

PREPRINT

Author-formatted, not peer-reviewed document posted on 25/02/2026

DOI: <https://doi.org/10.3897/arphapreprints.e189270>

**Cohort fertility database for regions of Russia and
countries**

Pavel Kishnin

Introduction

The analysis of fertility trends is fundamental to understanding demographic processes and planning social policy. In demographic research, two main approaches are distinguished: the analysis of period fertility and the analysis of cohort fertility. Period indicators, such as the period total fertility rate, reflect the intensity of births in a given calendar year. However, they are highly susceptible to structural changes and timing effects, such as the postponement of maternity to older ages. Consequently, period indicators may not accurately reflect the true number of children born to a cohort of women. Cohort fertility indicators, such as the completed cohort fertility rate, provide a more stable and accurate measure of reproductive behavior, as they summarize the fertility experience of women born in the same year throughout their entire reproductive life.

Despite the analytical value of cohort indicators, the availability of such data at the regional level remains limited. Existing international databases, such as the Human Fertility Database (HFD), provide high-quality data primarily at the national level. For the Russian Federation, the Russian Fertility and Mortality Database (RusFMD) offers extensive period data, but there is a lack of long-term series of cohort indicators for the subjects of the Federation. The vast territory of Russia is characterized by significant spatial heterogeneity in demographic development. Regions differ in the timing of the demographic transition, ethnic composition, and socio-economic conditions. Therefore, national averages often mask divergent trends at the regional level.

The «Cohort fertility database for countries and regions of Russia» presented in this article was created to fill this gap. It consolidates data for 278 countries and territories, based on the IASA-2025 estimates, and provides a unique detailed dataset for 83 regions of the Russian Federation. The database allows researchers to trace the evolution of fertility from the cohorts born in the 1930s to the cohorts currently active in reproductive behavior. The inclusion of projection data for cohorts born up to 2050 allows for the analysis of future demographic prospects. This database is useful for studying the transformation of the age profile of fertility, analyzing changes in family structure by the number of children, and forecasting the future labor force and population structure.

Data format and availability

The database is organized as a collection of separate files in the `.xlsx` (Microsoft Excel) format. This format ensures compatibility with most statistical software packages and easy of use for researchers. The data are available for download free of charge and without the need for registration. The primary access point is the website of the International Laboratory for Population and Health at HSE University.

Database structure and description of variables

The structure of the database is designed to facilitate both cross-country comparisons and in-depth regional analysis within Russia. The data are divided into two main blocks: international data and Russian regional data. The files are named according to the content and the specific demographic indicators they contain. The following list describes the variables and the corresponding file contents included in the database.

First, the database contains the completed cohort fertility rate (CCFR) for 278 countries and dependent territories. These data cover female cohorts born from 1935 to 2050. The variable represents the average number of children born to a woman of a specific birth cohort by the time

she reaches age 50. The data are based on age-specific fertility rates derived from the International Institute for Applied Systems Analysis database (version January 2025, IIASA-2025) using the SSP-2 forecast model (median fertility and mortality).

Second, the database includes the CCFR specifically for 83 subjects of the Russian Federation. This block is divided into two temporal resolutions. One file contains data for single-year cohorts of women born from 1935 to 2000. Another file contains data for five-year cohorts of women born from 1935–1940 to 2045–2050. This indicator allows for the assessment of the total fertility intensity of cohorts in each specific region.

Third, the database provides the completed cohort fertility rate by birth order for 83 subjects of the Russian Federation. The data are presented for five-year cohorts. The variable is disaggregated by birth order (1st child, 2nd child, 3rd child, 4th child and higher). This allows researchers to analyze structural changes in the family, such as the decline in the prevalence of large families or the increase in childlessness.

Fourth, the database includes the mean age of birth (MAB) by age 50. These data are provided for 83 subjects of the Russian Federation for five-year cohorts. The indicator reflects the timing of fertility and allows for the tracking of the process of aging of the fertility schedule (postponement of births).

Fifth, the database contains the mean age at birth by birth order (1, 2, 3, 4+) by age 50 for 83 subjects of the Russian Federation. These data are presented for five-year cohorts. This variable is crucial for understanding the timing of transition to the next parity.

Sixth, the database presents the parity progression ratios (PPR) to age 50. These data cover 83 subjects of the Russian Federation for five-year cohorts. The variable represents the probability that a woman with a certain number of children will give birth to another child.

Seventh, the database includes the cohort distribution of women by number of children born (parity distribution) by age 50. The data are provided for 83 subjects of the Russian Federation for five-year cohorts. The variable shows the proportion of women with 0, 1, 2, 3 or more children.

Eighth, the database contains the proportion of women who have birth of order N by age 50. These data cover 83 subjects of the Russian Federation for five-year cohorts. The variable indicates the cumulative proportion of women who have reached at least a specific parity (1, 2, 3+).

Methodology

The creation of the database required the integration of various data sources and the application of specific demographic analysis and econometric methods (calculation of age-specific fertility rates for Russian regions for 2024–2029 using regression models of economic growth of per capita income and expenditure of the population, adjusted for the potential change in the average age at birth). The methodology ensures the comparability of long-term time series and the reliability of projections for cohorts that have not yet completed their reproductive period.

Data

For the international comparison block, the data were derived from the IIASA-2025 database. The projection model SSP-2 (Shared Socioeconomic Pathways, Middle of the Road) was

selected as the baseline scenario. This model assumes median trends in fertility and mortality. The methodology for these projections is described in K.C. et al. (2024).

The primary sources for calculating the cohort and period fertility indicators for the Russian Federation were the annual numbers of births by age of mother and birth order for the period 1946–2023. Additionally, data on the population by sex and age for the period 1959–2023 were used (Churilova et al. 2024). These data were provided by the Federal State Statistics Service (Rosstat), which applies both to the annual number of births by mother's age and birth order, and to the population structure for all periods from 1989 to 2023 for which such data are available. For the earlier period of 1946–1958, the data on the sex and age composition of the Russian population were taken from the demographic reconstruction presented in the work of Andreev et al. (1998).

Estimates of the completed cohort fertility rate for 83 regions of Russia were taken from a study analyzing this indicator - a more detailed methodology was described by researchers earlier (Kishenin 2023). Primary data on the cohort mean age of birth and the distribution of the completed cohort fertility rate by birth order were taken from the 1979, 1989, 2002, 2010, and the 2020 Census, conducted in 2021 (its data were used only for regions for which cohort fertility rates are retrospectively comparable with the corresponding data series from the 1979, 1989, 2002 and 2010 censuses). Age-specific fertility rates from the RusFMD database were also used.

The data used and methodology for the one-year and five-year cohorts are described in more detail in Tables 1 and 2 (Appendix 1).

Limitations of the study and data

The key challenge of the study was the use of a fairly large number of sources, which required the use of both demographic analysis and statistics to adjust the completed cohort fertility rates in the constituent entities of the Russian Federation based on the constituent entities' share of the total number of women of reproductive age and the total number of children born in the corresponding birth cohort in the country to accurately calculate the completed cohort fertility rates for the country as a whole.

This adjustment also helped account for the average effects of internal female migration between regions, as we lack other sufficiently reliable methods for assessing the impact of internal migration on the cohort fertility indicators through systematic surveys. Some problems arose in estimating cumulative fertility rates for birth cohorts whose active reproductive age occurred before 1989/1990, when precise regional cumulative fertility rates are unavailable and only census results had to be relied upon. Following the recommendation of the Statistics Japan Bureau (Ogawa, Retherford 1993), data from five censuses were used to refine these data: 1979 and 1989, 2002, 2010, and partially 2021 (Rosstat). Using multiple censuses allows us to somewhat mitigate potential selective effects of fertility on mortality, i.e., those associated with survival to the time of the next census.

Certain shortcomings of the study also relate to distortions in population size and, consequently, cohort fertility indicators in a number of constituent entities of the Russian Federation, which is particularly relevant for the national republics in the eastern North Caucasus.

Calculation of completed cohort fertility rate

The completed cohort fertility rate is the fundamental indicator in the database. It represents the average number of children born to a woman belonging to a specific birth cohort by the end of her reproductive period, which is conventionally defined as age 50.

For cohorts that have already completed their childbearing (women born before the mid-1970s), the CCFR is calculated as the sum of the age-specific fertility rates (ASFR) observed for that cohort at each age from 15 to 49. The formula for calculation is:

$$CCFR_c = \sum_{x=15}^{49} f_{x,c}$$

where $CCFR_c$ is the completed fertility rate for cohort c , and $f_{x,c}$ is the fertility rate at age x for women of cohort c . The rate $f_{x,c}$ is calculated as the ratio of the number of births to women of age x from cohort c to the average annual population of women of that age and cohort.

However, for cohorts where annual registration data are incomplete or missing, the database utilizes the results of the population censuses conducted in 1979, 1989, 2002, 2010, and 2021. For these cohorts, the indicators of completed fertility are taken directly from the census data on the average number of children ever born to women who have reached age 50. This census-based approach is used as the primary source for the reconstruction of fertility for older cohorts in the Russian regions. This allows the database to provide reliable values even for periods before the establishment of detailed regional vital statistics.

For the younger cohorts that have not yet reached age 50, the CCFR is calculated as the sum of the accumulated fertility observed in recent years and the projected additional fertility. The projected rates are derived from the Bayesian hierarchical model of IIASA-2025, extrapolated to the data of 83 Russian regions. The CCFR is also calculated by birth order N ($CCFR_N$). This indicator represents the average number of children of order i born to a woman of the cohort. The total CCFR is the sum of CCFR by orders:

$$CCFR_N = \sum_N CCFR_{N,c}$$

Calculation of cohort mean age of birth

The mean age of birth (MAB) characterizes the timing of fertility in the cohort. For cohorts with complete annual registration series, the weighted average age is calculated using the age-specific fertility rates. The formula used in the database is:

$$MAB_c = \frac{\sum_{x=15}^{49} (x + 0,5) \times f_{x,c}}{\sum_{x=15}^{49} f_{x,c}}$$

where x represents the age in completed years, and 0,5 is the correction factor assuming that births are distributed evenly throughout the year of age.

In cases where annual registration data are insufficient to form a complete cohort profile, the mean age of birth is reconstructed using the retrospective information from the population censuses. This methodology involves the analysis of the cumulative fertility achievements at different ages as reported by women in different census years. For these cohorts, the age profile is adjusted to match the total completed fertility levels reported in the census. This ensures that the timing indicators are consistent with the final number of children born to the cohort. The

same logic is applied to the calculation of the mean age of birth for specific birth orders ($MAB_{N,c}$):

$$MAB_{N,c} = \frac{\sum_{x=15}^{49} (x + 0,5) \times f_{x,N,c}}{\sum_{x=15}^{49} f_{x,N,c}}$$

This indicator allows for the analysis of the postponement of first births separately from the postponement of second or third births.

Calculating the parity progression ratio

The parity progression ratio (PPR) measures the probability that a woman with i children will proceed to have an $N + 1$ child. In the context of this database, PPR is calculated for the cohort by the end of the reproductive period.

The calculation is primarily based on the distribution of women by the number of children born as reported in the population censuses. For the cohorts that have completed their reproductive life, the progression ratios are derived from the census distributions of women by parity. Let a_i be the parity progression ratio from N to $N + 1$ children. It is defined as:

$$a_N = \frac{\text{Proportion of woman with } N + 1 \text{ or more children}}{\text{Proportion of woman with } N \text{ or more children}}$$

Alternatively, using the completed cohort fertility rates by order ($CCFR_N$), the probability can be approximated in datasets where individual parity distribution is reconstructed from order-specific rates. The database relies on the processed distributions obtained from the synthesis of census data and vital statistics. For instance, the probability of having a first child (a_0) is the proportion of women who became mothers (1 minus the proportion of childless women). The probability of having a second child (a_1) is the proportion of women who had at least two children divided by the proportion of women who had at least one child.

Calculation of the cohort proportion and parity distribution of women by number of children born

The database provides the final distribution of women by the number of children ever born by age 50. This distribution is denoted as w_N , representing the proportion of women in the cohort who have exactly N children.

The calculation of this distribution involves the integration of census data and vital statistics. Census data provide benchmarks for the parity distribution at specific time points (1979, 1989, 2002, 2010, 2021). For the periods between censuses and for the projection period, the distribution is reconstructed using the age-specific fertility rates by birth order.

The proportion of women with n children is related to the birth order rates. The cumulative proportion of women who have had at least N children (denoted as $P_{\geq N}$) corresponds to the Completed Cohort Fertility Rate for order N ($CCFR_N$) only if multiple births (twins, triplets) are accounted for or assumed negligible for the specific rate definition. In this database, the calculation follows the standard demographic balance equations:

$$P_{\geq 1} = \sum_x f_{x,1,c}$$

$$P_{\geq 2} = \sum_x f_{x,2,c}$$

The proportion of women with exactly N children (w_N) is derived as:

$$w_0 = 1 - P_{\geq 1}$$

$$w_N = P_{\geq N} - P_{\geq N+1}$$

, for $N \geq 1$

This distribution allows for the detailed analysis of childlessness levels (w_0), the prevalence of the two-child family norm (w_2), and the frequency of large families (w_{3+}) across different regions and cohorts.

Application of data

The «Cohort fertility database for regions of Russia and countries » has a wide range of potential applications in scientific research and practical policy making.

First, the database allows for the testing of demographic theories at the regional level. Researchers can investigate the diffusion of the second demographic transition across the diverse territory of Russia. The data allow for the analysis of whether regions are converging towards a single low-fertility model or maintaining distinctive reproductive patterns.

Second, the data are essential for spatial analysis. The availability of data for 83 regions enables the construction of detailed demographic maps and the identification of clusters with similar fertility behaviors. This is particularly relevant for a multi-ethnic country like Russia, where cultural factors significantly influence fertility.

Third, the cohort perspective provides a more accurate basis for forecasting. Unlike period indicators, which fluctuate significantly due to calendar effects, cohort indicators reveal the fundamental trends in the number of children women actually have. This information is crucial for forecasting the future number of births, the size of the younger cohort, and the dependency ratios in the pension system. As demonstrated by Lutz et al. (2014), cohort based projections are superior for long term human capital assessments.

Fourth, the database supports social policy evaluation. By analyzing the completed fertility of cohorts exposed to different policy measures (e.g., the introduction of Maternity Capital), researchers can assess the long-term effectiveness of pronatalist policies, distinct from short-term tempo effects. Bongaarts and Sobotka (2012) emphasize that understanding these timing effects is critical for sustainable demographic planning.

List of reference

Andreev, E. M., Darsky, L. E., & Kharkova, T. L. (1998). Demographic History of Russia: 1927-1959. Informatics. Moscow.

Bongaarts J., Sobotka T. (2012) A Demographic Explanation for the Recent Rise in European Fertility. *Population and Development Review* 38(1): 83–120. <https://doi.org/10.1111/j.1728-4457.2012.00473.x>

Churilova, E. V., Andreev, E. M., Chertenkov, K. O., Kishenin, P. A., Rodina, O. A., & Sokolova, V. V. (2024). Russian Fertility Database. *Population and Economics*, 8(4), 138-149. <https://doi.org/10.3897/popecon.8.e135073>

K.C., S., Dhakad, M., Potančoková, M., Adhikari, S., Yildiz, D., Mamolo, M., Sobotka, T., Zeman, K., Abel, G., Lutz, W., & Goujon, A. (2024). Updating the Shared Socioeconomic Pathways (SSPs) Global Population and Human Capital Projections. IIASA Working Paper. Laxenburg, Austria: WP-24-003. <https://pure.iiasa.ac.at/id/eprint/19487/>

Kishenin P. A. (2023). Regional differentiation of fertility in the Russian Federation: cohort perspectives. *Demographic Review*, 10(4), 86-120. <https://doi.org/10.17323/demreview.v10i4.18810>

Lutz W., Butz W. P., KC S. (Eds). (2014). *World Population and Human Capital in the Twenty-First Century*. Oxford University Press, Oxford.

Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). (2023). *Human Fertility Database*.

Ogawa, N., & Retherford, R. D. (1993). The Resumption of Fertility Decline in Japan: 1973-92. *Population and Development Review*, 19(4), 703–741. <https://doi.org/10.2307/2938411>

List of censuses

USSR State Statistics Committee (1979). 1979 Union Population Census of the USSR.

USSR State Statistics Committee (1989). 1989 Union Population Census of the USSR.

Rosstat (2002). 2002 Russian Population Census. Volume 12 "Fertility".

Rosstat (2010). 2010 Russian Population Census. Volume 10 " Fertility".

Rosstat (2021). 2020 Russian Population Census. Volume 9 " Fertility".

Other sources of information

Human Fertility Database: <https://www.humanfertility.org>

Russian Fertility and Mortality Database (RusFMD):
http://demogr.nes.ru/index.php/ru/demogr_indicat/data_description

Federal State Statistics Service (Rosstat): <https://rosstat.gov.ru>

International Institute for Applied Systems Analysis (IIASA): <https://iiasa.ac.at>

IIASA-2025 (countries with a population of over 100 000 people):
<https://dataexplorer.wittgensteincentre.org/wcde-v3/>

IIASA-2025 (all countries and dependent territories): <https://zenodo.org/records/14718294>

Russian Fertility Database: <https://demogr.hse.ru/en/RFD>

Acknowledgments

The author expresses gratitude to HSE University for supporting this research. Special thanks are extended to the Federal State Statistics Service (Rosstat) for providing the initial data necessary for the calculations. The author is deeply grateful to Sergei Vladimirovich Zakharov for his invaluable guidance and expertise in demographic analysis. The author also thanks Elena Vladimirovna Churilova for her significant contribution to the development of the database methodology and data processing.

Appendix 1

Table 1. Data used for one-year and five-year birth cohorts

Cohort fertility indicator, one-year cohort 1935-2000	Cohort fertility indicator, five-year cohort 1935-2050
Data on age-specific fertility rates for Russia as a whole for 1-year periods of time from HFD/by regions of Russia for 1-year periods of time from RusFMD	Expected age-specific fertility rates constructed using a multi-level hierarchical regression model (forecast data on the socio-economic development of Russian regions for 2024 and 2025-2027 from the Ministry of Finance of Russia and for Russia as a whole for 2024-2029 from the World Bank and the IMF)
Data on the number of children born by age of mothers by regions of Russia by 5-year age groups for the 1979, 1989, 2002, 2010 and 2020 censuses	Data from the Bayesian hierarchical forecast model IIASA-2025 for constructing extrapolation models of expected completed cohort fertility rates from 2000-2005 to 2045-2050 for the population of 83 constituent entities of the Russian Federation
Data on the share of living women from the total number of women and the share of children from the total number of births in the constituent entities of the Russian Federation at the time of the census	
Estimates of the distribution of the completed cohort fertility rates by birth order and calculations of the average age of mother at childbirth	
Results of the sample survey of reproductive plans of the population in 2017 and 2022 from Rosstat	
Estimates of reproductive intentions based on the representative survey "Man, Family, Society" conducted by RANEPa in 2017 and 2020	
Sobotka's model tables for estimating residual fertility	

Table 2. Methodology used for one-year and five-year birth cohorts

Cohort fertility indicator, one-year cohort 1935-2000	Cohort fertility indicator, five-year cohort 1935-2050
Calculation of cumulative fertility rates based on data from RusFMD and the 1979, 1989, 2002, 2010, and 2020 population censuses for 83 Russian regions	Estimates of the completed cohort fertility rates by regions of the Russian Federation for the years 1935-2000 of birth, generalized from 1-year-olds to 5-year-olds

Application of correction for bringing regional data through the share of living women from the total number of women and the share of children from the total number of births in the subjects of the Russian Federation to the values of the country as a whole according to HFD data using simple double iteration methods	Construction of expected age-specific fertility rates using a multilevel hierarchical regression model for the calendar years 2024–2029
Calculation of residual fertility rates by the Sobotka-Zeman method using Sobotka model tables and application of the tempo-adjusted total fertility rate by region using the Bongaarts-Fini method, integrated estimates of reproductive intentions indicators and estimates of the possible realization of reproductive intentions through a system of second-order differential equations using the Chang-Kasturića method. Calculation of other indicators of cohort fertility from census data (mean age at birth in general and by order, distribution of women and completed cohort fertility rate by order, parity progression ratios)	Extrapolation of the completed cohort fertility rates from 1935-1940 to 1995-2000 birth years into the Bayesian hierarchical forecast model IIASA-2025: the arithmetic mean for each birth interval and region between the results of 3 methods of extrapolation inverse interpolation - the finite difference method in the extrapolation version (Aitken-Neville algorithm), the Lagrange polynomial method, the calculation of the spline function in R

Appendix 2

Table 3. Cohort fertility indicators and their corresponding files in the database

Indicator	Cohorts	File name
Completed cohort fertility rate to age 50 for 278 countries and dependent territories, total	1935-2050	IIASA-2025, completed cohort fertility rate.xlsx
Completed cohort fertility rate to age 50 for 83 subjects of the Russian Federation by single year cohorts, total	1935-2000	Russia, completed cohort fertility rate.xlsx, sheet «1935-2000»
Completed cohort fertility rate to age 50 for 83 subjects of the Russian Federation by five year cohorts, total	1935-2050	Russia, completed cohort fertility rate.xlsx, sheet «1935-2050»
Completed cohort fertility rate to age 50 for 83 subjects of the Russian Federation by five year cohorts, by birth order (1, 2, 3, 4+)	1935-2050	Russia, CCFR proportion.xlsx
Cohort mean age of mother of birth to age 50 for 83 subjects of the Russian Federation by five year cohorts, total	1935-2050	Russia, cohort MAB.xlsx, sheet «MAB 1935-2050»
Cohort mean age of mother of birth to age 50 for 83 subjects of the	1935-2050	Russia, cohort MAB.xlsx, sheet «MAB by parity 1935-2050»

Russian Federation by five year cohorts, by birth order (1, 2, 3, 4+)		
Parity progression ratios to age 50 for 83 subjects of the Russian Federation by five year cohorts	1935-2050	Russia, cohort PPR.xlsx
Cohort proportion of women by number of children born to age 50 for 83 subjects of the Russian Federation by five year cohorts, by birth order (0, 1, 2, 3+)	1935-2050	Russia, cohort proportion.xlsx, sheet «Cohort proportion 1935-2050»
Proportion of women (parity distribution) who have birth of order N to age 50 for 83 subjects of the Russian Federation by five year cohorts, by birth order (1, 2, 3+)	1935-2050	Russia, cohort proportion.xlsx, sheet «Parity distribution 1935-2050»

Appendix 3

Table 4. Description of the data structure and content

Indicator	Description
Completed (accumulated) cohort fertility rate to age 50 for 278 countries and dependent territories, total	Country or dependent territory (row). Cohort (column). All births, completed (accumulated) fertility rate to age 50.
Completed (accumulated) cohort fertility rate to age 50 for 83 subjects of the Russian Federation by single year cohorts, total	Region (row). Cohort (column). All births, completed (accumulated) fertility rate to age 50.
Completed (accumulated) cohort fertility rate to age 50 for 83 subjects of the Russian Federation by five year cohorts, total	Region (row). Cohort (column). All births, completed (accumulated) fertility rate to age 50.
Completed (accumulated) cohort fertility rate to age 50 for 83 subjects of the Russian Federation by five year cohorts, by birth order (1, 2, 3, 4 and higher)	Region (row). Cohort (column). All births, completed (accumulated) fertility rate to age 50. First births, completed fertility rate to age 50, and so on. Fourth and subsequent births, completed fertility rate to age 50.
Cohort mean age of mother of birth to age 50 for 83 subjects of the Russian Federation by five year cohorts, total	Region (row). Cohort (column). All births, mean age of mother at birth to age 50.
Cohort mean age of mother of birth to age 50 for 83 subjects of the Russian Federation by five year cohorts, by birth order (1, 2, 3, 4 and higher)	Region (row). Cohort (column). All births, mean age of mother at birth to age 50. First births, mean age of mother at birth to age 50, and so on. Fourth and subsequent births, mean age of mother at birth to age 50.
Parity progression ratios to age 50 for 83 subjects of the Russian Federation by five year cohorts	Region (row). Cohort (column). Probability of having the first child. Probability of having the second child. Probability of having the third and subsequent child.
Final distribution of women by number of children born to age 50 for 83 subjects of the Russian Federation by five year	Region (row). Cohort (column). By age 50, the proportions of women with no children (0) and those with children of order 1, 2, 3 and higher.

cohorts, by birth order (0, 1, 2, 3 and higher)	
Proportion of women who have birth of order N to age 50 for 83 subjects of the Russian Federation by five year cohorts (1, 2, 3 and higher)	Region (row). Cohort (column). By age 50, the proportions of women who have reached at least birth order 1, 2, 3 and higher.