering the whole region B, the “baseline” census parity distribution in the year  $t_{\text{cens}}$  – also called the golden census – is combined with age-specific fertility rates by birth order computed for all cohorts of reproductive age for  $t_{\text{cens}}$  and for the subsequent years. Thus, even when available, the parity distribution data from the censuses following the golden census are not used for updating the annual age-parity distributions of the female population. The series are reconstructed annually until the year when parity-specific exposures of the respective cohorts can be obtained for all reproductive ages until 45 or older purely by cumulating annual series of cohort fertility rates by birth order (region A in Figure 3.5). Therefore, in Figure 3.5 there is a break in the time series in the year  $t_0+30$ , from which the period fertility tables are entirely based on parity- and age-specific population exposures constructed by cumulating fertility rates over long periods of time.

The distributions of female population by age and parity provided by population censuses, population registers or large representative surveys are also used directly for the construction of period

fertility tables. Let us imagine a hypothetical country with two censuses in years  $t_1$  and  $t_2$ , and a population register functioning from the year  $t_r$  on (Figure 3.6).



**Figure 3.6. Lexis regions for the census- or register-based period fertility tables based on Lexis squares (year-age cells).**

In such a country, parity- and age-specific population exposures are available for the entire range of reproductive ages in the two census years and during the continuous time period lasting from the year  $t_r$  to the year  $T$ . These are the years for which the census- or register-based period fertility tables are being constructed.

#### 4. Common adjustments to input data on births

This section provides a description of the methods employed to adjust the data used for the production of the Lexis data classified by calendar years and single years of the mother's age and of the mother's year of birth.

Three common adjustments to the raw data are performed: 1) redistributing births with an unknown age of the mother or an unknown birth order, 2) splitting birth counts classified by Lexis squares into Lexis triangles, and 3) splitting birth counts classified by aggregated age groups into single years of age. All of these adjustments are being applied separately to each birth order (including all birth orders combined).

At the end of this section, we provide a description of the iterative proportional fitting (IPF) algorithm, which is used to ensure that, within each age group, the sum by birth orders is equal to a fixed total; and that, within each birth order, the sum by ages is equal to a fixed total. For example, if data are initially aggregated by five-year age groups, and are then split into single-year ages for each birth order separately and independently from other birth orders, IPF makes it possible to modify the age distributions of order-specific data to fit the age-specific totals of all of the birth orders. It should be noted that these calculations typically result in non-integer birth numbers for individual ages and Lexis triangles.

All of the methods for splitting births are based on the piece-wise cubic Hermite interpolation<sup>15</sup> of the cumulative fertility rate. The method for the splitting of aggregated age groups is a modification of the algorithm proposed by Dimiter Philipov and German Rodriguez. For the sake of simplicity, all of the aggregated age groups are assumed to be based on rectangular Lexis shapes. Nevertheless, all formulae can be easily re-written for vertical parallelograms.

The HFD data user should be aware that we do not introduce any adjustments in the birth data to correct possible errors due to age heaping.

#### **4.1 Distributing births with unknown parameters**

The first adjustment to raw data deals with distributing births for which the mother's age and/or birth order is unknown into specific age categories. These births are to be distributed proportionally across the age/birth order ranges. This adjustment is based on an assumption that the probability of the mother's age not being reported is independent of the mother's age.

Let us suppose that birth counts are available for individual Lexis triangles of the Lexis diagram, but that the age of the mother is unknown for a certain number of births. Let

$B(x, t, t - x)$  be the number of lower-triangle births recorded among those aged  $[x, x + 1)$  in year  $t$  ;

$B(x, t, t - x - 1)$  be the number of upper-triangle births recorded among those aged  $[x, x + 1)$  in year  $t$  ;

$B^{UNK}(t)$  be the number of births with an unknown age of the mother in year  $t$  ; and

$B^{TOT}(t)$  be the total number of births in year  $t$  that is equal to

$$\sum_x [B(x, t, t - x) + B(x, t, t - x - 1)] + B^{UNK}(t) .$$

---

<sup>15</sup> We used the R function `interp1` with option "pchip" from the package "signal". A detailed description of the Hermite interpolation can be found, for example, in Fritsch and Carlson (1980).

Then, the following set of equations redistributes births with an unknown age of the mother proportionally across the upper and the lower Lexis triangles over the full range of ages:

$$B^*(x, t, t-x) = B(x, t, t-x) \cdot \left( \frac{B^{TOT}(t)}{B^{TOT}(t) - B^{UNK}(t)} \right) \quad (4.1)$$

and

$$B^*(x, t, t-x-1) = B(x, t, t-x-1) \cdot \left( \frac{B^{TOT}(t)}{B^{TOT}(t) - B^{UNK}(t)} \right) \quad (4.2)$$

for all ages  $x$  in year  $t$ .

In Sections 4.2 to 4.4, all of the formulae are based on an assumption that births with an unknown age of the mother have already been distributed proportionally, if needed; the superscript “\*” used in this section is suppressed for the sake of simplicity.

If for some births the birth order was unknown, such births are to be distributed proportionally within the corresponding age group. For some births both the age of the mother and the birth order could be unknown. In such cases, we begin from the proportional distribution of births with an unknown birth order within the respective age groups, followed by the distribution of births with an unknown age across the age groups according to (4.1) and (4.2). After the redistribution of births with an unknown age, the IPF procedure should be applied in order to achieve the initial balance between totals (with respect to age) for specific birth orders and the known total for all birth orders combined.

It should be noted that these manipulations are being performed before the raw birth counts are split into single-year ages and/or into Lexis triangles. However, the final result would not change if aggregated data were first split into finer age categories before the redistribution of an unknown age.

#### **4.2 Splitting births classified by one-year age or cohort groups into Lexis triangles**

Birth counts are often available only split by the single age of the mother, without the additional split by the mother’s year of birth. This implies that the birth data are given in Lexis squares and should be further split into Lexis triangles before further calculations can be made. Following McNeil *et al.* (1977) and Wilmoth *et al.* (2005), the HFD splitting procedure is based on the interpolation of a cumulative fertility indicator, the cumulative fertility rate. Data for each birth order are treated separately and independently from other birth orders. The following formulae assume that the calendar year and birth order are fixed.

Let  $F(x)$  be the cumulative fertility over all ages up to age  $x$  calculated using known age-specific births  $B(x)$  and population exposures  $E(x)$ :

$$F(x) = \sum_{u=x_{\min}}^{x-1} f(u) = \sum_{u=x_{\min}}^{x-1} \frac{B(u)}{E(u)} \quad (4.3)$$

Function  $F(x)$  is known for each age  $x \in [x_{\min}, x_{\max}]$ . If  $\hat{F}(x)$ , the continuous approximation of the discrete  $F(x)$  is also known, then the fertility rate for the upper Lexis triangle can be expressed as

$$f_U(x) = 2 \int_x^{x+1} (s-x) d\hat{F}(s) \quad (4.4)$$

Formula (4.4) uses the fact that the area of an elementary triangle is equal to 1/2. Similarly, for the lower triangle

$$f_L(x) = 2 \int_x^{x+1} [1-(s-x)] d\hat{F}(s) . \quad (4.5)$$

The continuous cumulative birth rate  $\hat{F}(x)$  is estimated using an interpolation technique. First, we apply a logit transformation

$$Y(x) = \text{logit} \left( \frac{F(x)}{F(x_{\max})} \right) = \log \left( \frac{F(x)}{F(x_{\max}) - F(x)} \right) \quad (4.6)$$

At this stage, two data points are lost because the logarithm at the extremes is not defined. In order to retain the information, we set  $Y(x_{\min}) = -20$  and  $Y(x_{\max}) = 12$ , which is equivalent to assuming that the ratio  $P(x) = F(x)/F(x_{\max})$  differs from zero or one at the extreme ages by less than  $10^{-6}$  (corresponding to less than one birth in a population comparable in size to that of the USA). Using piecewise cubic Hermite interpolation, we get a continuous approximation  $\hat{Y}(x)$  of  $Y(x)$ . The continuous cumulative fertility function  $\hat{F}(x)$  is obtained via an inverse logit transformation of  $\hat{Y}(x)$ :

$$\hat{F}(x) = \left[ \frac{\exp(\hat{Y}(x))}{1 + \exp(\hat{Y}(x))} \right] F(x_{\max}) \quad (4.7)$$

Thus, all of the components for a calculation of fertility rates by triangles in formulae (4.4) and (4.5) are defined. The numbers of births  $\hat{B}_L(x)$  and  $\hat{B}_U(x)$  in the lower and the upper Lexis triangles are calculated as follows:

$$\hat{B}_L(x) = f_L(x)E_L(x) \quad (4.8)$$

$$\hat{B}_U(x) = f_U(x)E_U(x) = 2[f(x) - 0.5f_L(x)]E_U(x) \quad (4.9)$$

where  $E_L(x)$  and  $E_U(x)$  are exposures to risk in the respective Lexis triangles, and  $f_L(x)$  and  $f_U(x)$  are defined in (4.4) and (4.5). The mid-point numerical integration with one thousand intervals is used for the computation of integrals in (4.4) and (4.5).

It is clear that the sum of the estimated birth counts  $\widehat{B}_L(x)$  and  $\widehat{B}_U(x)$  is (in general) not exactly equal to the original birth count  $B(x)$ . Thus, as a final step, we make the following adjustment:

$$\begin{aligned} B_U(x) &= \widehat{B}_U(x) \frac{B(x)}{\widehat{B}_U(x) + \widehat{B}_L(x)} \\ B_L(x) &= \widehat{B}_L(x) \frac{B(x)}{\widehat{B}_U(x) + \widehat{B}_L(x)} \end{aligned} \tag{4.10}$$

This procedure returns birth data split by Lexis triangles for each birth order and (independently) for all birth orders combined. To reach the balance with respect to birth orders, the IPF is applied to data for all births orders combined used as the marginal age-specific totals over all birth orders.

### **4.3 Splitting aggregated age groups into one-year age groups**

Occasionally, the original birth counts are available in broader groups than single years of age. Usually, it is necessary to deal with five-year age groups as well as with open-age intervals. For example, the age scale in a given year can include the following age groups:

<15, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and 50+ .

In this example, there are seven five-year age groups and two open-ended age groups. Data may be available in different age groupings; for example, the last closed age group may be 40-49. Or, for some segments of the reproductive age span, single-year data may be available as well, while for other segments the age groups may be broader.

Whenever at least one of the age groups of birth counts is wider than one year, it is necessary to apply a procedure that disaggregates the data into single-year age groups.

The splitting algorithm described below is very similar to the algorithm described in Section 4.2.

Let us assume there are no births at ages 11 and younger, at ages 55 and older, or at ages 60 and older. Then, open-ended age intervals are treated as age groups with the fixed borders of 12 and 55 (or 12 and 60). Thus, rather than using the original scale, it is possible to work with the following open-ended age groups:

12-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, and 50-54.

In a more formal way, let us assume there are  $n$  age groups  $[x_i, x_{i+1}), i = 0, \dots, n$ , covering the age range  $[x_{\min}, x_{\max})$ . Then, the cumulative fertility schedule is calculated as follows:

$$F(x) = \sum_{i=0}^n (x_{i+1} - x_i) f(x_i; x_{i+1}) = \sum_{i=0}^n (x_{i+1} - x_i) \frac{B(x_i; x_{i+1})}{E(x_i; x_{i+1})} \quad (4.11)$$

where  $f(x_i; x_{i+1})$  is the mean unconditional fertility rate over the age interval  $[x_i, x_{i+1})$ ,  $B(x_i; x_{i+1})$  is the number of births, and  $E(x_i; x_{i+1})$  is the female population exposure within the age interval  $[x_i, x_{i+1})$ ,  $i = 0, \dots, n$ .

For an elementary age interval  $[x, x+1)$ , the fertility rate is estimated as

$$f(x) = \int_x^{x+1} d\hat{F}(s) = \hat{F}(x+1) - \hat{F}(x), \quad (4.12)$$

where  $\hat{F}(x)$  is a continuous approximation of  $F(x)$  estimated using the interpolation technique. The algorithm for interpolating  $F(x)$  is the same as in Section 4.2, and consists of three steps. First, a logit transformation (4.6) is applied. Second, using piecewise cubic Hermite interpolation, a continuous approximation of the logit function is obtained. Third, an inverse logit transformation (4.7) is used to feed  $\hat{F}(x)$  values into formula (4.12).

Once the fertility rate  $f(x)$  has been calculated, the number of births in a one-year age group can be estimated as

$$\hat{B}(x) = f(x)E(x) \quad (4.13)$$

The sum of the estimated births over the aggregated age interval is, in general, not exactly equal to the original number of births in this interval. To obtain the original aggregated birth number, the following adjustment is made:

$$B(x) = \hat{B}(x) \frac{B(x_i; x_{i+1})}{\sum_{z=x_i}^{x_{i+1}-1} \hat{B}(z)}, \quad x_i \leq x < x_{i+1} \quad (4.14)$$

This procedure returns birth data by single age groups for each birth order (independently) as well as for all birth orders combined. To balance the data with respect to birth order, the IPF algorithm is applied.

#### 4.4 Iterative proportional fitting (IPF)

In the previous sections, we showed that common adjustments to raw data often result in adjusted age- and order-specific data cells that do not correspond exactly to the marginal totals. To ensure a balance, the iterative proportional fitting (IPF) is applied.

The IPF algorithm—also known as bi-proportional fitting in statistics, the RAS algorithm in economics, and matrix raking or matrix scaling in computer science—is an iterative algorithm for recalculating the cell values in a contingency table in such a way that the row and column sums of the cells become equal to the fixed row and column totals. More details about the IPF technique can be found in Fienberg (1970) and Bishop et al. (1975).

Let index  $k$  be an iteration number and

$B_i^k(x)$  is the estimated (after  $k$  steps) number of births of birth order  $i$  at age  $x$ ;

$B(x)$  is a fixed number of births at age  $x$  (all birth orders combined); and

$B_i$  is a fixed sum (over ages) of all births of birth order  $i$ .

We begin ( $k=0$ ) from the order-specific birth numbers obtained after applying the splitting procedures, or after distributing births with an unknown age of the mother and/or birth order

$$B_i^0(x) = B_i(x). \quad (4.15)$$

For values of  $k = 0, 2, 4, \dots$ , the age- and order-specific birth estimates at steps  $k+1$  and  $k+2$  are computed as

$$B_i^{k+1}(x) = \frac{B_i^k(x)}{\sum_i B_i^k(x)} \cdot B(x) \quad (4.16)$$

$$B_i^{k+2}(x) = \frac{B_i^{k+1}(x)}{\sum_x B_i^{k+1}(x)} \cdot B_i. \quad (4.17)$$

For each  $k$ , the distance  $\Lambda^k$  between the fixed marginal values is computed by a summation of  $B_i^k(x)$  computed after each completed iteration (i.e., for  $k = 2, 4, 6, \dots$ )

$$\Lambda^k = \sum_x \left( \sum_i B_i^k(x) - B(x) \right)^2 + \sum_i \left( \sum_x B_i^k(x) - B_i \right)^2. \quad (4.18)$$

Steps (4.16) to (4.17) are to be repeated until  $\Lambda^k < 10^{-6}$ .



## 5. Adjustments of census parity distributions

### 5.1 Estimating the age-parity distributions of women for 1 January of the census year

Census and survey data usually pertain to a specific date during the year  $t$ , denoted  $T'$ . The distribution of women by age  $y$  (ARDY) and parity  $i$  as of 1 January of the year  $t$ —i.e., population weights  $w_i(y, T'^{Jan1})$ —have to be estimated from the set of the known census weights  $w_i(y, T')$ <sup>16</sup> and from the unconditional fertility rates  $f_i(y, t)$ .

In the case of population data, the standard HFD age scale is ages 12 to 55. In some cases, parity distribution data by one-year age intervals are not available for this complete range (e.g., ages 15 to 49 can be with one-year age intervals and ages 50+ with 5-year age intervals). In such cases we require that the data are available with one-year age intervals for the ages 15 to 49. For any ages outside this minimum required range, we assume that

- women younger than the lowest age present in the data with a single year age interval ( $\hat{x}_{\min}$  where  $\hat{x}_{\min} \leq 15$ ) are childless (i.e.,  $w_{i=0}(x < \hat{x}_{\min}, T') = 1$  and  $w_{i>0}(x < \hat{x}_{\min}, T') = 0$ ), and
- above the highest age present in the data with a single year age interval ( $\hat{x}_{\max}$  where  $\hat{x}_{\max} \geq 49$ ), the age-parity distribution of women is stable and is equal to the distribution either
  - at the highest age present in the data if this is part of the continuous range of one-year age intervals (i.e.,  $w_i(x > \hat{x}_{\max}, T') = w_i(\hat{x}_{\max}, T')$  if  $\hat{x}_{\max}$  is with a single year age interval).
  - in the first multi-year age interval if one exists (i.e.,  $w_i(x > \hat{x}_{\max}, T') = w_i(\hat{x}_{\max} + 1, T')$  if  $\hat{x}_{\max} + 1$  is the first multi-year interval).

Where the minimum age range of one-year age interval data is not present for a given year, this parity data is not used as an input to the HFD.

If the census weights are originally given by age in completed years (ACY), they are first recalculated into the ARDY age:

$$w_i(y, T') = (1 - Z) \cdot w_i(x - 1, T') + Z \cdot w_i(x, T') \quad (5.1)$$

where  $Z$  is a fraction of the base year elapsing between the beginning of the year (1 January) and the date of the census. For a non-leap year this means,  $Z = D / 365$ , where  $D$  represents the number of days between 1 January and the date of the census. For instance, if the census took place on 15 May of a

<sup>16</sup> Before generating the census weights, women with unknown age or unknown parity are redistributed across the known categories applying the same procedure as described in Section 4.1.

non-leap year,  $Z$  would be computed as  $Z = 134 / 365 = 0.367$ , where 134 is the number of days that elapsed from 1 January through 15 May.

Assuming a linear distribution of births and birth rates,  $f_i(y,t)$ , over the census year  $t$ , the age-parity weights are estimated by progressing backwards from the census date to the beginning of the year  $t$ :

$$w_0(y,T) = w_0(y,T) + f_1(y,t) \cdot Z \quad (5.2)$$

$$w_i(y,T) = w_i(y,T) - f_i(y,t) \cdot Z + f_{i+1}(y,t) \cdot Z \quad \text{for } i=1, 2, 3; \quad (5.3)$$

$$w_{4+}(y,T) = w_{4+}(y,T) - f_4(y,t) \cdot Z \quad (5.4)$$

However, the fertility intensity varies during the year  $t$  since fertility changes with age. We solve this problem by performing for each age  $y$  a local cubic spline interpolation<sup>17</sup> across data points taken from the fertility schedule of cohorts observed during three consecutive years  $f_i(y-1,t-1)$ ,  $f_i(y,t)$ ,  $f_i(y+1,t+1)$  based on vertical parallelograms that are allocated in the middle of the corresponding age intervals  $y-0.5$ ,  $y+0.5$ ,  $y+1.5$ .

In the next step, the value of the spline function at point  $T=y+Z$  is deducted and labeled  $f'_i(y,t)$ . If the estimated fertility rates  $f'_i(y,t)$  turn negative (which is possible at very young or very old reproductive ages where two data points are zeros and the third is a non-zero), these values are replaced by zero. The spline-based values  $f'_i(y,t)$  are to be replaced by values  $f''_i(y,t)$ :

$$f''_i(y,t) = f'_i(y,t) \text{ if } f'_i(y,t) \geq 0, \text{ and} \quad (5.5)$$

$$f''_i(y,T) = 0 \text{ if } f'_i(y,T) < 0. \quad (5.6)$$

The procedure for estimating the age-parity distribution progressing backwards from the census date is:

$$w_0(y,T) = w_0(y,T) + f''_1(y,t) \cdot Z \quad (5.7)$$

$$w_i(y,T) = w_i(y,T) - f''_i(y,t) \cdot Z + f''_{i+1}(y,t) \cdot Z, \quad i=1, 2, 3 \quad (5.8)$$

$$w_{4+}(y,T) = w_{4+}(y,T) - f''_4(y,t) \cdot Z \quad (5.9)$$

If the estimated weights  $w_i(y,T)$  turn negative for some age and parity groups, the age-parity distribution of the female population is restricted to zero, and the adjusted distributions  $w'_i(y,T)$  are computed as follows:

$$w'_i(y,T) = w_i(y,T) \text{ if } w_i(y,T) \geq 0, \text{ and} \quad (5.10)$$

$$w'_i(y,T) = 0 \text{ if } w_i(y,T) < 0 \quad (5.11)$$

<sup>17</sup> For this purpose, the function `spline( age, fx, ..., method = "fmm" )` of the statistical package *R* is used.

Subsequently, the non-zero  $w'_i(y, T)$  are redistributed to ensure that their total equals one:

$$w''_i(y, T) = \frac{w'_i(y, T)}{\sum_i w'_i(y, T)} \quad (5.12)$$

## 5.2 Estimating the age-parity distribution from the base year forward

Once the age-parity distribution,  $w_i(y, T)$ , for  $T=1$  January of the census year  $t$  is obtained, the age-parity distribution on 1 January of the year  $t+1$  (denoted  $T+1$ ) is derived in the following way:

$w_0(y_{\min}) = 1$  and  $w_i(y_{\min}) = 0$ , where  $i=1, 2, 3, 4+$  and  $y_{\min}$  is a certain very young age at which all women are childless;<sup>18</sup>

$$w_0(y+1, T+1) = w_0(y, T) - f_1(y, t); \quad (5.13)$$

$$w_i(y+1, T+1) = w_i(y, T) + f_i(y, t) - f_{i+1}(y, t), \text{ for } i=1, 2, 3 \quad (5.14)$$

$$w_{4+}(y+1, T+1) = w_{4+}(y, T) + f_4(y, t). \quad (5.15)$$

It should be noted that the rates  $f_4(y, t)$  that are used to compute the distribution of women at the highest parity 4+ refer to birth order 4 (not to birth orders 4+).

In some instances, especially for higher birth orders at very young ages, the estimated relative age-parity distribution on 1 January of the year  $t+1$  may turn negative. This can happen due to an understatement of births by very young women in the census data, or because of the selective influence of migration, which was not accounted for in the estimation procedure. Therefore, we restrict the age-parity weights to the non-negative, and by replacing  $w_i(y, T)$  by  $w'_i(y, T)$ , we obtain  $w'_i(y, T) = w_i(y, T)$

$$\text{if } w_i(y, T) \geq 0, \text{ and} \quad (5.16)$$

$$w'_i(y, T) = 0 \text{ if } w_i(y, T) < 0 \quad (5.17)$$

Then all of the weights  $w'_i(y, T)$  are further transformed to ensure that their total over  $i$  equals one:

$$w''_i(y, T) = \frac{w'_i(y, T)}{\sum_i w'_i(y, T)} \quad (5.18)$$

---

<sup>18</sup>  $y_{\min} = 11$  whenever data are available or any age between 11 and 15, depending on the youngest age for which the census parity weights are available  $y_{\min} = y_{\text{youngest}} - 1$ .

## 6. Unconditional fertility rates and summary indicators

All of the calculation procedures described in this section, except for the computation of the crude birth rate, are based on birth data  $B(x,t,c)$  and female population exposures  $E(x,t,c)$  by Lexis triangles that are available from the Lexis Database. For most of the countries, order-specific data on births  $B_i(x,t,c)$  are available as well, although frequently not for the whole period covered by the data for all of the birth orders combined. The crude birth rate is based on total live births  $B(t)$  and total population exposure  $E_{total}(t)$ .

### 6.1 Calculation of population exposure

Female population exposure  $E(x,t,c)$  is calculated for each Lexis triangle and then aggregated as necessary. The method of calculation used depends on the availability of monthly birth data for the cohort  $c$ . Since the duration of exposure to risk at age  $x$  within a calendar year varies from one individual to another depending on the date of birth, if the births are distributed evenly throughout the year, then each individual will on average contribute an exposure of 0.5 years. Where the mean date of birth for a cohort,  $\bar{b}(c)$ , measured in years from 1 January, is markedly different from 0.5<sup>19</sup>, the standard estimate of exposure as one half of the population on the vertical edge of the triangle is repeatedly affected for all the Lexis triangles for this cohort.

For cohorts for whom the mean date of birth can be calculated, exposure is therefore calculated using Method 1, based on the method presented by Calot and Sardon (2003), which assumes constancy of risk according to age within the triangle. The assumptions inherent in this method are detailed in Calot and Sardon (2003). For cohorts for whom no monthly birth data are available and the mean date of birth must therefore be assumed to be 0.5, a slight improvement in the exposure to risk estimation is obtained by using mortality data (i.e., adjustment of exposure due to deaths in the triangle) according to the method used in the HMD (Wilmoth et al, 2007).

Method	$E(x,t,t-x-1)$	$E(x,t,t-x)$
1	$P(x,t) \cdot \bar{b}(t-x-1)$	$P(x,t+1) \cdot (1 - \bar{b}(t-x))$
2	$\frac{P(x,t)}{2} - \frac{D(x,t,c)}{6}$	$\frac{P(x,t+1)}{2} + \frac{D(x,t-1,c)}{6}$

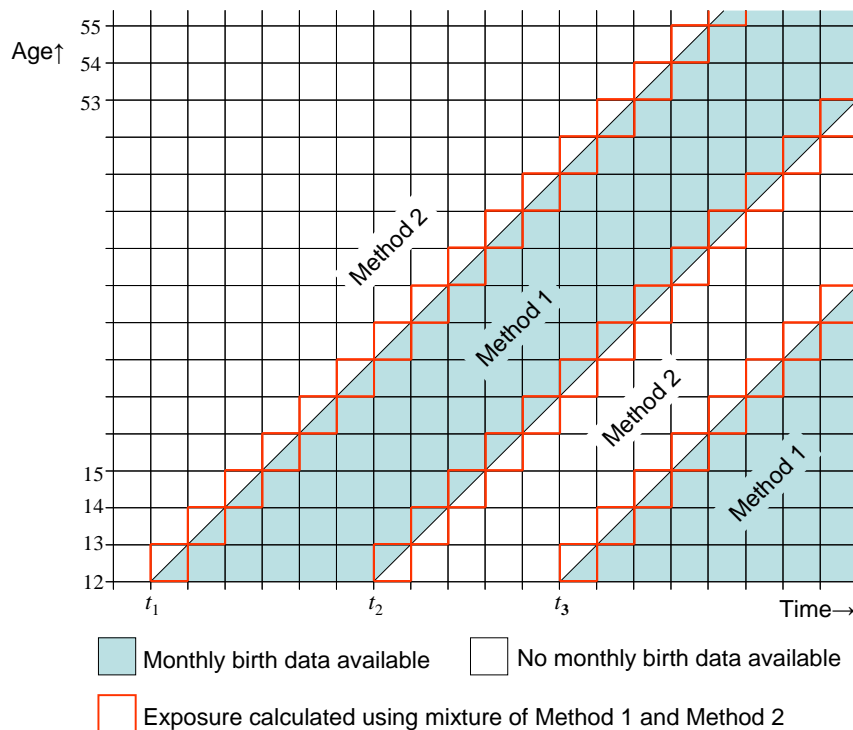
$E(x,t,c)$  Female population exposure at age  $x$  in year  $t$  for cohort  $c$

<sup>19</sup> A considerably higher concentration of birthdays in the early months of the year implies a value below 0.5; and, conversely, many birthdays in the last months of the year imply a value above 0.5

$P(x,t)$	Population size at age $x$ on 1 January in year $t$
$D(x,t,c)$	Deaths at age $x$ in year $t$ for cohort $c$
$\bar{b}(c)$	Mean date of birth for cohort $c$

Since in some cases the series of monthly birth data are non-continuous (with, for example, breaks for war years), there may be multiple boundaries between the two methods for a single country. Figure 1 shows an example in which monthly birth data are available for the period  $t_1 - t_2$  and from  $t_3$  onwards: in this case, for all cohorts born before  $t_1$  and from  $t_2 - t_3$ , Method 2 is used; whereas for those cohorts born from  $t_1 - t_2$  and after  $t_3$ , Method 1 using  $\bar{b}(c)$  is used. For each country in the HFD, the availability of monthly birth data is described in the country documentation. The method used for calculating exposure for any given triangle may be determined from this documentation.

It should be noted that population exposure is calculated by triangles and then aggregated to parallelograms and rectangles as required. The exposure in rectangles along the boundary between regions is therefore calculated by different methods (highlighted in Figure 1) using a mixture of the two methods.



**Figure 6.1. Lexis map showing the method used for calculating female population exposure where monthly birth data are not continuously available**

Total population exposure  $E_{total}(t)$ , used for the computation of crude birth rates, is obtained by aggregating female and male population exposures  $E(x, t, c)$  over ages and adding the sums together. Male population exposure is estimated by the same method as female population exposure (see the description of the two possible methods above).

## 6.2 Unconditional age-specific fertility rates

Age-specific fertility rates and the indicators derived from them are the most commonly employed measures of fertility, and have been in use for about a century (Kuczynski, 1932).

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.

Having both birth counts and female population exposure split by Lexis triangles allows us to compute unconditional fertility rates for all Lexis shapes (Lexis triangles, Lexis squares, vertical parallelograms, and horizontal parallelograms) as follows:

*Unconditional age-specific fertility rates for year  $t$ , cohort  $c$ , age  $x$ , and birth order  $i$  (Lexis triangle):*

$$f_i(x, c, t) = \frac{B_i(x, c, t)}{E(x, c, t)}. \quad (6.1)$$

*Unconditional age-specific fertility rates for year  $t$ , age  $x$ , and birth order  $i$  (Lexis square):*

$$f_i(x, t) = \frac{B_i(x, t, t - x) + B_i(x, t, t - x - 1)}{E(x, t)}. \quad (6.2)$$

*Unconditional age-specific fertility rates for year  $t$ , birth cohort  $c$ , and birth order  $i$  (vertical parallelogram):*

$$f_i(t, c) = \frac{B_i(x - 1, t, t - x) + B_i(x, t, t - x)}{E(t, c)}. \quad (6.3)$$

*Unconditional age-specific fertility rates for cohort  $c$ , age  $x$ , and birth order  $i$  (horizontal parallelogram):*

$$f_i(x, c) = \frac{B_i(x, t, t - x) + B_i(x, t + 1, t - x)}{E(x, c)}. \quad (6.4)$$

Equivalent rates can be computed for all birth orders combined by placing the births of all of the birth orders in the numerators of formulae (6.1)–(6.4).

### 6.3 Cumulative and total fertility rates

Cumulative fertility rates computed for birth cohorts refer to the average number of children born to a woman by a certain age. They are usually shown for all birth orders combined, but they can also be disaggregated by birth order. When computed from period fertility rates, cumulative fertility 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.

In the HFD, cumulative fertility rates are calculated from unconditional age-specific fertility rates sorted by Lexis squares and vertical parallelograms (period dimension) and horizontal parallelograms (cohort dimension):

*Cumulative period fertility rates* by age  $x$  for year  $t$  for all birth orders combined (Lexis squares and vertical parallelograms):

$$CPFR(x, t) = \sum_{z=x_{\min}}^{x-1} f(z, t), \quad (6.5a)$$

*Cumulative period fertility rates* by age  $x$  for year  $t$  for birth order  $i$  (Lexis squares and vertical parallelograms):

$$CPFR_i(x, t) = \sum_{z=x_{\min}}^{x-1} f_i(z, t). \quad (6.5b)$$

In formulae (6.5a) and (6.5b),  $x$  and  $z$  refer to the age in completed years (ACY) in case of the Lexis squares and the age reached during the year (ARDY) for Lexis vertical parallelograms;  $x_{\min}$  corresponds to age 12 or younger. If the upper age limit of the summation is equal or very close to the maximum reproductive age (i.e., if it is 50 or higher), the cumulative fertility rate equals the total fertility rate (*TFR*).

The cumulative cohort fertility rate (*CCFR*) refers to the average number of children born to a woman from birth cohort  $c$  by age  $x$ , and is computed by summing up the set of age-specific fertility rates of the cohort  $c$  observed over their reproductive lives up to age  $x$ . *CCFRs* are calculated for all cohorts  $c$  who are observed from age  $x_{\min}$  that is equal to 15 or younger.

*Cumulative cohort fertility rates* by age  $x$  for cohort  $c$  for all birth orders combined (horizontal parallelogram) is

$$CCFR(x, c) = \sum_{z=x_{\min}}^{x-1} f(z, c). \quad (6.6a)$$

*Cumulative cohort fertility rate* by age  $x$  for cohort  $c$  and birth order  $i$  (horizontal parallelogram) is

$$CCFR_i(x, c) = \sum_{z=x_{\min}}^{x-1} f_i(z, c). \quad (6.6b)$$

The *period total fertility rate* for all birth orders combined and by birth order is computed as follows:

$$TFR(t) = \sum_{x=x_{\min}}^{x_{\max}} f(x,t), \quad (6.7a)$$

$$TFR_i(t) = \sum_{x=x_{\min}}^{x_{\max}} f_i(x,t). \quad (6.7b)$$

In formulae (6.7a) and (6.7b),  $x_{\min}$  corresponds to 12 years or younger. The values of the *TFR* and *TFR<sub>i</sub>* are computed for age  $x_{\max}=55+$  years; i.e., for the age span covering all reproductive ages. The HFD also lists a parallel estimate based on the sum of the observed fertility rates by age 40; i.e., with  $x_{\max}=39$  years. This information is more useful for cohort fertility analysis, where the cumulated fertility rates of cohorts nearing the end of their reproductive period provide a valuable approximation of their future completed fertility.

For birth cohorts, the corresponding quantities represent the completed cohort fertility (*CCF*). The *completed cohort fertility* for all birth orders combined and by birth order is computed as follows:

$$CCF(c) = \sum_{x=x_{\min}}^{x_{\max}} f(x,c), \quad (6.8a)$$

$$CCF_i(c) = \sum_{x=x_{\min}}^{x_{\max}} f_i(x,c) \quad (6.8b)$$

The *CCF* is calculated for all cohorts  $c$  that are observed from age  $x_{\min}$  that is equal to age 15 or younger until age 50 or older. Again, two types of the *CCF* are shown. The first one represents the *CCF* at age 50 or older ( $x_{\max}=49+$  years), whereas the second one shows the *CCF* (or, more correctly, cumulated cohort fertility) by age 40 (with  $x_{\max}=39$  years) and thus represents an incomplete approximation of the future *CCF*.

#### **6.4 Tempo-adjusted total fertility rate**

Changes in period fertility measures are often driven by the temporary postponement or advancement of births. It is therefore difficult to identify to what extent fluctuations seen in the period *TFR* result from such “timing” changes, and to what extent these are “real” (quantum) changes that would influence the completed fertility of real birth cohorts. A comparison of period and cohort fertility measures reveals that tempo distortions can cause a substantial gap between the two indicators for an extended period of time (Sobotka, 2004a, 2004b).

Tempo distortions in period fertility measures have inspired efforts to develop an adjustment method that would help to eliminate them. A simple and widely used *TFR* adjustment, based on order-specific *TFRs* and changes in order-specific mean ages at birth, was proposed by Bongaarts and Feeney



(1998). The Bongaarts-Feeney tempo-adjusted TFR is computed as a sum of order-specific TFRs adjusted for changes in the mean age of order-specific fertility schedule,  $r_i(t)$  as shown in formulae (6.9a) and (6.9b).

$$adjTFR(t) = \sum_i adjTFR_i(t) \quad (6.9a)$$

$$adjTFR_i(t) = \frac{TFR_i(t)}{1 - r_i(t)} \quad (6.9b)$$

Following Bongaarts and Feeney (2000: 563), the adjustment factor  $r_i(t)$  is estimated as follows:

$$r_i(t) = \frac{MAB_i(t+1) - MAB_i(t-1)}{2} \quad (6.10)$$

where  $MAB_i(t)$  is the mean age at birth order  $i$  calculated from unconditional age- and order-specific fertility rates (see equation 6.11b below). The tempo distortion in the observed  $TFR$  then equals  $adjTFR(t) - TFR(t)$ .

Due to its simplicity and lower data requirements the Bongaarts-Feeney adjustment remains more widely used than any alternative method of this kind. This was one of the reasons why it was included in the HFD. However, the tempo-adjusted TFR often displays large year-to-year fluctuations (Sobotka 2003), which can make its use for specific years problematic. We therefore recommend smoothing the time series of the tempo-adjusted TFR by using, for instance, simple three-year or five-year moving averages (see Goldstein et al. 2009 and VID 2010 for illustrations).

## 6.5 Mean ages at birth

*Mean age at birth (MAB)* refers to the average age of the mother at childbearing, standardised for the age structure of the female population of reproductive age. It is based on the schedule of unconditional age-specific fertility rates. Like in the case with the TFR, the HFD also provides the mean age at birth by age 40; it is based on the schedule of age-specific fertility rates with  $x_{\max}=39$  years.

The *period mean age at birth* for all birth orders combined and by birth order  $i$  are:

$$MAB(t) = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot f(x,t)}{\sum_{x=x_{\min}}^{x_{\max}} f(x,t)}, \quad (6.11a)$$

$$MAB_i(t) = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot f_i(x,t)}{\sum_{x=x_{\min}}^{x_{\max}} f_i(x,t)}. \quad (6.11b)$$

In formulae (6.12a) and (6.12b),  $x_{\min}$  and  $x_{\max}$  are exactly the same as those in formulae (6.7a) and (6.7b).

*The cohort mean age at birth* for all birth orders combined and by birth order  $i$  are:

$$CMAB(c) = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot f(x,c)}{\sum_{x=x_{\min}}^{x_{\max}} f(x,c)}, \quad (6.12a)$$

$$CMAB_i(c) = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot f_i(x,c)}{\sum_{x=x_{\min}}^{x_{\max}} f_i(x,c)} \quad (6.12b)$$

$CMABs$  in formulae (6.12a) and (6.12b) are computed for exactly the same cohorts and over the same ranges of ages as those in formulae (6.8a) and (6.8b).

Value  $\bar{x}$  in formulae (6.11a), (6.11b), (6.12a) and (6.12b) is the mean age at birth within the elementary age interval  $[x, x+1)$ :

$$\bar{x} = x + a(x), \quad (6.13)$$

where  $a(x)$  is the average share of the age interval  $[x, x+1)$  lived before giving birth to a child. We assume that all  $a(x)$  values are equal to 0.5 for any completed age  $x$  and birth order  $i$  (for data organised by Lexis squares and horizontal parallelograms) and zero for any age  $x$  reached during the year and birth order  $i$  (for data organised by vertical parallelograms).

## 6.6 Standard deviation in the mean age at birth

The standard deviation in the mean age at birth shows the extent of variation in the timing of childbearing. A low value of standard deviation points to fertility concentration around the mean age at birth, whereas a high value of standard deviation suggests that fertility is spread over a broader range of ages. The standard deviation in the mean age at birth is also computed by age 40.

*The standard deviation in the mean age at birth* for all birth orders combined and by birth order  $i$  for year  $t$  are calculated as follows:

$$sdMAB(t) = \sqrt{\left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x}^2 \cdot \frac{f(x,t)}{\sum_{x=x_{\min}}^{x_{\max}} f(x,t)} \right] - \left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot \frac{f(x,t)}{\sum_{x=x_{\min}}^{x_{\max}} f(x,t)} \right]^2} \quad (6.14a)$$

$$sdMAB_i(t) = \sqrt{\left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x}^2 \cdot \frac{f_i(x,t)}{\sum_{x=x_{\min}}^{x_{\max}} f_i(x,t)} \right] - \left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot \frac{f_i(x,t)}{\sum_{x=x_{\min}}^{x_{\max}} f_i(x,t)} \right]^2} \quad (6.14b)$$

The analogous indicators for cohort  $c$  are calculated using these formulae:

$$sdCMAB(c) = \sqrt{\left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x}^2 \cdot \frac{f(x,c)}{\sum_{x=x_{\min}}^{x_{\max}} f(x,c)} \right] - \left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot \frac{f(x,c)}{\sum_{x=x_{\min}}^{x_{\max}} f(x,c)} \right]^2} \quad (6.15a)$$

$$sdCMAB_i(c) = \sqrt{\left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x}^2 \cdot \frac{f_i(x,c)}{\sum_{x=x_{\min}}^{x_{\max}} f_i(x,c)} \right] - \left[ \sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot \frac{f_i(x,c)}{\sum_{x=x_{\min}}^{x_{\max}} f_i(x,c)} \right]^2} \quad (6.15b)$$

We recall that value  $\bar{x}$  in formulae (6.14a), (6.14b), (6.15a), and (6.15b) is the mean age at birth within the age interval  $[x, x+1)$  and equals  $x + a(x)$ , where  $a(x)$  is the average share of the age interval  $[x, x+1)$  lived before giving birth to a child. We assume that all  $a(x)$  values are equal to 0.5 for any completed age  $x$  and birth order  $i$  (for data organised by Lexis squares and horizontal parallelograms).

## 6.7 Cohort parity progression ratios

Childbearing behaviour can be analysed not only by following a woman's movement along the age scale, but also by following her progression to higher parities. The most common summary measures describing the movement from one parity to the next are *parity progression ratios (PPRs)*. The parity progression ratio expresses the probability of giving birth to another child ( $i+1$ ), conditional on reaching parity  $i$ . In the HFD, in the same way as for the other summary indicators of cohort fertility, the parity progression ratios are displayed for the cohorts who have completed 49 years of age.

$$PPR_{0,1}(c) = CTFR_1(c) \quad (6.16a)$$

$$PPR_{i-1,i}(c) = \frac{CTFR_i(c)}{CTFR_{i-1}(c)}, \text{ for } i > 1 \quad (6.16b)$$

It should be noted that the highest parity-progression ratio computed in the HFD pertains to the progression from third to the fourth birth ( $PPR_{3,4}$ ), as the data for the highest birth order category included in the HFD lump together all fifth and subsequent births, and do not allow fifth births to be placed in a separate category.

## 6.8 Crude birth rates

The crude birth rate is the simplest measure of birth intensity. Unlike the above summary measures, it is not standardised for the age structure of the population. The rate is calculated by dividing the total number of live births occurring in a given year  $t$  by person-years lived by all population in that year. The crude birth rate is often expressed as the number of live births per 1,000 of population.

The *crude birth rate* for all birth orders combined:

$$CBR(t) = \frac{B(t)}{E_{total}(t)} \times 1000 \quad (6.17a)$$

The *crude birth rate* by birth order  $i$ :

$$CBR_i(t) = \frac{B_i(t)}{E_{total}(t)} \times 1000 \quad (6.17b)$$

where  $E_{total}(t)$  stands for the total population across all ages; i.e., including men and women and also including non-reproductive ages. Apart from the intensity of births, the *CBR* depends on the population distributions by age and sex, and should be used with caution.

## 7. Cohort fertility tables

Cohort fertility tables are increment-decrement life tables that model the process of childbearing in female birth cohorts. In the HFD, they describe a two-dimensional cohort progression of women toward older age and higher parities. Women in each cohort of interest are moving from parity zero to parity one and to subsequent parities by giving birth to children of the corresponding birth orders as they age. The distribution of births by the age of the mother and birth order in the cohort fertility table corresponds to the observed fertility trajectories of real cohorts.

The cohort fertility tables are constructed only if sufficiently long time series of data on fertility by birth order are available (Figure 3.4). These data should allow us to observe female cohorts from the beginning of their reproductive age (age 15 or younger) until age 25 or an older age.

### 7.1 Construction of the cohort table

For each cohort  $c$ , all of the functions of the cohort fertility table are computed from the schedule of unconditional age-specific fertility rates by birth order  $f_i(x, c)$ , which are computed according to (6.4).

The following functions (columns) of the cohort life tables are computed and displayed in the HFD:  $x$ ,  $b_i(x)$ ,  $l_{i-1}(x)$ ,  $q_i(x)$ ,  $m_i(x)$ ,  $Sb_i(x)$ ,  $chi(x)$  (see also Appendix 1).

Working with a table population of 10,000 women, table births by birth order are computed for each age  $x$ :

$$b_i(x) = 10,000 \cdot f_i(x, c) \quad (7.1)$$

This formula considers birth as a repeatable event. This implies that the total number of individuals in the life table, which is 10,000 in the HFD, remains the same at any age. The individuals move, however, towards higher parities, in line with a given schedule of unconditional rates by birth order.

At any age  $x$ , the life table cohort of the size 10,000 is divided into parity-specific sub-cohorts,  $l_i(x)$ . The cohort progresses over ages and parities, starting from the initial childless status at the minimum age at childbearing  $x_{min}$ , as follows:

$$l_0(x_{min}) = 10,000 \text{ (radix of the cohort)} \quad (7.2)$$

$$l_i(x_{min}) = 0, \text{ for } i = 1, 2, 3, 4 \quad (7.3)$$

$$l_i(x) = l_i(x-1) - b_{i+1}(x-1), \text{ for } i = 0 \quad (7.4)$$

$$l_i(x) = l_i(x-1) + b_i(x-1) - b_{i+1}(x-1), \text{ for } i = 1, 2, 3 \quad (7.5)$$

$$l_{i+}(x) = l_{i+}(x-1) + b_i(x-1), \text{ for } i = 4 \text{ (} i+ \text{ stands for women at parities } i \text{ and higher)} \quad (7.6)$$

Life table age- and parity-specific fertility rates (conditional rates, occurrence-exposure rates) for women aged  $x$  and at parity  $i$  are obtained by relating births of the order  $i$  at age  $x$  to person-years lived at this age at parity  $i-1$ :

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) - a(x) \cdot b_i(x)}, \text{ for } i = 1 \quad (7.7)$$

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) - a(x) \cdot b_i(x) + (1 - a(x)) \cdot b_{i-1}(x)}, \text{ for } i = 2, 3, 4 \quad (7.8)$$

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) + (1 - a(x)) \cdot b_{i-1}(x)}, \text{ for } i = 5+ \quad (7.9)$$

The conditional probability of giving an  $i$ -th birth at age  $x$  for a woman of parity  $i-1$  is expressed as:

$$q_i(x) = \frac{b_i(x)}{l_{i-1}(x)}, \quad i=1, 2, 3, 4, 5+ \quad (7.10)$$

Cumulative births by age  $x$  and birth order  $i$  can be computed simply by summing up the table births of order  $i$  at all ages through  $x-1$ :

$$Sb_i(x) = \sum_{z=x_{\min}}^{x-1} b_i(z), \quad i=1, 2, 3, 4, 5+ \quad (7.11)$$

The average number of children born to a woman in the highest parity category 4+ by the time she reaches age  $x$ :

$$chi(x) = \frac{4 \cdot l_4(x) + \sum_{z=x_{\min}}^{x-1} b_{5+}(z)}{l_4(x)} \quad (7.12)$$

At very young ages, the number of births observed at higher birth orders in a population is very low and fluctuates strongly. This may lead to negative values of  $l_{i-1}(x)$  or to  $q_i(x)$  exceeding 1. In such cases, we do the following adjustments in the computed values for  $l_{i-1}(x)$ ,  $b_i(x)$ ,  $q_i(x)$ , and  $m_i(x)$ <sup>20</sup>:

$$\text{If } l_{i-1}(x) < 0 \text{ then } l_{i-1}(x) = 0, b_i(x) = 0, m_i(x) = ., q_i(x) = . \quad (7.13)$$

$$\text{If } l_{i-1}(x) = 0 \text{ then } b_i(x) = 0, m_i(x) = ., q_i(x) = . \quad (7.14)$$

$$\text{If } l_{i-1}(x) < b_i(x) \text{ then } b_i(x) = l_{i-1}(x), m_i(x) = ., q_i(x) = . \quad (7.15)$$

## 8. Period fertility tables

Like the cohort fertility tables, the period fertility tables used in the HFD are increment-decrement life tables, which model the process of childbearing in hypothetical (synthetic) cohorts of women specified by age and parity. The period fertility tables describe the childbearing histories of synthetic cohorts of women who would live their whole lives under the fertility conditions of a given year  $t$ , assuming the absence of mortality and migration. They are built using a multi-state life table method, where states are defined by current parity. The basic parameters from which all of the other estimates are derived are the conditional age- and order-specific fertility rates and probabilities.

Period fertility tables controlling for age and the parity composition of the female population of reproductive ages provide a rich set of indicators that enable us to conduct a thorough analysis of fertility level and timing. So far, the period fertility tables have not been widely used, in part due to limited data availability. There are many analogies to the commonly used mortality tables, but the main function of fertility tables, unlike of mortality tables, is not the analysis of the timing of births (an

<sup>20</sup> Dots provided for  $q_i(x)$  and  $m_i(x)$  show that the values in these cells are implausible.

analogy to life expectancy would be waiting time to the birth of the next child), but the levels and trends of fertility across different ages and parity categories.

### 8.1 Parity-specific population exposure

Annual time series of female population exposure by age and parity, which are necessary for computing conditional fertility rates  $m_i(x)$  and the period fertility tables, are obtained from the input data on female population exposures  $E(x,t)$  and the mid-year estimates of the age- and parity-specific distribution of women  $w_i(x,t)$ :

$$E_{i-1}(x,t) = w_{i-1}(x,t) \cdot E(x,t) \quad (8.1)$$

Depending on data availability, different approaches are used for obtaining the estimates of  $w_i(x)$ . They can be defined by the following hierarchy:

#### 8.1.1 Cumulating cohort fertility rates over long periods of time

For countries that have a sufficiently long time series of period data on births by the age of the mother and birth order, annual estimates of the age-parity distribution  $w_{i1}(x)$  are reconstructed from these data by cumulating fertility of cohorts over their reproductive age span:

$$w_i(x_{\min}, t) = 1, \text{ for } i = 0 \quad (8.2)$$

$$w_i(x_{\min}, t) = 0, \text{ for } i = 1, 2, 3, 4+ \quad (8.3)$$

$$w_i(x, t) = \frac{l_i(x, t-x) + l_i(x+1, t-x-1)}{2 \cdot l_0(x_{\min}, t-x_{\min})}, \text{ for } x_{\min} < x < x_{\max}$$

(note that  $l_0(x_{\min}, t-x_{\min})$  is the radix and equals 10,000 and  $l_i(x, t-x)$  is a size of cohort  $t-x$  by age  $x$ )

$$(8.4)$$

It should be noted that, in incomplete period fertility tables constructed for the years  $t_0+k$ ,  $30 \leq k < 40$  (see Figure 3.5), the last computed  $w_i(x)$  is for the age  $\hat{x}_{\max} - 1$  where  $\hat{x}_{\max} = k + 15$ <sup>21</sup> is the last age for which  $l_i(x)$  is calculated. This last  $w_i(\hat{x}_{\max} - 1)$  is calculated according to formula (8.4).

Starting from the year  $t_0+40$ ,  $\hat{x}_{\max}$  always equals  $x_{\max}$  and  $w_i(x_{\max})$  is estimated as follows:

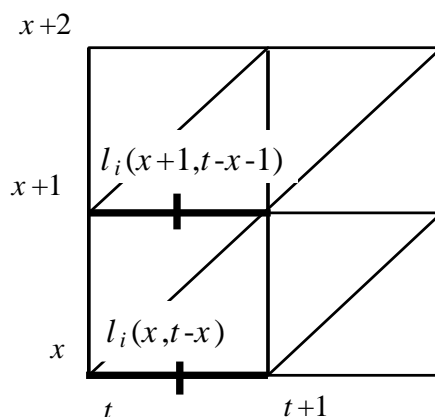
$$w_i(x_{\max}, t) = \frac{l_i(x, t-x)}{l_0(x_{\min}, t-x_{\min})} \quad (\text{note that } l_0(x_{\min}, t-x_{\min}) \text{ is the radix and equals 10,000}) \quad (8.5)$$

where  $l_i(x)$  is the table cohort size specific for age and parity, derived from the cohort fertility tables (see Figure 8.1). In order to ensure that the sum of the weights obtained in formulae (8.1)–(8.5) is equal to one in each age category, the values of weights are adjusted analogously to formula (5.18).

---

<sup>21</sup> i.e.  $45 \leq \hat{x}_{\max} \leq 55+$

Having the age-parity distribution  $w_i(x)$  estimated by this approach, the period fertility tables can be built beginning from the year when the first cohort observed from the minimum age reaches age 45. As shown in Figure 3.5, it covers the Lexis region A and extends over the period  $t_0+30$  to  $T$ .



**Figure 8.1. Estimation of  $w_i(x)$  using the life table cohort size by age and parity,  $l_i(x)$  from the cohort fertility tables**

### 8.1.2 Use of a “golden” census

For countries where age- and order-specific data on births are available for a short period only, population census or population register data (called the “golden” census thereafter) are used to derive the initial age-parity distribution (for one starting year, the “base year”), which is then annually updated by cumulating fertility of cohorts over their childbearing ages, as described in Section 8.1.1. In exceptional cases, when no other method for deriving the initial age-parity distribution can be employed, survey data may be considered. Such a survey must cover a large and representative sample of at least 1% of the female population of reproductive age. The use of a “golden” census often makes it possible to construct the period fertility tables for an extended period of time prior to the year when the age-parity distribution for women aged 45 or older can be constructed entirely from the time series of cohort fertility rates by age and birth order.

The decision about which data to use as the “golden” census may be based on several considerations, which are not clear-cut. The decision should be made in consultation with the country expert, and will depend both on the length of the time series of the available order-specific data in the country, as well as on the quality and reliability of the data of the potential “golden” census.

The calculation of the census-based age-parity distribution involves cohorts born in the years  $t_{\text{cens}}-x_{\text{max}}$  through  $t_{\text{cens}}-12$  (Figure 3.5). For these cohorts, the cohort fertility tables (not featured in the HFD) are left-censored and start from certain non-null values of  $l_i(x)$  that are computed using the census weights  $w_i(x, t_{\text{cens}}-x)$ :

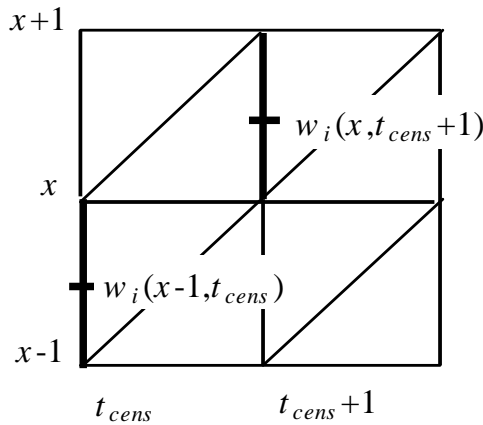


$$l_i(x, t_{cens} - x) = 10,000 \cdot w_i(x, t_{cens} - x), \tag{8.6}$$

where 10,000 is the radix of the cohort born in the year  $t_{cens}-x$ . The census weights  $w_i(x, t_{cens}-x)$  are calculated as (see also Figure 8.2):

$$w_i(x, t_{cens} - x) = \frac{w_i(x-1, t_{cens}^{Jan1}) + w_i(x, [t_{cens} + 1]^{Jan1})}{2}. \tag{8.7}$$

The values of  $l_i(x)$  are estimated for all of the cohorts of reproductive age (observable in the Lexis region B in Figure 3.5), and the mid-year age-parity distribution  $w_i(x)$  is computed as expressed in formulae (8.2)–(8.5) and shown in Figure 8.1. Based on these  $w_i(x)$  estimates, combining census-based values with the subsequent cohort fertility time series, the female age- and parity-specific exposure in period fertility tables is computed for the year  $t_{cens}$  and the subsequent years. Starting from the first year when the age-parity distribution can be obtained and updated for all of the subsequent years for women aged 45 or older *purely* by cumulating cohort population exposures (year  $t_0+30$  and region A in Figure 3.5), the age-parity distribution is obtained and updated using these cumulated cohort fertility series, as described in Section 8.1. This introduces a one-time series break in fertility tables, and also temporarily reduces the range of high reproductive ages covered by the tables, as ages 46-55 are excluded until the year when the cohort parity distribution can also be estimated for these women (see Figure 3.5).



**Figure 8.2. Estimation of the census weights used to produce the left-censored cohort fertility tables**

It is important to note that the census or register data usually pertain to a specific date in a given year, while many HFD computations, including the application of the golden census method, are based on the age-parity distribution of women at the beginning of the year (e.g., formula 8.7). The procedures for reconstructing the data for the age-parity distribution on 1 January of the census year and of the subsequent year are described in Sections 5.1 and 5.2 of this document.

### 8.1.3 Direct use of census or register data

Finally, for countries where census or high-quality population register data on the parity distribution of women by age exist, census- or register-based fertility tables can be constructed for the years covered by these data. These alternative fertility tables are featured in the HFD alongside the main set of fertility tables described above. The computation procedure of the mid-year value of  $w_i(x)$  in each census or register year is identical to that applied in the case of the golden census, described in Section 8.1.2. It should be noted that, for the register-based tables, unlike for the census-based tables, annual series of these data estimated for 1 January of each year are readily available, and no additional adjustments to derive the age-parity weights for 1 January are needed<sup>22</sup>. The computation can, however, be performed for the last calendar year in the series, when the data on births in this year are available and when the age-parity distribution of women is available only for 1 January. In this case, preliminary estimates of  $w_i(x)$  are computed for 1 January of the subsequent year following the procedure described in Section 5.2.

## 8.2 Conditional age-specific fertility rates

Conditional rates are obtained by dividing the number of  $i$ th births to women at age  $x$  in a year  $t$  by the person-years lived by women aged  $x$  and at parity  $i-1$ , and who are thus exposed to the risk of having an  $i$ th birth in the year  $t$ :

$$m_i(x, t) = \frac{B_i(x, t, t-x-1) + B_i(x, t, t-x)}{E_{i-1}(x, t)} = \frac{B_i(x, t)}{E_{i-1}(x, t)} \quad (8.8)$$

To eliminate huge fluctuations in  $m_i(x)$  at the lowest and highest childbearing ages, which are attributable to a very low observed numbers of births, these rates are computed in the HFD only for age and parity combinations where  $E_{i-1}(x) \geq 5$ ,  $m_i(x) < 1$  and  $q_i(x) < 1$ ; otherwise, the values of  $E_{i-1}(x)$  are replaced by zero and the values of  $q_i(x)$  and  $m_i(x)$  are presented by dots (implausible data) in the fertility tables.

From a theoretical perspective, conditional rates (rates of the first type, occurrence-exposure rates) are preferred to unconditional rates because they are in line with the principle of correspondence between the nominator and the denominator. Specifically, only women who are de facto at risk of having an  $i$ -th birth (i.e., those at parity  $i-1$ ) are included in the denominator when fertility rates for birth order  $i$  are computed (this assumption ignores multiple births). This gives conditional period

---

<sup>22</sup> Register data on women by age and parity can also refer to 31 December. Such data are treated as pertaining to 1 January of the subsequent year.

fertility rates  $m_i(x)$  an advantage over the unconditional rates  $f_i(x)$  that may be distorted by compositional effects due to the changing parity structure of the female population.

### 8.3 Construction of the period table

Conditional fertility rates,  $m_i(x,t)$ , which are further converted into probabilities,  $q_i(x,t)$ , serve as the major input for the construction of the period fertility tables. This approach was selected for the HFD because it allows us to at least partly account for the effects of mortality and migration on population exposure  $E_{i-1}(x,t)$  over the year (period)  $t$ .

The functions (columns) of the period life tables computed and displayed in the HFD are as follows:  $x$ ,  $w_{i-1}(x)$ ,  $m_i(x)$ ,  $q_i(x)$ ,  $l_{i-1}(x)$ ,  $b_i(x)$ ,  $L_{i-1}(x)$ ,  $Sb_i(x)$  (see also Appendix 1). These functions are analogous to those in the cohort life table, described in Section 7. They are computed by using the following sequence of formulae:

$$q_i(x) = \frac{m_i(x)}{1 + [1 - a(x)] \cdot m_i(x)} \quad (8.9)$$

$$l_0(x_{\min}) = 10,000 \text{ (the radix)} \quad (8.10)$$

$$l_i(x_{\min}) = 0, \text{ for } i = 1, 2, 3, 4 \quad (8.11)$$

$$l_i(x) = l_i(x-1) \cdot [1 - q_{i+1}(x-1)], \text{ for } i = 0 \quad (8.12)$$

$$l_i(x) = l_i(x-1) - b_{i+1}(x-1) + L_{i-1}(x-1) \cdot m_i(x-1), \text{ for } i = 1, 2, 3 \quad (8.13)$$

$$l_{i+}(x) = l_{i+}(x-1) + L_{i-1}(x-1) \cdot m_i(x-1),$$

for  $i = 4$  (recall that  $i+$  stands for women at parities  $i$  and higher) (8.14)

$$b_i(x) = L_{i-1}(x) \cdot m_i(x) \quad (8.15)$$

$$L_i(x) = l_i(x) - l_i(x) \cdot q_{i+1}(x) \cdot [1 - a(x)], \text{ for } i = 0 \quad (8.16)$$

$$L_i(x) = l_i(x) + l_{i-1}(x) \cdot q_i(x) \cdot [1 - a(x)] - l_i(x) \cdot q_{i+1}(x) \cdot [1 - a(x)], \text{ for } i = 1, 2, 3 \quad (8.17)$$

$$L_{i+}(x) = l_{i+}(x) + l_{i-1}(x) \cdot q_i(x) \cdot [1 - a(x)], \text{ for } i = 4 \quad (8.18)$$

$$Sb_i(x) = \sum_{z=x_{\min}}^{x-1} b_i(z), \text{ } i=1, 2, 3, 4, 5+ \quad (8.19)$$

We recall that  $a(x)$  is the average share of the age interval  $[x, x+1)$  lived before giving birth to a child. We assume that all  $a(x)$  values are equal to 0.5 for any age  $x$  and birth order  $i$ .

In incomplete period fertility tables constructed using the parity weights  $w_i(x)$  generated by cumulating cohort fertility (Section 8.1.1) and covering the years  $t_0+30$  through  $t_0+39$  (see Figure 3.5),

all fertility table functions, except  $l_i(x)$ , are computed until the age  $\hat{x}_{\max} - 1$  where  $45 \leq \hat{x}_{\max} \leq 55+$ , while function  $l_i(x)$  is calculated also for the age  $\hat{x}_{\max}$ . The first complete period fertility table with all of the functions expanding over the entire age range (from 12 and younger until 55 and older) is generated for the year  $t_0+40$ .

#### **8.4 Summary indicators based on the period fertility table**

On the basis of period fertility tables, the following summary indicators are computed in the HFD: the parity- and age-adjusted total fertility rate (*PATFR*, also called “the summary index of period fertility controlling for age and parity”; see Rallu and Toulemon 1994, and footnote 12) and its parity-specific components ( $PATFR_i$ ), and table mean ages at birth (*TMAB*) and table mean ages at birth by birth order ( $TMAB_i$ ). Formulae (8.20a), (8.20b), (8.21a), and (8.21b) define their computation.

*The parity- and age-adjusted total fertility rate for all birth orders combined:*

$$PATFR = \frac{\sum_{x=x_{\min}}^{x_{\max}} b(x)}{10,000} \quad (8.20a)$$

*The parity- and age-adjusted total fertility rate by birth order  $i$ :*

$$PATFR_i = \frac{\sum_{x=x_{\min}}^{x_{\max}} b_i(x)}{10,000} \quad (8.20b)$$

*The table mean age at birth for all birth orders combined:*

$$TMAB = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot b(x)}{\sum_{x=x_{\min}}^{x_{\max}} b(x)} \quad (8.21a)$$

*The table mean age at birth by birth order  $i$ :*

$$TMAB_i = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot b_i(x)}{\sum_{x=x_{\min}}^{x_{\max}} b_i(x)} \quad (8.21b)$$

*PATFRs* and *TMABs* are calculated for the period cohorts that are observed from age  $x_{\min}=12$  until  $x_{\max}$  that ranges from 44 to 54 in the incomplete period fertility tables and equals 55+ in the complete fertility tables (see Figure 3.5 for Lexis regions covered in the period fertility tables).

Values  $\bar{x}$  in formulae (8.21a) and (8.20b) are the mean ages at birth within age intervals  $[x, x+1)$ . We assume that, within each such age interval, the mean age can be approximated as  $x + 0.5$ .

## Acknowledgments

We are sincerely grateful to Joshua Goldstein, former Director of the Max Planck Institute for Demographic Research, who launched and headed the Human Fertility Project (HFP) at MPIDR; and to Wolfgang Lutz, Director of the Vienna Institute of Demography, Austrian Academy of Sciences, who supported the HFD activities at the VID. We are also grateful to HFD Advisory Board members Jean-Paul Sardon, Laurent Toulemon, and John Wilmoth for their thoughtful comments and advice. We would like to thank Sigrid Gellers-Barkmann and Jörn Lübke (MPIDR) for their graphic and editing assistance. 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).

## References

- Bishop, Y., S. Fienberg, and P. Holland (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT University Press.
- Bongaarts, J. and G. Feeney (1998). "On the quantum and tempo of fertility". *Population and Development Review* 24 (2): 271-291.
- Bongaarts, J. and G. Feeney (2000). "On the quantum and tempo of fertility: Reply". *Population and Development Review* 26 (3): 560-564.
- Caselli, G. and J. Vallin (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.
- Calot, G. and J.-P. Sardon (2003). "Methodology for the calculation of Eurostat's demographic indicators. Detailed report by the European Demographic Observatory (EDO)". *Population and social conditions* 3/2003/F/no 26, 146 p.
- Fienberg, S. (1970): An iterative procedure for estimation in contingency tables, *The Annals of Mathematical Statistics*, Vol. 41, N. 3, 907-917.
- Frejka, T. and J.-P. Sardon (2006). "First birth trends in developed countries: persistent parenthood postponement." *Demographic Research* 15(6): 147-180
- Fritsch, F. N. and R.E. Carlson (1980). "Monotone Piecewise Cubic Interpolation". *SIAM Journal on Numerical Analysis* 17 (2): 238-246. SIAM.
- Goldstein J.R., T. Sobotka, A. Jasilioniene (2009). "The End of "Lowest-Low" Fertility?" *Population and Development Review* 35(4): 663-399.
- Kohler, H.-P., F.C. Billari, and J. A. Ortega (2002). "The emergence of lowest-low fertility in Europe during the 1990s". *Population and Development Review* 28 (4): 641-680.
- Kuczynski, R.R. (1932) [1982]. *Fertility and Reproduction*. Reprinted in 1982 in Berlin (GDR), Akademie-Verlag (*Beitrage zur Demographie*) No. 6..
- McNeil, Donald R., T. James Trussell, and John C. Turner (1977). "Spline interpolation of demographic data." *Demography* 14(2): 245-252.
- 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.
- Sobotka, T. (2003). "Tempo-quantum and period-cohort interplay in fertility changes in Europe. Evidence from the Czech Republic, Italy, the Netherlands and Sweden". *Demographic Research* 8(6): 151-213.
- Sobotka, T. (2004a). Is lowest-low fertility explained by the postponement of childbearing? *Population and Development Review* 30(2): 195-220.
- Sobotka, T. (2004b). *Postponement of childbearing and low fertility in Europe*. PhD Thesis, University of Groningen. Amsterdam: Dutch University Press.
- VID 2010: European demographic data sheet 2010. Vienna Institute of Demography and IIASA. ([http://www.oeaw.ac.at/vid/datasheet/download\\_2010.shtml](http://www.oeaw.ac.at/vid/datasheet/download_2010.shtml)).
- Wilmoth, J. R., K. Andreev, D. Jdanov, and D.A. Gleijeses with the assistance of C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, and P. Vachon (2005). "Methods Protocol for the Human Mortality Database.", Retrieved February 20, 2007 (<http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>).

## Appendix 1. Notations

<b>General</b>	
$x$	Age at childbearing
$x_{\min}$	Lowest age at childbearing considered in the analysis
$x_{\max}$	Highest age at childbearing considered in the analysis
$t$	Calendar year
$c$	Cohort
$i$	Parity and birth order
$i_{\max}$	Highest parity (birth order) used in the analysis
<b>Empirical data</b>	
$B(t)$	Total live births
$B(x, t, c), B_i(x, t, c)$	Number of live births (all birth orders combined and by birth order): $B(x, t), B(x, t, \circ); B_i(x, t), B_i(x, t, \circ)$ – for rectangular (age, year) $B(t, c), B(\circ, t, c); B_i(t, c), B_i(\circ, t, c)$ – for vertical parallelogram (year, cohort) $B(x, c), B(x, \circ, c); B_i(x, c), B_i(x, \circ, c)$ – for horizontal parallelogram (age, cohort)
$P(x, t)$	Female population size on 1 January
$E(x, t, c)$	Female population exposure: $E(x, t), E(x, t, \circ)$ – for rectangular $E(t, c), E(\circ, t, c)$ – for vertical parallelogram $E(x, c), E(x, \circ, c)$ – for horizontal parallelogram
$E_{\text{total}}(t)$	Total population exposure, including both men and women, for all ages combined
$f(x, t, c), f_i(x, t, c)$	Unconditional age-specific fertility rates (all birth orders combined and by birth order): $f(x, c, t), f(x, c, t)$ – for Lexis triangle $f(x, t), f(x, t, \circ); f_i(x, t), f_i(x, t, \circ)$ – for rectangular $f(t, c), f(\circ, t, c); f_i(t, c), f_i(\circ, t, c)$ – for vertical parallelogram $f(x, c), f(x, \circ, c); f_i(x, c), f_i(x, \circ, c)$ – for horizontal parallelogram
<b>Fertility table</b>	
$w_i(x)$	Relative distribution of female population exposure by parity (population weights): $\sum_i w_i(x) = 1$
$E_i(x)$	Female population exposure by parity: $E_i(x) = w_i(x) \cdot E(x)$
$m_i(x)$	Conditional age-specific fertility rates in age interval $[x, x+1)$
$q_i(x)$	Probability of having an $i$ th birth in age interval $[x, x+1)$
$l_i(x)$	Table population of parity $i$ at age $x$
$b_i(x)$	Table number of births of order $i$ in age interval $[x, x+1)$
$L_i(x)$	Table population exposure of women of parity $i$ within age interval $[x, x+1)$
$chi(x)$	Average number of children by exact age $x$ in the highest parity category $i_{\max+}$
$Sb_i(x)$	Cumulative (in respect to age) births of order $i$ by exact age $x$

$a_i(x)$	Mean duration (time) spent at parity $i$ within age interval $[x, x+1)$ by women progressing to parity $i$ during this age interval
<b>Summary measures</b>	
$CBR$	Crude birth rate for all birth orders combined
$CBR_i$	Crude birth rate by birth order
$TFR, TFR_i$	Period total fertility rate based on unconditional age-specific fertility rates for all birth orders combined $f(x)$ and by birth order $f_i(x)$
$adjTFR, adjTFR_i$	Tempo-adjusted total fertility rate for all birth orders combined and by birth order, using Bongaarts and Feeney's (1998) method
$MAB, MAB_i$	Period mean age at birth based on unconditional age-specific fertility rates for all birth orders combined $f(x)$ and by birth order $f_i(x)$
$sdMAB, sdMAB_i$	Standard deviation in period mean age at birth for all birth orders combined and by birth order
$CPFR(x), CPFR_i(x)$	Cumulative period fertility rate for all birth orders combined and by birth order by exact age $x$
$CCF, CCF_i$	Completed cohort fertility for all birth orders combined and by birth order
$CMAB, CMAB_i$	Cohort mean age at birth for all birth orders combined and by birth order
$sdCMAB, sdCMAB_i$	Standard deviation in cohort mean age at birth for all birth orders combined and by birth order
$CCFR, CCFR_i$	Cumulative cohort fertility rate for all birth orders combined and by birth order by exact age $x$
$PPR_{i-1,i}$	Cohort parity progression ratios (lifetime probability of transition from parity $i-1$ to parity $i$ )
$PATFR, PATFR_i$	Parity- and age-adjusted total fertility rate for all birth orders combined and by birth order. Rallu and Toulemon (1994) call it "the summary index of fertility controlling for age and parity". It is an alternative indicator to the period total fertility rate.
$TMAB, TMAB_i$	Table mean age at birth for all birth orders combined and by birth order based on the age distribution of the table number of births