

HUMAN FERTILITY DATABASE DOCUMENTATION: POLAND

Krzysztof Tymicki

Warsaw School of Economics, Demography Unit

E-mail: krzysztof.tymicki@sggw.waw.pl

Kryštof Zeman

Vienna Institute of Demography, Austrian Academy of Sciences

Email: krystof.zeman@oeaw.ac.at

Last revision: 26 August 2019

WARNING:

Due to high out-migration from Poland, which is not fully captured in the official population statistics, the fertility indicators for the period since 2000 can be underestimated by as much as 10%. Cohort fertility indicators for the cohorts born after 1965 (especially childlessness) should be used only with special caution. Please see details in section 5.

1 General information

This report documents the data for Poland included in the Human Fertility Database Project. The Central Statistical Office (CSO) in Warsaw is the main body responsible for population statistics in Poland. All the data covered in the report were provided by the Demography Department of the CSO or collected from the tables published in the Demographic Yearbook of Poland published by the CSO. The data cover the period 1971–2016.

1.1 Territorial coverage

The Polish data collected for the HFD cover the whole territory of Poland. There were no territorial changes during the period these data cover. There was a major reform of administrative units (voivodships) carried out on 1 January 1999, after which their number, together with the number of regional statistical offices, was reduced from 46 to 16, but this did not affect the completeness of registration of births.

1.2 Data collection and availability

The data on births, included in the HFD, correspond to the period 1971–2016. Data for an earlier period are not available in any official statistical publication. The existing data on births come from the official registration of births conducted by the Central Statistical Office and are tabulated by age of mother and birth order. Starting from 1985, the Central Statistical Office has collected individual level data on births, which for the purposes of the HFD have been tabulated in the format and detail most suitable for the HFD.

Until 2014 the primary source of data on live births and still births in Poland was the "Notification of birth" maintained by the Ministry of Health, which serves primarily as the basic document for civil status acts and secondarily for national statistics utilization. Since 2015 the new "Birth certificate" maintained by the Ministry of Health has served as a source of medical data on live births, while

socio-demographic data come from the register of civil status and the central population register. Data on still births are reported separately, through the hospitals (CSO 2016: 184).

Population censuses carried out in 1970, 1988, and 2002 provide additional information on fertility and parity distribution. For the HFD, however, only the female age-parity distribution from the 2002 census is used (see section 3.2).

The monthly numbers of live births have been available from the official statistical publications since 1985.

Population data by age and sex used in the HFD come from the Human Mortality Database (HMD, www.mortality.org). The HMD provides data on the male and female population data for the period since 1958.

All the input data which are used for generating the HFD output data and indicators are specified in Appendix 1.

2 Birth count data

The data on live births by age of mother and birth order for each calendar year were prepared from the tabulations published in the Demographic Yearbook of Poland for the period 1971–1984. These tables contain data on live births by single years of age of mother and birth order. These data were already processed and used by Holzer and Holzer-Żelażewska (1997) and Holzer-Żelażewska and Tymicki (2010). As reported by Holzer and Holzer-Żelażewska (1997), the preparation of these data involved problems with missing information on the exact age and cohort of mother and in the definition of live birth across the whole period analysed. These issues are described in detail in section 4.

As noted before, the birth count data for the period since 1985 were computed using individual registration files provided by the Central Statistical Office. For each calendar year, these files contain the following information:

- year of registration (calendar year)
- administrative unit (voivodship)
- mother's residence
- birth date of child (day, month, year)
- child's birth weight
- information concerning viability of birth
- birth order (total birth order, live birth order, and birth order within marriage among marital births)
- whether birth is extramarital / marital
- birth date of mother (day, month, year)
- mother's marital status
- mother's education
- date of marriage (day, month, year)
- date of previous birth (day, month, year)
- viability of previous birth (day, month, year)

These data sets enabled production of tabulations of the number of live births by birth order and age of mother and mother's birth cohort for each year (Lexis triangles). The original data files provided by the Central Statistical Office contained both the births actually delivered in a given calendar year and those births registered in a given calendar year. Only a small fraction of all births were registered in a different year than the year of the delivery. Usually, these were cases of children who were born in December but who were not registered until the following year.

However, in order to obtain precise information, we have corrected the data and moved those births registered in the year other than the actual year of birth to the corrected data file. This issue is discussed in section 5.

3 Population count data

3.1 Population count data by age

The annual age structure of women is taken from the Human Mortality Database (HMD). Data quality issues as related to the population estimates are discussed separately in section 5.2.

3.2 Population count data by age/cohort and parity

The question of the number of live-born children is not included in the Polish population censuses, but was asked in the series of Fertility Surveys carried out as a part of the population censuses in 1970, 1988, and 2002. These large surveys collected retrospective data on reproduction and marriage among women. In the case of the 2002 survey, it was possible to generalize the results for the whole population with the use of inverse probability weights. Data on the female parity distribution from the 1970 and 1988 censuses were not used for the HFD computations for various reasons (see below for details). The 2011 census did not include a question on the number of children ever born.

Census 1970

The Fertility Survey carried out during the National Census 1970 (which began on 8 December) covered 412 thousand of ever-married women who were younger than age 70 at the time of the survey. The main results were tabulated in a publication by the Central Statistical Office (1971), further output was published in Bolesławski (1974 and 1993). However, as noted by Paradysz and Szymkowiak (2006), the quality of the published tables is not satisfactory and their use for the purposes of cohort and period analysis is limited, also because they do not cover single women. Therefore, these data are not used in the HFD.

Census 1988

During the National Census 1988 (which began on 6 December), the Fertility Survey was limited to 89,391 females aged 18+. The main results are tabulated in Paradysz (1992). Unfortunately, due to a high number of missing records, it was not possible to generalize the results for the entire population with the use of inverse probability weights (Paradysz 1992, Paradysz and Szymkowiak 2006). Therefore, these data are not used in the HFD.

Census 2002

Data on the distribution of women by age and parity have been calculated from the Fertility Survey carried out during the National Census of 2002. Representative sample includes 264,845 cases, which is 3.3% of female population. The National Census was carried out between 21 May and 8 June 2002. The core questionnaire included questions about full birth histories (up to the 20th birth) as well as union histories. This data source provides the most reliable source of information concerning parity distribution of women among all the fertility surveys carried out in Poland. The survey also collected information about the date of woman's birth, her level of education at the time of the survey, her fertility intentions, size of residence and region (voivodship) of residence. The data set includes women born between 1896 and 1986. There were very few missing cases:

information on birth histories was missing only for 1.2% of women included in the survey (0.9% of women at age 12–55).

According to the information obtained from the Polish Central Statistical Office, the sample was drawn using a two-stage procedure. Because the survey took place during the National Census, first sampling level contained census districts (27 thousand districts were considered for sampling). At the second stage, sampling was performed on the level of households, covering 280 thousand households. At the first stage a stratified sampling scheme was applied. Units were sampled with probability proportional to size (number of households in a given district). At the second stage units were sampled according to a simple random sampling scheme. This sampling scheme was applied in subpopulations selected according to rural and urban parts of the voivodships, in which the main units were defined as towns in urban areas and counties (*powiat*) in rural areas.

For each observation, the data contained weights which are equal to the inverse probability of being sampled in respective strata. The value of the inverse probability weight (p-weight) could be interpreted as equivalent to the number in the population represented. The inverse probability weights were aimed at making the sample representative of the population with respect to residence and age of surveyed women. The resulting distribution of women by parity then contains information on 8,409,090 women born between 1896 and 1986. However, due to selectivity linked to mortality, we do not recommend using these data for women born before and during the Second World War, i.e., the cohorts born before 1945 (aged 57 at the time of the interview). The quality of the data and the usefulness for analysing cohort fertility was validated by Tymicki (2010).

4 Specific details

4.1 Definition of age

The birth counts pertaining to the period 1971–1984 are organized by mother's age in completed years (Lexis squares). More details about these data can be found in Holzer and Holzer-Zelażewska (1997).

For the period since 1985, the birth counts have been tabulated using files on individual birth records obtained from the Central Statistical Office and are provided by both age in completed years and by birth cohort of mother (Lexis triangles).

4.2 Definition of birth order

Live birth order is defined by considering the total number of live births a woman had previously had, without counting stillbirths. In the case of multiple delivery (e.g. twin, triplet) every newborn baby is included in the statistical records as a separate birth (CSO 2016: 185). Since 1985, official statistics recognise two definitions of birth order – the one considering the total number of live births a woman had previously had; and the second where birth order considers all previous live births and still births. Data used in HFD always consider the previous number of live births.

The difference between the given totals of births and sum of births by birth order is assigned as with unknown birth order – these cases are however negligible (the highest number was 274 cases in 1974). There was a higher number of births of unknown birth order in 1987–1993 (about 2 to 3 thousands, i.e. 0.4–0.5% of total births). Since 1994 the number of unknown cases has been negligible.

4.3 Changes in the definition of live birth

During the period analysed, three different definitions of live birth were used.

- Until 1962 a live birth was defined as the complete expulsion or extraction from its mother, irrespective of the duration of pregnancy, of a newborn who at the moment of the cutting of the umbilical cord showed any signs of life, or who did not show any signs of life but was brought back to life.
- In 1963–1993 births with a birth weight higher than 600g and lower than 1001g that did not survive the first 24 hours were excluded from the live birth statistics. These births were classified as a separate group called “non-viable births with signs of life”. Births that weighed less than 600g were excluded from the births statistics.
- In 1994, following the recommendations of the World Health Organization (WHO), a new modified live birth definition was introduced, based on three elements: duration of pregnancy, birth weight, and birth height. Since 1994 a newborn is considered a live birth if it weighs at least 501g or – if the birth weight is unknown – it has been extracted from the mother’s body after at least 22 weeks of gestation or is at least 25 cm long.

In the data prepared for the HFD we have used the numbers of live births that were reported according to the live birth definitions used at the time of data collection/registration. It was not possible to recalculate the number of live births by single years of mother’s age according to the current live birth definition. This implies that the numbers of live births in the period before 1994 are slightly underestimated. Due to the change in definition, infant deaths were underestimated by about 20% (Fihel and Jasilionis 2016). However, the change of the definition of birth and infant death had only a small influence on data on live births (CSO 2015: 185). The error due to this slight underestimation has no significant influence on the estimates of period and cohort fertility rates.

5 Data quality issues

5.1 Births registered vs. the actual number of births in a year

As mentioned in section 2, the number of births from registration data may differ when taking criterion of being registered or being actually born in the year of registration into account. For the period before 1985 there is only information about children registered in the given calendar year, without any separation between births actually delivered in a given calendar year and births only registered in a given calendar year. This imprecision in properly accounting for the actual year of birth has only a minor effect on the estimated fertility rates. The minor difference between fertility rates based on the reported vs. the actual year of birth is due to the fact that births delivered in one year and erroneously reported in the following year (year of registration) are “compensated for” by the “extra” births reported in a given year, but actually born in the previous year.

From 1985 onwards, information is available on both the registration date and the actual date of birth of each child. Although parents are obliged, according to Polish law, to register their children within 14 days of delivery, this rule is not always followed. In 1985–2016, 97.4% of all children were registered in the year of the actual delivery. In order to provide precise data for the HFD, individual data files were used to clean the datasets and to provide the data on the number of children actually delivered in a given calendar year. The data for 2016 thus also include information on children delivered in 2016 but registered in 2017. Data for births in 2017 have not been included in the HFD due to unavailability of the data on births registered in 2018.

Monthly data for 1985–2016 provided for the HFD are also tabulated by the month and year of the actual delivery. The monthly numbers of births are therefore different than the numbers published by the Central Statistical Office, which always use the month of registration.

5.2 Population estimates: External migration and births/exposures mismatch

The quality of population estimates for Poland by age and sex is influenced by changing definitions of population in the official population statistics, as well as by the reliability and completeness of these population estimates, especially with respect to properly accounting for outmigration.

During the period when population estimates were covered by the Human Mortality Database (HMD), 1958–2016, four different definitions of population were used in the official demographic statistics (see Fihel and Jasilionis 2016 for details). In 1958–1982, population statistics referred to the currently resident (*de facto*) population. In 1983–2005, population statistics covered the population registered as permanent or temporary (residing for at least 2 months) residents. In 2006–2010, the definition was modified so that temporary residence referred to those staying in the country for at least 3 months. Finally, a more recent change was introduced with the 2011 census, when a new concept of “usual residence” was introduced in population statistics. The new official definition includes as usual residents all those individuals who are living in Poland (immigrants holding EU citizenship or temporary residence permits) and those declaring their intention to live in Poland for at least one year. This definition excludes people who officially declared to the authorities their departure abroad for a period of more than one year. The HMD assumes—in line with the official data produced by the Central Statistical Office—that emigrants who do not officially declare their departure to the authorities are still “usual residents” in the country (despite *de facto* residence abroad for a period longer than one year).

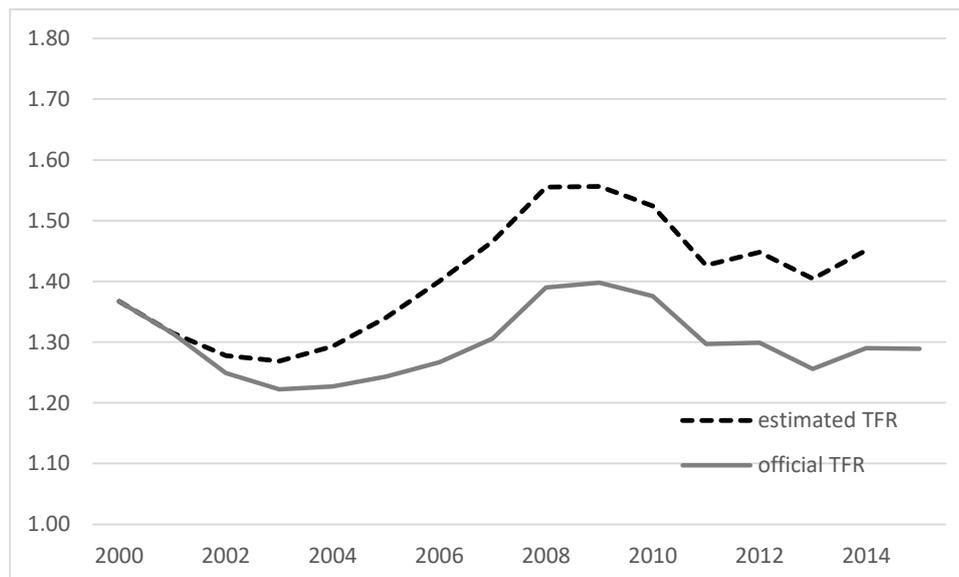
For the most recent inter-censal period, 2002–2010, and for the two years preceding the 2002 census, 2000–2001, the HMD used inter-censal population estimates recalculated back from the most recent census of 2011, based on resident population according to the official definition. However, according to the Central Statistical Office (2015), the number of Polish citizens who were included in the official resident population, but were in fact living abroad increased from 0.4 million in 2002, to 1 million in 2004 (just after the accession of Poland into the European Union) and then increased again to 1.95 million in 2006 and to 2.3 million in 2014. Moreover, this population is predominantly concentrated in peak productive and reproductive ages, especially in the age group 25–34 (Kaczmarczyk 2014, Gołata 2016). As a result, the “overestimation” of the actual resident population at these ages amounts to roughly 12% for 2011. To avoid the births/exposures bias in estimating fertility rates in the resident population, the births statistics would have to cover all births to those Polish citizens living abroad and registered as “usual residents” in Poland. This is, however, not the case – statistical reporting of births includes all the births registered by civil registration offices that took place in Poland in a given year (excluding births to mothers staying in Poland temporarily, but permanently residing abroad). The number of births to Polish citizens living abroad is difficult to estimate. For instance, in 2014 the number of births to mothers with Polish citizenship in England and Wales was 22.1 thousand (ONS 2016), while in Germany it amounted to 10.0 thousand (DESTATIS 2016). It is not known how many of these mothers were still registered as Polish residents at that time, but including these births in the birth statistics for Poland would boost the registered number of live births in the country (375 thousand in 2015) by almost 9%.

Thus the fertility indicators for Poland are underestimated and this underestimation has progressively worsened since 2000. In 2011 the number of women aged 15–49 and registered as residing in Poland who actually lived abroad reached 760 thousands. If we exclude these women from the estimated exposure population used for computing the period TFR, this indicator jumps from the officially reported level of 1.30 to 1.46 in 2011, i.e., by 11% (see Figure 1). We estimate that during 2004–2014, the births/exposures bias has led to an increasing underestimation of the

period TFR in Poland in the order of 5–10%. Our estimations are confirmed by Gołata (2016), who estimates the TFR for 2011 to be 1.45–1.46.¹

In a similar manner the cohort summary indicators can be underestimated as well. For example, the cumulated cohort fertility for first birth is suspiciously low among the women born in the late 1960s and the early 1970s. This suggests that the cohort childlessness calculated in the HFD—increasing from 10–14% among cohorts born in 1956–1967 up to 20% in the 1974 cohort—is overestimated for these women. Cohort fertility indicators for the cohorts born after 1965 (especially cohort childlessness) should thus be used only with special caution.

Figure 1 Official total fertility rate (TFR) for Poland and the estimated value that takes the under-registration of births abroad into account



6 Acknowledgements

We would like to acknowledge the work of Dorota Holzer Żelażewska in preparing the data from births registration as well as from the Fertility Surveys. We would also like to acknowledge the assistance of Małgorzata Waligórska from the Demography Department of the Central Statistical Office in acquiring data on births from the registration of vital events.

¹ Dormon (2014) estimates that the TFR of Polish migrants to England and Wales was 2.1 in 2011, i.e. by 0.8 higher than in Poland. Waller et al. (2014) estimates TFR of Polish migrants to England and Wales to be 2.15 in 2010–2011 but 1.4 for the entire period of 2004–2012.

7 References

- Bolesławski, L. 1974. Tablice płodności kobiet według generacji. [Fertility Tables by Generation.] Polish Statistics/Statystyka Polski, Central Statistical Office, Warsaw.
- Bolesławski, L. 1993. Polskie tablice dzietności kobiet 1971–1992. [Polish fertility tables 1971–1992.] Central Statistical Office, Warsaw.
- Central Statistical Office. 1945–1984. Rocznik Demograficzny 1945–1966, 1967–1968, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1984. [Demographic Yearbook...], Główny Urząd Statystyczny, Warsaw.
- Central Statistical Office. 1971. Dzietność kobiet, t. 1, 2. [Fertility of women, vol. 1, 2.]. Główny Urząd Statystyczny, Warsaw.
- Central Statistical Office. 2015. Informacja o rozmiarach i kierunkach czasowej emigracji z Polski w latach 2004–2014. Notatka informacyjna 5.10.2015 r. Główny Urząd Statystyczny, Warszawa.
- Central Statistical Office. 2016. Demographic Yearbook of Poland 2016. Central Statistical Office, Warsaw.
- DESTATIS. 2016. Live births: Germany, years, citizenship. GENESIS-Online Datenbank. Retrieved on 3.4.2017 from <https://www-genesis.destatis.de/genesis/online>
- Dormon, O. 2014. Childbearing of UK and non-UK born women living in the UK: 2011 Census data. Office for National Statistics. Retrieved on 3.4.2017 from <http://webarchive.nationalarchives.gov.uk/20160105160709/http://ons.gov.uk/ons/rel/fertility-analysis/childbearing-of-uk-and-non-uk-born-women-living-in-the-uk/2011-census-data/index.html>
- Fihel, A. and Jasilionis, D. 2016. About mortality data for Poland. Human Mortality Database country documentation file.
- Gołata, E. 2016. Estimation of fertility in Poland and of Polish born women in the United Kingdom. *Studia Demograficzne* 1(169): 13–38.
- Holzer, J. Z. and Holzer-Żelażewska, D. 1997. Płodność kohortowa kobiet w Polsce w latach 1945–1994. [Cohort Fertility of Polish Women, 1945-1994.] *Studia Demograficzne* 2(134): 3–23.
- Holzer-Żelażewska, D. and Tymicki, K. 2010. Cohort and Period Fertility of Polish Women, 1945–2008. *Studia Demograficzne* 1.
- Kaczmarczyk, P. 2014. Recent Trends in International Migration in Poland. The 2012 SOPEMI Report. CMR Working Papers 71/129. Centre of Migration Research, University of Warsaw.
- ONS. 2016. Parents' country of birth, 2015, Office for National Statistics. Retrieved on 3.4.2017 from <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/parentscountryofbirth>
- Paradysz, J. 1992. Dzietność kobiet w Polsce. [Fertility of women in Poland], Central Statistical Office, Warsaw.
- Paradysz, J. and Szymkowiak, M. 2006. Źródła danych ludnościowych, [Sources for population studies]. Metodologia Badań Demograficznych Zeszyt nr 15. Sekcja Analiz Demograficznych PAN., Warsaw.
- Tymicki, K. 2010. Validation of data quality from Polish Fertility Survey 2002 with use of cohort fertility rates. *Studia Demograficzne* 1–2(157–158): 61–77.
- Waller L., Berrington A. and Raymer J. 2014. New insights into the fertility patterns of recent Polish migrants in the United Kingdom. *Journal of Population Research* 2(31): 131–150.

**APPENDIX 1
INPUT DATA USED FOR HFD CALCULATIONS**

BIRTHS

Period	Type of data	Age range	Birth order	RefCode(s)
1971–1974	Annual number of live births by age of mother and birth order (Lexis squares)	-15, 16,..., 40–44, 45+	1, 2,..., 8+, UNK	1
1975–1984	Annual number of live births by age of mother and birth order (Lexis squares)	-15, 16,..., 44, 45+	1, 2,..., 8+, UNK	1, 2
1985–2016	Annual number of live births by age of mother, mother's year of birth and birth order (Lexis triangles)	12,..., 58	1, 2,..., 10+, UNK	3, 6
1985–2016	Annual number of live births by month	–	–	3, 6

FEMALE POPULATION: Distribution by age and parity

Census date	Type of data	Age range	Year of birth, range	Parity	RefCode(s)	Notes
01.06.2002	Number of women by age and parity	–	1896–1986	1,..., 8+	4	unknown parity redistributed proportionally

FEMALE POPULATION: Exposure by age and year of birth

Female exposure population by calendar year, age, and year of birth (Lexis triangles) is estimated using data on population size and deaths from the Human Mortality Database, which is available at <http://www.mortality.org> or <http://www.humanmortality.de>.