FERTILITY TABLES IN THE HUMAN FERTILITY DATABASE: CONSTRUCTION AND ILLUSTRATIONS

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

First Draft: 23 December 2009

(This paper is still in progress. Please do not cite without authors' permission)

Abstract (short)

The Human Fertility Database (HFD) has been created following the example of the Human Mortality Database (HMD). Life table method applied in the HMD is well known, thoroughly described in the literature and widely used among demographers and other researchers. Meanwhile, fertility tables, which are featured in the HFD, are much less familiar and their methodology is not standardized. We explain in detail the procedure for building cohort and period fertility tables specific for age and parity. We show how each table function is computed and how selected summary indicators are derived. In addition, we provide illustrations based on the cohort and period fertility tables and indicators in the HFD, which demonstrate their practical usefulness and help to get a better insight into their functions and their interpretation. These standardized sets of fertility tables in the HFD aim to facilitate new fertility research and make a significant contribution to comparative fertility analysis.

1. Introduction

The Human Fertility Database (HFD) has been created following the example of the Human Mortality Database (HMD). Both projects share the same aim to provide free access to detailed and high-quality data and keep to the same principles of comparability, flexibility, accessibility, and reproducibility. Life table method applied in the HMD is well known, thoroughly described in the literature and widely used among demographers and other researchers. Meanwhile, fertility tables, which are featured in the HFD, are much less familiar and their methodology is not standardized. While in mortality research the main focus is on the indicator of the 'tempo' of mortality, namely, life expectancy by age, in fertility the main interest lies in the 'quantum' (childbearing intensity and total fertility). An indicator of fertility 'tempo' comparable to life expectancy in mortality tables – *mean number of years until the next birth* – appears to be of relatively little research interest. Another important distinction can be drawn with respect to the likelihood of an event: in mortality analysis, death is unavoidable and occurs only once; therefore, the overall quantum of mortality is always 1 and the ordinary mortality table is a decrement table, where the initial population defined at age 0 is depleted through mortality. In fertility analysis, birth is a repeatable event, but it is also 'avoidable' and many women (or men) may remain childless; the 'quantum' of first birth rates always stays by definition below 1. Therefore, fertility tables are more complex. Usually, multistate tables are used, where each state represents progression to one parity, until the highest parity considered.

In this paper we explain in detail the procedure for building cohort and period fertility tables specific for age and parity. We show how each table function is computed and how selected summary indicators are derived. In addition, we provide illustrations based on the period fertility tables in the HFD data, which demonstrate their practical usefulness and help to get a better insight into their functions and their interpretation

2. Parity-specific analysis of fertility: developments and applications

In comparative fertility research, fertility indicators based on age-specific fertility rates, especially the period total fertility rate (TFR), dominate the analysis as well as statistical reporting by the official statistical agencies. When birth order dimension is discussed, frequently, age-specific fertility rates are computed by age and birth order, but do not control for the parity distribution of the female population by age. That is, for each age x considered, age-specific fertility rate for birth order i, $f_i(x)$, relates births specified by age and birth order ($B_i(x)$ to all women of a given age $(E(x))$ irrespective of their parity status: $f_i(x) = B_i(x) / E(x)$. In agreement with common terminology, we call these fertility rates "incidence rates", "unconditional fertility rates" or "rates of the second type". These rates have some obvious advantages: First, they are less data demanding as female population by single years of age is routinely published by all the statistical offices in the developed countries, but annual time series of the female parity composition by age are usually unavailable. Second, incidence rates can be readily summed up to order-specific components of the TFR. Third, computing indicators controlling for age and parity dimensions (and/or duration dimension for the second and higher-order births) is more work-intensive and no widely established set of procedures or indicators exists. At the same time, indicators not controlling for the parity distribution of the female population have a number of drawbacks, especially in low-fertility countries where relatively small differences in parity-specific fertility among childless women and among women with one child may lead to substantial aggregate differences in fertility rates. Recently, Sobotka and Lutz (2009) have argued that the excessive reliance on the period TFR in contemporary fertility analysis, and policy-relevant discussions may lead to misinterpretations of period fertility trends and levels, to incorrect inferences about the presumed gap between fertility intentions and realised fertility and to an erroneous evaluation of family policy effects.

Over time, many scholars have warned against an excessive use of the period TFR and the prominence of fertility rates controlling for age only (Whelpton 1946, Ryder 1990, Ní Bhrolcháin 1992, Toulemon 1994, Ortega and Kohler 2002) and substantial body of contributions that discuss and use fertility rates controlling for age, parity, and/or duration since the previous birth has accumulated. As early as in 1946 Whelpton discussed fertility rates controlling in addition to age for parity, marriage, and sterility. He was the first one to point out some potentially obscure results in the period TFRs disaggregated by birth order, which can seemingly imply more than 100% of women having a first birth (when the first-order TFR surpasses 1). In the 1950s, Henry has pioneered parity-specific analysis of fertility within marriage, taking duration between marriage and first birth and interbirth intervals as main controlling factors (e.g., Henry 1953). Since the 1970s many researchers have computed fertility indexes and tables controlling for parity based on vital statistics (e.g., Park 1976 for the United States, Kojima and Rallu 1997 for Japan, Boleslawski 1993 for Poland, Barkalov and Dorbritz 1996 for Eastern Germany (former GDR)), Schoen 2003 for the United States, and Sobotka 2003 for four European countries), survey data (Rallu 1986 for France, Feeney and Yu 1987 for China, Ní Bhrolcháin 1987 for England and Wales, Rallu and Toulemon

1994 for France, Smallwood 2002 for England and Wales, and Barkalov 2005 for Russia), and detailed census data (Neels 2006 for Belgium, McDonald and Kippen 2007 for Australia). Besides that, fertility analysis based on individual data using intensity regression methods commonly applies hazard rates standardized for age, parity and/or duration (since these methods are outside the scope of our paper, we do not discuss them further). Fertility indicators controlling for age and parity have been first elaborated in detail Park (1976) and later by Rallu and Toulemon (1994).

Clearly, parity-specific models and tables of fertility analysis have been repeatedly used and discussed, but, so far, they have not become a mainstream methodology in fertility research. Therefore, differently from mortality tables, researchers interested in parity-specific fertility tables will not find established terminology, notations, textbooks, handbooks or, for that matter, standard indicators published by the statistical offices. The HFD aims to build on the existing literature and create internally logical, unified and standardised system of methodology, computation and indicators of age-parity-specific fertility and fertility tables that can be replicated and used in comparative analysis. We hope to set a standard for this type of analysis and thus also promote the wider use of these indicators.

The main principles and assumptions applied in the HFD to age and parity-specific fertility tables and indicators are listed here. Note that we refer here to women only because the HFD focuses on female fertility analysis, but all the principles described here relate equally to men.

- Childbirth is a repeatable event, but each woman can have only one child of each birth order (multiple births are assigned separate birth order each). In the HFD, fertility tables are constructed for birth orders up to an open-ended category 5+.[1]

- Fertility tables in the HFD are single-decrement tables; i.e., the only path through which women move out of one parity category is by giving birth.

- Fertility tables can be seen as multistate tables, where different parities represent different states in the fertility process. However, the movement between states is unidirectional, always from one parity category to the next higher one. In the HFD, the possibility of giving multiple births is disregarded and these births are counted separate transitions between two neighbouring parities.

- Tables for first births are only decrement tables, where, in analogy to mortality tables, there is an initial starting population defined at the beginning of reproductive age (age 12 in the HFD), when all women are supposed to be childless. The tables for the second and higher-order births are increment-decrement tables, where the exposure population of women of parity i at any reproductive age considered (ages 12-55 in the HFD) simultaneously increases through transitions from the lower parity i-1 and decreases through giving births of order $i+1$ (and thus 'transiting' to the parity category $i+1$).

---

[1] Following the established terminology, we use the term 'parity' when referring to the number of children born to a woman and 'birth order' when referring to the birth order of the child. Thus, a woman at parity 3 is exposed to give birth to a child of birth order 4 (and thus also make a transition to parity 4). In analogy, we refer to fertility indicators as 'parity-specific' when they restrict the denominator to the women of the parity at risk. We use the term 'fertility rates by birth order' when we refer to rates and indicators that do not control for the parity status of the female population (see also Multilingual demographic dictionary, IUSSP 1982, par. 634).

- The fertility tables in the HFD are constructed from the set of age- and parity-specific fertility rates, $m_i(x)$, defined for all birth orders (1 through 5+) and reproductive ages (12-55) covered in the HFD. These rates are computed for the average exposure population for the given year or age and thus better capture the effects of migration that may change the female population composition during the year. This contrasts with the main alternative approach, where age-parity birth probabilities are computed directly on the basis of births recorded in a calendar year and related to the initial exposure at the beginning of that year.

- Fertility tables for the highest parity category included in the HFD are increment tables based on fertility rates for birth order 5+ computed for the subset of women at parities 4+. This computation is based on an implicit assumption that age- and parity-specific fertility rates for women at parities 4 and higher remain constant with parity and that women at parities 4+ remain in the exposure population when they experience birth. Since only a small number of women in the developed countries progress to parities 4 or higher, model assumptions applied in the HFD should not have a large influence on the overall fertility indicators for all parities combined.

We are convinced that the standardized sets of fertility tables in the HFD will stimulate new fertility research and will make a significant contribution to comparative fertility analysis. Because the fertility tables methodology is not well established, we devote the next three sections of the paper to describing the methodology used in the HFD in detail for both cohort- and period- based fertility tables controlling for age and parity.

3. Initial data

The HFD is based on one and the same type of initial data: original, officially registered live birth counts by calendar year, age of the mother (and/or mother's year of birth, i.e. birth cohort) and, whenever possible, biological birth order. These data, together with the exposure population estimates mostly extracted from the HMD, selected population censuses and register data, are processed using a uniform set of methods. The HFD methodology allows transforming the input data that vary in many respects across countries and time into uniform Lexis data.

The raw data on births are often classified only by calendar year and age of the mother or by calendar year and birth cohort of the mother. For some countries and calendar years, 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 unknown age of the mother or 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 these raw data into a uniform universe of data classified by single years of age ranging from age ≤12 to 55+, by single-year birth cohorts, and (whenever possible) by birth orders varying from 1 to 5+. Births with unknown age of the mother are distributed proportionally according to the birth data where age of the mother is specified. Within each age, births with unknown birth order are distributed proportionally across known birth orders. Aggregated age groups are additionally split into single-year ages. Birth orders higher than five are combined into birth order 5+.

Adjustments are also applied to population exposures and data on the age- and parity-specific distribution of women. In the Lexis data, population exposures are classified by calendar year-age-birth cohort cells. Age varies from age 12 to age 55. Age-parity distributions of women are given by one year age groups as of the 1st of January. Parity varies from 0 to 4+.

The Lexis data are of paramount importance since they form the basis for all further calculations. Having data by Lexis triangles makes it possible to compute fertility rates and other fertility indicators in any configuration desired – be it *horizontal parallelogram*, *vertical parallelogram or square* (*rectangle*).

4. Cohort fertility table

Cohort fertility tables are increment-decrement life tables, which model the process of childbearing in female cohorts by age and parity. In principle, they describe a two-dimensional cohort progression toward older age and higher parities. Women of the cohort of interest are moving from parity zero (i.e., from being childless) to parity one, from parity one to parity two, and to subsequent parities, by giving births of the corresponding birth orders.

In general, the construction of cohort fertility tables is much less sophisticated than of period fertility tables. For each cohort, the table functions are computed from the schedule of age- and order-specific fertility rates (horizontal parallelograms) as the major input data. The distribution of births by age of the mother and birth order in the table and the parity distribution of the table population of females correspond to the observed fertility trajectories of cohorts analyzed.

The crucial matter in constructing cohort fertility tables is data availability. Cohort fertility tables can be built only if sufficiently long time series of period data on fertility by birth order are available. The data should allow observing female cohorts from the beginning, and if possible, until the end of their reproductive life span. In the HFD, cohort fertility tables are constructed for cohorts that are observed from age 15 or a younger age until age 25 or an older age. Cohorts satisfying these conditions belong to the colored diagonal region in Figure 4.1. The youngest and the oldest cohorts for which the cohort fertility tables are constructed are T-25 and t_0-15, respectively.

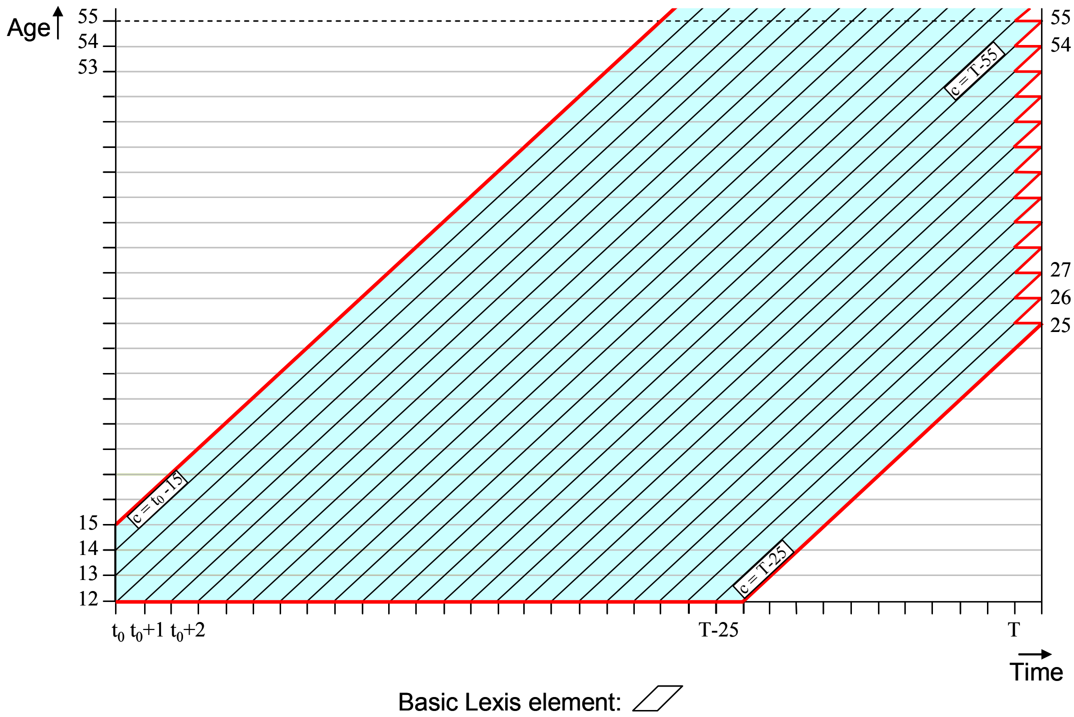


Figure 4.1. Lexis region for the cohort fertility tables based on horizontal parallelograms (cohort-age cells)

4.1 Construction of the cohort fertility table

For each cohort c, all the functions of the cohort fertility table are computed from the schedule of unconditional age-specific fertility rates by birth order $f_i(x,c)$ that are computed according to Formula 4.1.

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

$$f_i(x,c) = \frac{B_i(x,t,t-x) + B_i(x,t+1,t-x)}{E(x,c)} \qquad (4.1)$$

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

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

$$b_i(x) = 10{,}000 \cdot f_i(x,c) \qquad (4.2)$$

This formula considers birth as a repeatable event. This implies that total number of individuals in the life table (say 10,000) remains the same at any age. The individuals, however, move 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:

$l_0(x_{\min}) = 10{,}000$ (radix of the cohort) (4.3)

$l_i(x_{\min}) = 0$, for i = 1, 2, 3, 4 (4.4)

$l_i(x) = l_i(x-1) - b_{i+1}(x-1)$, for $i = 0$ (4.5)

$l_i(x) = l_i(x-1) + b_i(x-1) - b_{i+1}(x-1)$, for i = 1, 2, 3 (4.6)

$l_{i+}(x) = l_{i+}(x-1) + b_i(x-1)$ [2], for $i = 4$ (4.7)

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:

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) - a(x) \cdot b_i(x)}, \text{ for } i = 1 \quad (4.8)$$

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) - a(x) \cdot b_i(x) + (1 - a(x)) \cdot b_{i-1}(x)}, \text{ for } i = 2, 3, 4 \quad (4.9)$$

$$m_i(x) = \frac{b_i(x)}{l_{i-1}(x) + (1 - a(x)) \cdot b_{i-1}(x)}, \text{ for } i = 5+ \quad (4.10)$$

Conditional probability of giving an i-th birth at age x for a woman of parity i-1 is expressed:

$$q_i(x) = \frac{b_i(x)}{l_{i-1}(x)}, \; i\text{=1, 2, 3, 4, 5+} \quad (4.11)$$

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

$$Sb_i(x) = \sum_{z=x_{\min}}^{x-1} b_i(z), \; i\text{=1, 2, 3, 4, 5+} \quad (4.12)$$

Average number of children born by age x in the highest parity category 5+ is expressed:

$$chi(x) = \frac{4 \cdot l_4(x) + Sb_{5+}(x)}{l_4(x)} \quad (4.13)$$

In addition, we can easily compute lifetime probabilities of having a birth of a given order after a certain age, although in the HFD it is not provided:

$$Q_i(x) = 1 - \frac{l_{i-1}(x_{\max})}{l_{i-1}(x)} \quad (4.14)$$

At very young ages, number of births observed at higher birth orders in a population is very low and strongly fluctuating. This may lead to negative values of $l_{i-1}(x)$ or to $q_i(x)$ exceeding 1. In such cases, the computed values are replaced by zero for $l_{i-1}(x)$, $b_i(x)$, $q_i(x)$, and $m_i(x)$.

[2] i+ stands for women at parities i and higher.

4.2 Summary indicators based on the cohort fertility table

In the HFD, cohort summary indicators are computed on the basis of unconditional age-specific fertility rates by birth order $f_i(x,c)$. However, the same results would be obtained using cohort fertility functions for the computation as in both cases the data pertain to and reflect fertility of actual cohorts of women.

On the basis of cohort fertility tables, the following summary indicators can be computed: the cohort total fertility rate (*CTFR*) and its order-specific components ($CTFR_i$); cohort mean ages at birth (*CMAB*) and cohort mean ages at birth by birth order ($TMAB_i$); and cohort parity progression ratios (*CPPR*s). Formulae 4.15a, 4.15b, 4.16a, 4.16b, 4.17a and 4.17b define the computation of these cohort summary indicators. It is noteworthy that cohort summary indicators are to be computed for cohorts of women that have already completed their childbearing. In the HFD, they are computed over the range of ages from 15 or younger through age 50 or older. But since many cohorts that are approaching the end of their reproductive span have practically completed their fertility histories, the HFD also displays cohort total fertility rates and mean ages at childbearing achieved by age 40; these indicators should serve researchers for estimating or projecting completed fertility rates of those cohorts.

The completed cohort total fertility rates for all birth orders combined and by birth order:

$$CTFR = \frac{\sum_{x_{\min}}^{x_{\max}} b(x)}{10{,}000} \qquad (4.15a)$$

$$CTFR_i = \frac{\sum_{x_{\min}}^{x_{\max}} b_i(x)}{10{,}000}, \quad i=1,2,3,4,5+ \qquad (4.15b)$$

The cohort mean ages at birth for all birth orders combined and by birth order:

$$CMAB = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot b(x)}{\sum_{x=x_{\min}}^{x_{\max}} b(x)} \qquad (4.16b)$$

$$CMAB_i = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot b_i(x)}{\sum_{x=x_{\min}}^{x_{\max}} b_i(x)}, \qquad i=1,2,3,4,5+ \qquad (4.16b)$$

Values $\bar{x}$ in formulae (4.16a) and (4.16b) 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.

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

The cohort parity progression ratios:

$$PPR_{0,1}(c) = CTFR_1(c) \qquad (4.17a)$$

$$PPR_{i-1,i}(c) = \frac{CTFR_i(c)}{CTFR_{i-1}(c)}, \text{ for } i>1 \qquad (4.17b)$$

Note that the highest parity-progression ratio computed 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 do not allow separating fifth births as a special category.

5. Period fertility table

Many functions in period fertility tables are identical to those in cohort fertility tables, and their construction is based on comparable formulas. In analogy to cohort fertility tables, period fertility tables are increment-decrement life tables, which model the process of childbearing in hypothetical (synthetic) cohorts of women specified by age and parity. In other words, they give a period snapshot of fertility of many female birth cohorts and do not correspond to childbearing history of any real cohort.

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 a thorough analysis of fertility level and timing.

The period fertility tables in the HFD are based on the conditional age- and order-specific fertility rates by Lexis squares (year-age cells). In order to compute these rates, female age- and parity-specific population exposures must be estimated. These distributions are obtained either from cohort fertility tables, "golden" censuses that provide the parity distribution in one base year or directly from population censuses or registers (see Section 5.1 below). In the latter case, the fertility tables are called census- or register-based fertility tables.

We assume that age 45 is an age by which fertility of cohorts is nearly completed and women's parities are very close to their final values – therefore, the HFD period fertility tables are constructed for all the years when cohort parity distribution can be observed for ages 15 or younger through 45 or older. In Figure 5.1, 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).

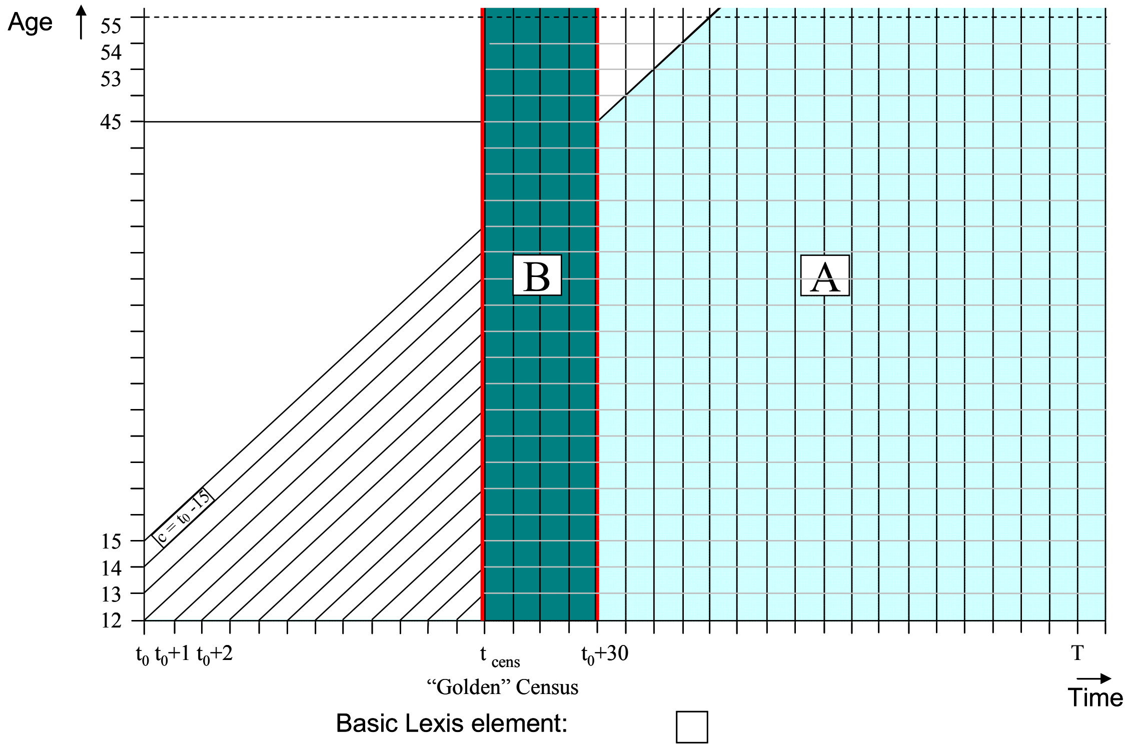


Figure 5.1 Lexis regions for the period fertility tables based on Lexis squares (year-age cells)

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 ("golden" census). If the census takes place in the year t_{cens}, then the period fertility tables can be constructed for an extended region A+B in Figure 5.1. In the year t_{cens}, parity- and age-specific population exposures are known, and they are annually updated until the year t_0+30 using population exposures of respective cohorts (region B in Figure 5.1). Starting from the year t_0+30, the period fertility tables are based on parity- and age-specific population exposures obtained purely by cumulating fertility of cohorts over their reproductive ages (region A in Figure 5.1).

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. 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 5.2).

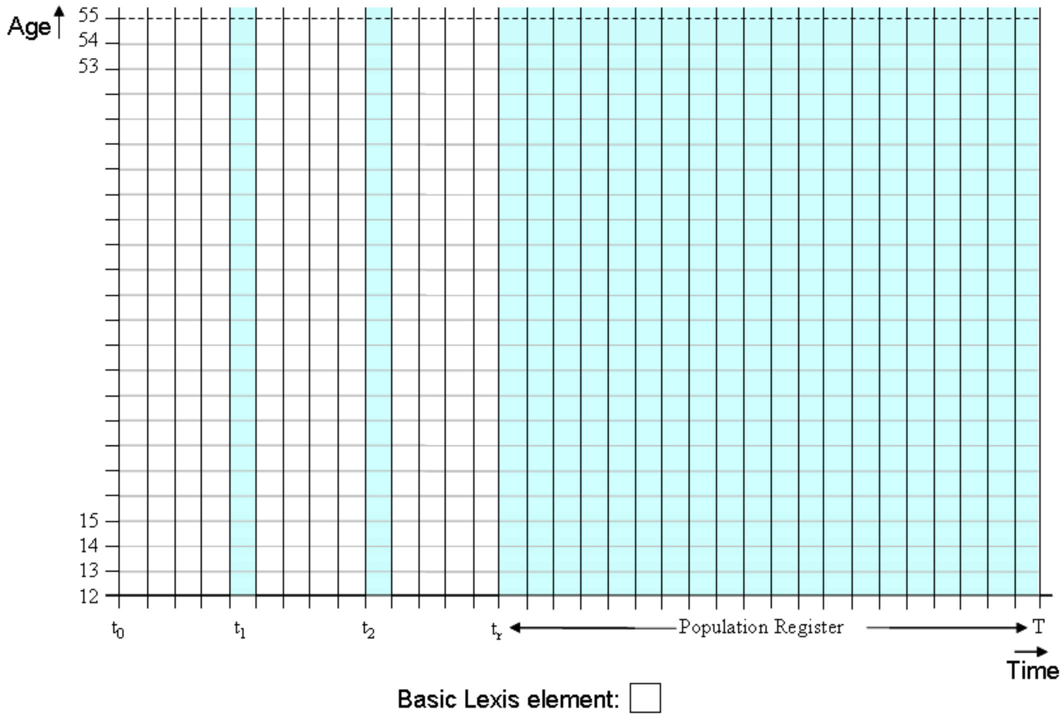


Figure 5.2. Lexis regions for the census- or register-based period fertility tables based on Lexis squares (year-age cells)

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.

5.1 Parity-specific population exposure

Annual time series of female population exposure by age and parity, 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 $w_i(x,t)$:

$$E_{i-1}(x,t) = w_{i-1}(x,t) \cdot E(x,t) \tag{5.1}$$

As noted above, depending on data availability, different approaches are used for obtaining the estimates of $w_i(x,t)$. These approaches are discussed below according to the hierarchy of their application.

5.1.1 Cumulating cohort fertility rates over long periods of time

For countries that have a sufficiently long time series of period data on births by age of the mother and birth order, annual estimates of the age-parity distribution $w_{i-1}(x)$ are reconstructed from these data by cumulating fertility of cohorts over their reproductive age span:

$$w_0(x_{\min},t) = 1, \tag{5.2}$$

$$w_i(x,t) = \frac{l_i(x,t-x) + l_i(x+1,t-x-1)}{2 \cdot l_0(x_{\min}, t-x_{\min})}, \qquad (5.3)$$

$$w_1(x_{\max}, t) = \frac{l_i(x_{\max}, t-x_{\max})}{l_0(x_{\min}, t-x_{\min})}, \qquad (5.4)$$

where $l_i(x)$ is the table cohort size specific for age and parity, derived from the cohort fertility tables (Figure 5.3). 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 the age 45. It covers the Lexis region A and extends over the period t_0+30 to T in Figure 5.1.

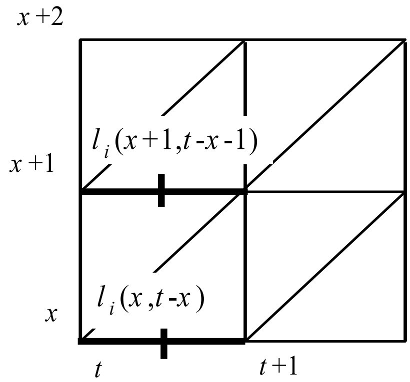


Figure 5.3. Estimation of $w_i(x)$ using the life table cohort size by age and parity, $l_i(x)$ from the cohort fertility tables

5.1.2 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) can be 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. In exceptional cases, when no other method for deriving the initial age-parity distribution can be employed, survey data can be considered. Such a survey must cover a large and representative sample of at least 1% of the female population of reproductive age.

Use of a "golden" census enables to construct the period fertility tables for an extended period t_{cens} to T instead of the period t_0+30 to T (see Figure 5.1). The calculation of the census-based age-parity distribution involves cohorts born in the years t_{cens}-x_{max} through t_{cens}-12. For these cohorts, the cohort fertility tables (not posted in the HFD) are left-censored and start from certain non-null values of $l_i(x)$ that are computed using the census weights $w_i(x, t_{cens}-x)$:

$$l_i(x, t_{cens} - x) = 10{,}000 \cdot w_i(x, t_{cens} - x), \qquad (5.5)$$

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 5.4):

$$w_i(x, t_{cens} - x) = \frac{w_i(x-1, t_{cens}^{Jan1}) + w_i(x, t_{cens} + 1^{Jan1})}{2}. \qquad (5.6)$$

The values of $l_i(x)$ are estimated for all cohorts observable in the Lexis region B in Figure 5.1, and the mid-year age-parity distribution $w_i(x)$ is calculated as expressed in Formulae 5.2, 5.3

and 5.4 and shown in Figure 5.3. Based on these $w_i(x)$ estimates, combining census-based and cohort-based values, the period fertility tables are build for the period t_{cens} to t_0+29. Starting from the year t_0+30 (region A in Figure 5.1), the age-parity distribution is obtained and updated for consequent years *purely* by cumulating cohort population exposures, i.e. exactly in the way explained in Section 5.1.1.

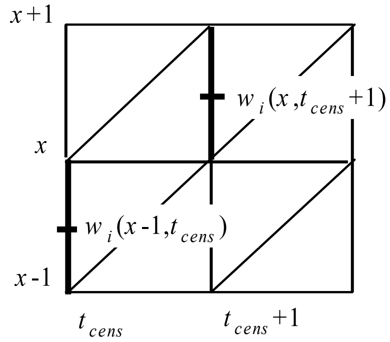


Figure 5.4. Estimation of the census weights used to produce the left-censored cohort fertility tables

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 5.6). The HFD uses procedures that allow reconstructing the data for the age-parity distribution on January 1st of the census year and of subsequent years; they are described in detail in the HFD Methods Protocol (see Jasilioniene et al. 2009).

5.1.3 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. In census-based tables the computation of the mid-year value of $w_i(x)$ in the census year is identical to that applied in the case of the "golden"census, except that the period fertility table is constructed for the census year only. In the case of register data, the computation procedure (see Formula 5.7) is simpler because annual series of the age-parity distribution for the 1st of January are readily available:

$$w_{i-1}(x,t) = \frac{w_i(x,t^{Jan1}) + w_i(x,t+1^{Jan1})}{2} \qquad (5.7)$$

5.2 Construction of the period fertility 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 at least partly to account for the effects of mortality and migration on population exposure $E_{i-1}(x,t)$ over the year (period) t.

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

$$m_i(x,t) = \frac{B_i(x,t,t-x-1) + B_i(x,t,t-x)}{E_{i-1}(x,t)} = \frac{B_i(x,t)}{E_{i-1}(x,t)} \qquad (5.8)$$

To eliminate huge fluctuations in $m_i(x)$ at lowest- and highest-childbearing ages, attributable to very low observed numbers of births, these rates are computed in the HFD only for age and parity combinations where more than 5 births are observed: $E_{i-1}>5$; otherwise, the values of E_{i-1} are replaced by zero 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 meet 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 fertility rates $m_i(x)$ 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.

Functions (columns) of the HFD period life tables are as follows: *Year*, *x*, $w_{i-1}(x)$, $m_i(x)$, $q_i(x)$, $l_{i-1}(x)$, $b_i(x)$, $L_{i-1}(x)$, $Sb_i(x)$ (see Appendix 1 for the notation). These functions are analogous to those in the cohort life table, described in Section 4. They are computed by using the following sequence of formulae:

$$q_i(x) = \frac{m_i(x)}{1 + [1 - a(x)] \cdot m_i(x)} \qquad (5.9)$$

$$l_0(x_{\min}) = 10{,}000 \text{ (the radix)} \qquad (5.10)$$

$$l_i(x_{\min}) = 0\text{, for } i = 1, 2, 3, 4 \qquad (5.11)$$

$$l_i(x) = l_i(x-1) \cdot [1 - q_{i+1}(x-1)]\text{, for } i = 0 \qquad (5.12)$$

$$l_i(x) = l_i(x-1) - b_{i+1}(x-1) + L_{i-1}(x-1) \cdot m_i(x-1)\text{, for } i = 1, 2, 3 \qquad (5.13)$$

$$l_{i+}(x) = l_{i+}(x-1) + L_{i-1}(x-1) \cdot m_i(x-1)^3\text{, for } i = 4 \qquad (5.14)$$

$$b_i(x) = L_{i-1}(x) \cdot m_i(x) \qquad (5.15)$$

$$L_i(x) = l_i(x) - l_i(x) \cdot q_{i+1}(x) \cdot [1 - a(x)]\text{, for } i = 0 \qquad (5.16)$$

$$L_i(x) = l_i(x) + l_{i-1}(x) \cdot q_i(x) \cdot [1 - a(x)] - l_i(x) \cdot q_{i+1}(x) \cdot [1 - a(x)]\text{, for } i = 1, 2, 3 \qquad (5.17)$$

$$L_{i+}(x) = l_{i+}(x) + l_{i-1}(x) \cdot q_i(x) \cdot [1 - a(x)]\text{, for } i = 4 \qquad (5.18)$$

$$Sb_i(x) = \sum_{x_{\min}}^{x} b_i(x) \qquad (5.19)$$

As in the case of the cohort fertility table, lifetime probabilities of having a birth of a given order after a certain age are not included in the HFD, but their computation is very simple:

$$Q_i(x) = 1 - \frac{l_{i-1}(x_{\max})}{l_{i-1}(x)} \qquad (5.20)$$

---

[3] Recall that *i*+ stands for women at parities *i* and higher.

Note 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.

5.3 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 summary index of period fertility controlling for age and parity (termed *PATFR* by Rallu and Toulemon (1994)) and its parity-specific components ($PATFR_i$); table mean ages at birth (*TMAB*) and table mean ages at birth by birth order ($TMAB_i$). Formulae 5.20a, 5.20b, 5.21a and 5.21b define their computation.

Summary index of period fertility controlling for age and parity for all birth orders combined:

$$PATFR = \frac{\sum_{x_{\min}}^{x_{\max}} b(x)}{10{,}000} \quad \text{, where } x_{\min}\text{=12 or younger, } x_{\max}\text{=55+} \qquad (5.20a)$$

Summary index of period fertility controlling for age and parity by birth order:

$$PATFR_i = \frac{\sum_{x_{\min}}^{x_{\max}} b_i(x)}{10{,}000}, \quad i\text{=1, 2, 3, 4, 5+} \qquad (5.20b)$$

Table mean ages at birth for all birth orders combined and by birth order:

$$TMAB = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot b(x)}{\sum_{x=x_{\min}}^{x_{\max}} b(x)} \qquad (5.21a)$$

$$TMAB_i = \frac{\sum_{x=x_{\min}}^{x_{\max}} \bar{x} \cdot b_i(x)}{\sum_{x=x_{\min}}^{x_{\max}} b_i(x)} \qquad (5.21b)$$

Values $\bar{x}$ in formulae (5.21a) and (5.21b) 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 $x+0.5$.

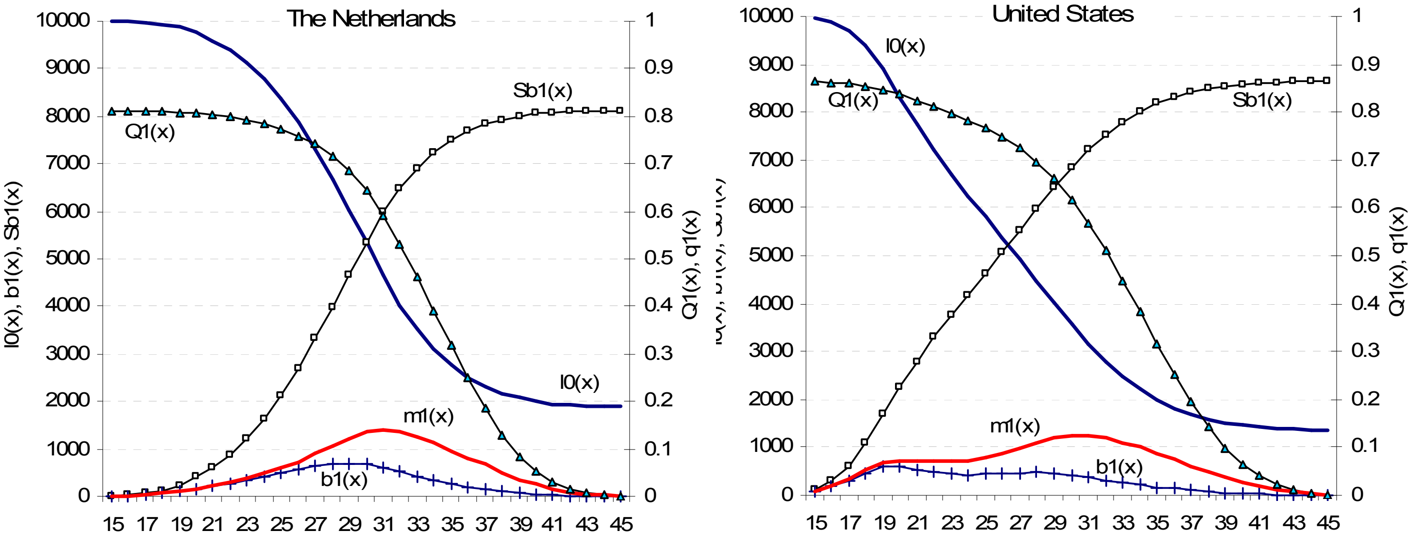
6. Graphical illustrations of different fertility table indicators

This section aims to provide practical illustrations of the use of various fertility indicators in the HFD fertility tables and demonstrate both the richness of data and a number of research issues that can be addressed with these indicators. First, to give a first detailed look at the data, we show various period indicators for first births. Then we provide illustrations based on cohort fertility indicators. Finally, we discuss selected issues that can be analysed with period fertility tables.

6.1 Period fertility: functions in the table for first births

Fertility tables provided in the HFD contain numerous indicators pertaining to parity-specific fertility levels, schedules and distributions by age. A look at the period indicators for first births in the Netherlands in 2007 and the United States in 2005 gives a first glimpse at age-specific indicators contained in the HFD and illustrate the differences in first birth patterns in these two countries (Figure 6.1). Inspecting conditional age-specific first birth rates, $m_1(x)$, a contrast between smooth and symmetrical curve for the Netherlands, peaking at age 31 and an asymmetric curve for the United States, with a local maximum at age 20, followed by a plateau and a main maximum at ages 30-31, can be observed. The considerably younger schedule of first birth rates in the U.S. is also apparent in the age distribution of table first births, $b_1(x)$, which peaks at age 19 and contrasts strongly with the Dutch pattern of $b_1(x)$, which peaks among women who are ten years older. Furthermore, the U.S. first birth pattern can also be illustrated by a rapid rise in the cumulative table first birth function, $Sb_1(x)$, which shows that one half of American women become mothers before reaching age 27. United States has overall higher first birth rates than the Netherlands, and according to the 2005 fertility tables, fewer than 14% of the initial cohort of 10,000 childless women would remain childless at age 45, whereas the corresponding figure for the Netherlands is 19%. (curve $l_0(x)$, due to very low first birth rates after age 45 we do not show results for ages 46-55). Finally, using the HFD data, a cumulated lifetime probability of having first birth for women childless at a given age, $Q_1(x)$, can be computed by subtracting the share of women who will stay at a given parity until the end of their reproductive life (defined as age 55) from 1: $Q_1(x) = 1 - l_0(x_{max}=55) / l_0(x)$. This function is particularly useful for depicting the cumulated chances at higher reproductive ages of ever having a(nother) child. In this respect, the Dutch and the U.S. first birth patterns are very similar after age 30: 32% of American as well as Dutch women still childless at age 35 would give birth to a child later in life if the recently observed first birth rates remained constant.

Figure 6.1. Selected functions from the period fertility table for first births in the Netherlands (2007) and the Unites States (2005).

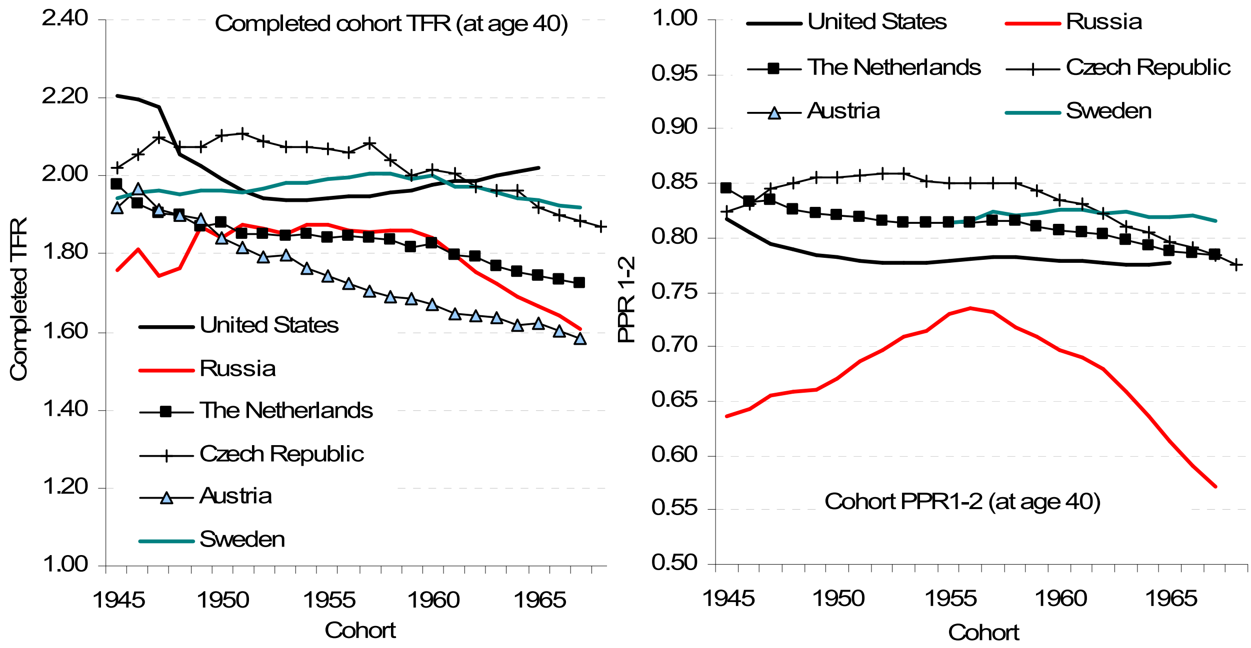


6.2 *Cohort fertility: quantum and timing*

Example 1: Completed fertility and progression rate to second birth

Billari and Kohler (2002) have suggested that European countries with 'lowest-low' fertility level are characterized by a low progression to second child rather than by high childlessness rates. In the future, the HFD data will make it possible to test such hypotheses and generate new ones. With a limited number of countries available at present, such hypotheses cannot be examined yet. Nevertheless, we provide these data as an illustration. Figure 6.2 (left panel) shows cumulated cohort fertility rates at age 40 among women born in 1945-68 in Austria, Czech Republic, the Netherlands, Russia, Sweden and the United States. Very small portion of fertility rates took place after that age, therefore, these data can be considered as an approximation of completed cohort fertility and they are also listed in the HFD under completed fertility rates. Right panel of Figure 6.2 shows second birth parity progression ratio in these cohorts (data for Austria are not available in the HFD). At a first glance, a strong relationship between falling completed fertility and declining second birth progression rates exists only in Russia among women born after 1960, who reached comparatively very low progression rate to second birth, falling below 55% among the youngest cohorts. A gradual decline in completed fertility in the Czech Republic and the Netherlands is also partly mirrored in a slow decline in second birth progression rates. In contrast, the U.S. women born in 1960-65 had an increasing completed TFR that reached higher level than in the other five countries, but stagnating and not particularly high levels of second birth rates.

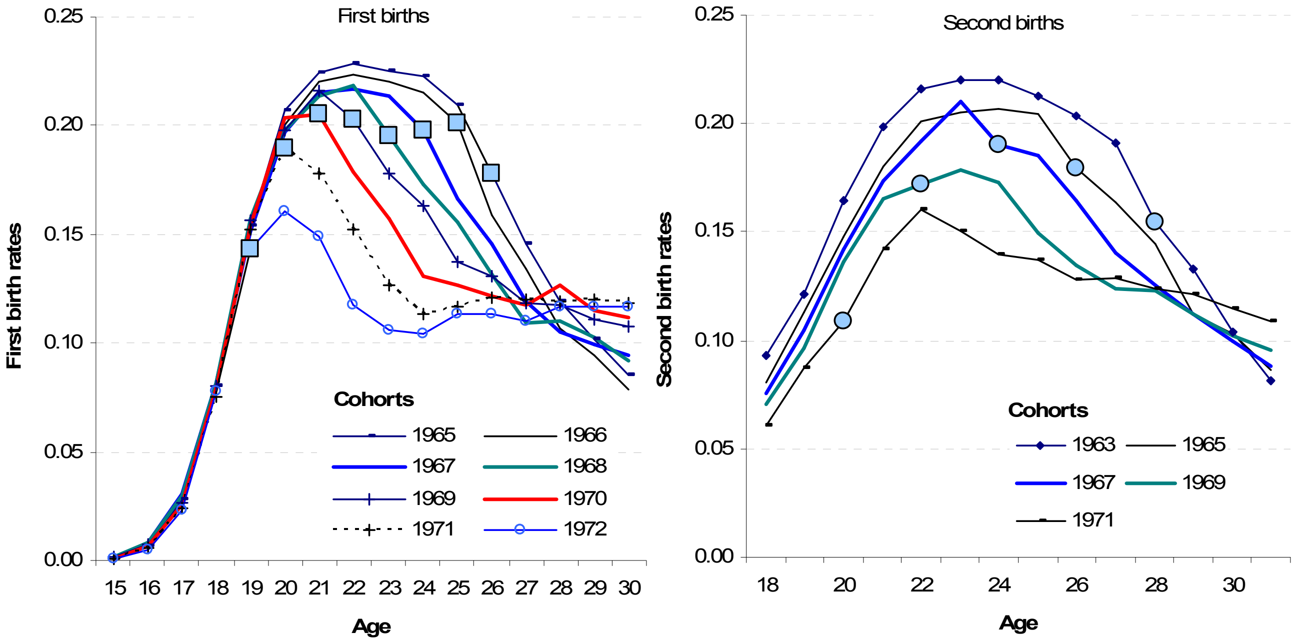
Figure 6.2. Cumulated cohort fertility rate at age 40 and second birth parity progression ratio; women born in 1945-1968



Example 2: Political regime change and the transformation in first and second birth trajectories

The HFD data also allow a detailed analysis of fertility shifts in individual countries. Sobotka et al. (2008) discuss rapid changes in period and cohort fertility patterns in the Czech Republic after the collapse of the state socialist system in 1989. The HFD data are well suited for tracing the shifts in cohort fertility related to such societal transformations. Because fertility pattern in the Czech Republic was characterized by early childbearing and a strong orientation towards a two-child family norm, one can expect that especially the first and second birth trajectories among relatively young women aged 18-28 were most affected by the political regime change. Figure 6.3 plots first and second parity-specific fertility rates [$m_1(x)$ and $m_2(x)$] at ages through 30 for selected cohorts whose childbearing trajectory was presumably most affected. Due to the time elapsing between conception and giving birth as well as some period needed for the young people to adjust their reproductive behaviour to the new social and political conditions, we assume that first changes can be observed in the year 1991, e.g., more than one year after the start of the political transformation in November 1989. In Figure 6.3 (left panel), depicting first birth rates among the childless women by age and birth cohort, age reached in 1991 is marked by an enlarged square. For the cohorts born in 1970 and older, i.e., those above age 20 in 1991, this was indeed the first year when their first birth rates diverged quite sharply from the previous trajectory and were set on a considerably lower level than the first birth rates among each of the previous cohorts. Only the youngest cohorts shown, born in 1971 and 1972 strayed off their initial trajectory by one year (1971) and two years (1972) later, respectively. In the case of second birth rates among women at parity 1 (right panel of Figure 6.3) we observe a similar effect of the larger-than-expected fall in second birth rates in 1991 (marked by large blue circles) except for the youngest women born after 1970 (to keep the figure relatively easy to read, only every second cohort is shown). However, the transformation of second birth schedule was not as pronounced as for the first births, perhaps because women progressing rather rapidly to the first birth have become increasingly select group.

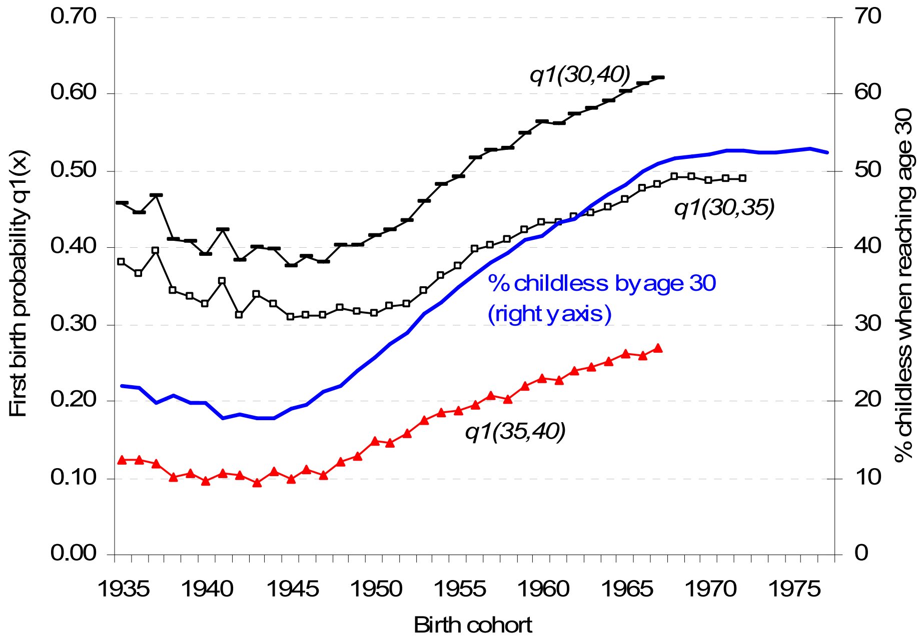
Figure 6.3. Parity-specific first and second birth rates *($m_i(x)$* in the Czech Republic among the cohorts born in 1963-72. Enlarged markers refer to age reached in 1991.



Example 3: Recuperation of 'delayed' first births

First births have been postponed to ever later ages across the whole developed world (Kohler et al. 2002, Sobotka 2004). A pertinent research question that emerged in relationship to this trend is what portion of the presumably delayed births eventually took place at later reproductive ages. This issue has been frequently addressed in a cohort perspective, especially in the work of Frejka and Sardon (2004, 2005, 2007). The HFD offers numerous ways how to analyse this process. One of them is looking to what extend an increase in cohort childlessness at younger reproductive ages corresponds with the rising probabilities of giving birth at a later age. If indeed most of the rise in childlessness at younger ages is due to fertility postponement rather than due to declining cohort first birth rates, these two trends should be closely related. We analyse them in Figure 6.4 for the Netherlands, which has experienced relatively long period of first birth postponement, initiated by the women born around 1945. Corresponding to that trend, the figure shows a massive rise in the percentage of women who are childless when reaching age 30, from fewer than 20% for the cohorts born in the mid-1940s to more than 50% for the 1970s cohorts (right y axis). We compare these data with conditional cohort first birth probabilities above age 30, computed for three broader age ranges: 30-34 ($q_1(30,35)$, 35-39 ($q_1(35,40)$), and, combining the two, for the range 30-40 ($q_1(30,40)$). These probabilities, $q_1(x, x+a)$, express likelihood that a woman childless at age x will give birth to a child before reaching age $x+a$. They can be computed using fertility table function $l_0(x)$ as follows: $q_1(x, x+a) = [l_0(x) - l_0(x+a)] / l_0(x)$. The comparison shows a close relationship between childlessness at age 30 and the likelihood of giving first birth after that age, even at ages above 35. Especially when a broader age range, 30-39, is considered, the two curves depict remarkably similar trends, indicating that rising first birth rates at ages above 30 have been closely linked to the previous first birth postponement among these cohorts of Dutch women.

Figure 6.4. Percent of women childless by age 30 and conditional probabilities of having first birth at ages 30-34, 35-39 and 30-39. The Netherlands, cohorts 1935-1977.



6.3 Period fertility: timing and quantum

[This part to be drafted]

References

[This part is to be completed later]

1. Jasilioniene, A., D.A. Jdanov, T. Sobotka, E.M. Andreev, K. Zeman and V.M. Shkolnikov, J.R. Goldstein, D. Philipov, and G. Rodriguez (2009). Methods Protocol for the Human Fertility Database. Rostock, MPIDR, 51 p. (http://www.humanfertility.org/Docs/methods.pdf)
2. Rallu, J.-L. and Toulemon, L. (1994). Period Fertility Measures: The Construction of Different Indices and their Application to France, 1946-89. *Population: An English Selection*, Vol. 6, pp. 59-63.
3. Sobotka, T. (2004). Postponement of Childbearing and Low Fertility in Europe. Doctoral thesis, University of Groningen. Amsterdam: Dutch University Press, 298 p.

Acknowledgments

We would like to thank Sigrid Gellers-Barkmann and Jörn Lübke (MPIDR) for their graphic assistance.

Appendix 1. Notations

	General
x	Age at childbearing
$x_{\min}$	Lowest age at childbearing considered in the analysis
$x_{\max}$	Highest age at childbearing
t	Calendar year
c	Cohort
i	Parity and birth order
$i_{\max}$	Highest parity (birth order) used in the analysis
	Empirical data
$B_i(x,t,c)$	Number of live births of order i
	$B_i(x,t)$, $B_i(x,t,\circ)$ – for rectangular (age, year)
	$B_i(t,c)$, $B_i(\circ,t,c)$ – for vertical parallelogram (year, cohort)
	$B_i(x,c)$, $B_i(x,\circ,c)$ – for horizontal parallelogram (age, cohort)
$P(x,t)$	Population size on January 1
$E(x,t,c)$	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
$f_i(x,t,c)$	Unconditional age-specific fertility rates
	$f_i(x,t)$, $f_i(x,t,\circ)$ – for rectangular
	$f_i(t,c)$, $f_i(\circ,t,c)$ – for vertical parallelogram
	$f_i(x,c)$, $f_i(x,\circ,c)$ – for horizontal parallelogram
	Fertility table
$w_i(x)$	Relative distribution of female population exposure by parity (population weights); $\sum_i w_i(x)=1$
$m_i(x)$	Conditional age-specific fertility rates in age interval [x, x+1)
$q_i(x)$	Probability of having an ith 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 age x in the highest parity category i_{max+}
$Sb_i(x)$	Cumulative (in respect to age) births of order i by 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
	Summary measures
$PPR_{i-1,i}$	Parity progression ratio (lifetime probability of transition from parity i-1 to parity i)
TFR_i	Total fertility rate for birth order i based on unconditional (non-exposure) age-specific fertility rates $f_i(x)$
TFR	Total fertility rate based on unconditional age-specific fertility rates $f(x)$
$CTFR$	Completed cohort fertility rate
$PATFR_i$	Period fertility index for birth order i based on conditional age-specific fertility rates or birth probabilities, derived from the fertility table. Rallu and Toulemon

	(1994) call it the summary index of fertility controlling for age and parity. It is an alternative indicator to the period TFR_i
PATFR	Period fertility index of total fertility based on conditional age-specific fertility rates or birth probabilities
$TMAB_i$	Table mean age at childbearing for birth order *i* based on the age distribution of the table number of births of order *i*, $b_i(x)$
MAB_i	Mean age at childbearing for birth order *i* based on unconditional age-specific fertility rates $f_i(x)$
MAB	Mean age at childbearing for all birth orders based on unconditional age-specific fertility rates $f(x)$