SOCIOLOGY · СОЦИОЛОГИЯ

Vestnik MIRBIS. 2026; 1 (45)': 182–192
Вестник МИРБИС. 2026. № 1 (45)'. С. 182–192

Original article
DOI: 10.25634/MIRBIS.2026.1.20

Error Detection in Russian Infant Mortality Data: Algorithm Development

Violetta S. Rusova[1], **Mariya R. Bikchenteeva**[2]

Abstract. The article presents a new development of an evidence-based methodology for an algorithmic approach to identify anomalies in data on infant mortality in Russia for the period 2015–2024. The present investigation relies upon a comprehensive analysis involving regulatory research, the application of a system of linear equations, a matrix data structure, and the Z-score method, based on official Rosstat statistics. The developed universal methodology for detecting data anomalies in raw data combines a comprehensive approach to diagnosis, specialized linear equation methods for imputation of missing data, a matrix structure for spatiotemporal analysis and data standardization as well as a Z-score for visualizing results. This analysis identified key demographic trends: a steady decline in the number of live births in all regions of Russia, specific features of the infant mortality rate (until 2021 — coefficient K, after 2021 — coefficient Kr) and presence of anomalous values. The research offers an integrated approach to diagnosing anomalies in infant mortality data that combines original methodological techniques and analytical tools. The practical significance of the developed methodology is manifested in improving the accuracy of data defect diagnosis, optimizing the management decision-making system in healthcare and improving quality control mechanisms for medical care. This research primarily targets the identification of systemic issues in data collection and processing, alongside the development of effective methods to mitigate and eliminate these problems.

Key words: infant mortality, legal-regulatory analysis, linear equations, matrix structure, Z-score.

For citation: Rusova V. S. Error Detection in Russian Infant Mortality Data: Algorithm Development. By V. S. Rusova, M. R. Bikchenteeva. DOI: 10.25634/MIRBIS.2026.1.20. *Vestnik MIRBIS*. 2026; 1:182–192.

Научная статья
УДК: 314.14:004.8

DOI: 10.25634/MIRBIS.2025.4.2

Обнаружение ошибок в данных о младенческой смертности в России: разработка алгоритма

Виолетта Сёмавна Русова[3], **Мария Рашидовна Бикчентеева**[4]

Аннотация. В статье представлена новая разработка научно обоснованной методологии алгоритмического подхода к выявлению аномалий в данных о младенческой смертности в России за период 2015–2024 гг. Настоящее исследование опирается на комплексный анализ, включающий нормативно-правовое исследование, применение системы линейных уравнений, матричную структуру данных и метод Z-оценки на основе официальной статистики Росстата. Разработанная универсальная методология обнаружения аномалий в исходных данных сочетает комплексный подход к диагностике, включающий метод линейных уравнений для восполнения пропущенных данных, матричную структуру для пространственно-временного анализа и стандартизации данных, а также Z-оценку для визуализации результатов. В ходе исследования выявлены ключевые демографические тенденции: устойчивое снижение числа живорождённых во всех регионах России, специфические особенности коэффициента младенческой смертности (до 2021 г. — коэффициент K, после 2021 г. — коэффициент Kr), а также наличие аномальных значений. Исследование предлагает интегрированный подход к диагностике аномалий в данных о младенческой смертности, который объединяет оригинальные методологические приёмы и аналитические инструменты. Практическая значимость разработанного алгоритма заключается в повышении точности диагностики аномалий данных, оптимизации системы принятия управленческих решений в здравоохранении и совершенствовании механизмов контроля качества медицинской

1 independent researcher (Moscow, Russia), madam.auditor@gmail.com, vio.rus.info@mail.ru, https://orcid.org/0000-0002-1816-4295
2 independent researcher (Saint-Petersburg, Russia), mary.bic4enteeva.ru@yandex.ru, https://orcid.org/0009-0000-6523-6194
3 независимый исследователь (Москва, Россия), madam.auditor@gmail.com, vio.rus.info@mail.ru, https://orcid.org/0000-0002-1816-4295
4 независимый исследователь (Санкт-Петербург, Россия), mary.bic4enteeva.ru@yandex.ru, https://orcid.org/0009-0000-6523-6194

помощи. Данное исследование в первую очередь нацелено на выявление системных проблем в сборе и обработке данных, а также на разработку эффективных методов их снижения и устранения.

Ключевые слова: младенческая смертность, нормативно-правовой анализ, линейные уравнения, матричная структура, Z-оценка.

Для цитирования: Rusova V. S. Error Detection in Russian Infant Mortality Data: Algorithm Development / V. S. Rusova, M. R. Bikchenteeva. DOI: 10.25634/MIRBIS.2026.1.20 // Вестник МИРБИС. 2026; 2:182–192.

Introduction

Today, the development of an effective algorithm for diagnosing anomalies in the infant mortality data monitoring system is of particular importance for the development of the healthcare system. Infant mortality is an integral indicator of the functioning of the entire medical system, reflecting not only the availability of high-quality medical care, but also the effectiveness of government programs, prompt reaction to threats to public health as well as general health policy [World health statistics 2023]. An algorithm serves as a diagnostic tool for comprehensive analysis of the healthcare system. The detection of anomalies in one of the system components inevitably entails the need to assess the functioning of the entire structure [Burstein 2019]. These anomalies could signify either localized malfunctions or deeper systemic issues, including violations in resource management, a shortage of qualified medical care, insufficient preventive work and poor planning system. Developing an algorithm to detect data anomalies holds practical importance because of several critical reasons. First, there is the need for an objective assessment of the effectiveness of national healthcare projects. Second, continuous surveillance of key public health indicators is necessary, as it enables timely identification of problem areas and facilitates rapid response [Liu 2012]. Third, the importance of prompt response to identified deviations in order to prevent further deterioration of the situation. A scientifically based approach to the diagnosis of infant mortality provides a reliable basis for managerial decision-making.

As noted by J. C. Whipple and S. A. Novoselsky, for the full-fledged development of the state, scrupulous consideration of its demographic potential is necessary [Whipple 1929]. This process implies a comprehensive approach to registering all demographic changes: from recording deaths and citizens traveling abroad to accounting for new births and the influx of immigrants. The authors emphasize that the real prosperity of a state is measured not so much by its natural wealth or material assets as by the health of the population and the general level of its well-being. Demographic statistics, according to researchers, acts as a fundamental tool for analyzing and managing the country's human capital, playing an important role in the economic sphere. Just as accounting records reflect the financial condition of an enterprise, the demographic service continuously monitors the population based on census data and systematic accounting. In a 1929 book, S.A. Novoselsky and J.C. Whipple consider the important issue of the reliability of statistical analysis. The authors refute the widespread opinion that statistical data can be interpreted in any way in favor of certain conclusions. In their opinion, erroneous or distorted results can only be obtained in two cases: if a person is poorly versed in data analysis or is deliberately trying to distort them. At the same time, statistical methods themselves are just a scientific tool that is not prone to manipulation. The authors believe that the main thing in obtaining accurate statistical results is to have proper knowledge of data analysis methods. A professional approach helps: to find errors in the source materials; to avoid mistakes in the processing of information; to obtain accurate and reasonable results. Thus, according to Novoselsky and Whipple, the accuracy of statistical conclusions depends on how well a specialist knows his business and correctly applies scientific methods of analysis [ibid].

The quality of demographic statistics assumes particular importance in the development of an algorithm for error detection in infant mortality data [Principles and recommendations... 2014].

Demographic monitoring is a fundamental tool for assessing the vital capital of the state, where every indicator, including infant mortality, requires the most accurate accounting and analysis. The need for high-quality and transparent statistics is enshrined in the principles of official statistics established by the UN General Assembly, which emphasizes the importance of reliable information to assess various aspects of the state's life, including demographic processes [Fundamental Principles... 2014]. Statistical analysis in the area of infant mortality should be based on strict adherence to methodological principles because the effectiveness of measures to improve the demographic situation depends on the correctness of the data collected and processed [Wilmoth 2007].

Semashko highlights the critical importance of reliable medical statistics in his seminal work «Essays on the organization of Soviet healthcare», emphasizing their role in identifying genuine issues within the healthcare system [Semashko 1947]. Since it is a comprehensive approach to the analysis of statistical measures, including careful quality control of the source data, regular verification of information and consideration of regional characteristics, allows not only to detect existing shortcomings in the data monitoring system but also to develop effective measures to reduce mortality, as well as to evaluate the effectiveness of the measures taken. Therefore, establishing stringent protocols for controlling the quality of source data is essential for generating trustworthy statistical outputs that underpin informed policymaking and sustainable improvements in public health outcomes.

Trust in high-quality and reliable data is paramount for successful health research, particularly in the context of infant mortality research. The issue of data anomalies attracts the close attention of the scientific community due to their negative impact on the validity of ongoing research. The classic work of Glasser shows that «a significant proportion of the recorded causes of death are incorrect or insufficiently complete («dirty codes»), which creates serious obstacles for further analytical work and reduces the scientific validity of the findings» [Glasser 1981]. Isermann expands the problem through an analysis of the mechanisms of errors in data caused by both sudden jumps and long-term accumulated deviations, noting the need for special algorithms for early detection of such phenomena [Isermann 2004]. Johnson et al. propose a strategy for processing incorrectly entered information, combining a wide range of analytical tools such as multiple cause of death analysis, negative correlation detection, and proportional analysis [Johnson 2021]. This approach improves the overall quality and reliability of data, creating the prerequisites for high-quality statistics and reliable analysis. In addition, Gassman et al. note that reducing the frequency of data anomalies depends on continuous professional training and training of those involved in data collection and processing [Gassman 1995]. Lack of qualifications and ignorance of the standard requirements for proper documentation leads to the repetition of typical errors, which negatively affects the quality of the source data.

The development of the algorithm allows not only to identify existing problems, but also to form a reasoned basis for evaluating current programs, planning future directions of healthcare development as well as improving the quality control system of medical care. The goal of the research is to develop a scientifically based and universal methodology in the form of an algorithmic approach for the effective detection of anomalies in infant mortality data, which allows to increase the accuracy of diagnosis and a comprehensive assessment of the health system efficiency in the Russian Federation. The algorithm for detecting anomalies in infant mortality data is a standardized methodology for the analytical processing of statistical information designed to detect inconsistencies and anomalies in data sets on infant mortality. The essence of the methodology is the consistent implementation of a set of analytical procedures aimed at a comprehensive assessment of data quality. The algorithm functions as an objective diagnostic tool for statistical arrays. Thus, the development of an algorithm for detecting anomalies in infant mortality is an actual and essential task that will make it possible to diagnose and record the state of the time-series indicators across the Russian Federation subjects for the period 2015–2024. Identifying errors may indicate inefficiencies in the healthcare system, specifically concerning the quality of medical care, resource optimization, forecasting, planning, and management. Our methodology considers theoretical, methodological, analytical, diagnostic, practical, evaluative, and interpretative aspects, ensuring a marked improvement in diagnostic accuracy and objectively assessing the effectiveness of the Russian healthcare system.

The paper introduces a universal methodology for infant mortality data quality assessment, notable for its distinctive methodological attributes:

- It introduces a comprehensive approach to diagnosing anomalies in infant mortality data.
- Missing data on the number of deceased infants and infant mortality rates were identified, and appropriate linear equation methods were employed to reliably impute these missing values.
- A special matrix structure has been developed, enabling the examination of infant mortality data in spatial and temporal dimensions. This structure facilitates comparison across regions and tracking of changes over time.
- The Z-score allows marking the matrix and visually displaying the dynamics of the rating of the Russian Federation subjects in time directly on the matrix itself.
- The universality of the proposed matrix structure methodology has been successfully demonstrated by applying the Z-score, which has shown its effectiveness in labeling and visualizing data, making it possible to clearly identify temporal and regional patterns of infant mortality.
- The proposed methodology provides a new opportunity for flexible matrix design, which allows you to individually customize the data structure depending on the research tasks and the specifics of the territory. This feature makes it possible to adapt to any changes in the information environment and increases the versatility of the proposed approach for analyzing infant mortality in various regions of the Russian Federation.

This research delivers a unified framework for error detection in infant mortality data, characterized by an original combination of methodological techniques and tools. The developed specialized methods of linear equations allowed to fill in the gaps in the source data and ensure the calculation of key indicators with high accuracy. The proposed matrix data structure provides new opportunities for spatial and temporal analysis, enabling detailed regional comparisons and effective tracking of change dynamics. The Z-score method was a unique tool for labeling and visualizing data, which made it possible to visualize changes and identify patterns in infant mortality rates [Fisher 1992]. The versatility and scalability of the proposed methodology was confirmed by its successful application on a large amount of data, demonstrating the readiness of the approach to solve diverse tasks and adapt to new conditions.

The proposed methodology, combining a comprehensive approach, specialized methods, a matrix structure, and the Z-score method, represents a robust and effective tool for analyzing infant mortality data, contributing to informed management decisions and enhancing the Russian healthcare system.

Materials and methods

Official statistical data from Federal State Statistics Service, processed and systematized by the researchers[1], served as the initial dataset for the analysis Federal State Statistics Service 2025[2].

Glossary of indicators:

A — the number of live births in the year n

B — the number of live births per year n–1

M — the number of children who died in the year n

K — infant mortality rate according to the formula $K = (x/A + y/B)1000$

x — children born and died in year n

y — children born in year n–1 and died in year n

Kr — infant mortality rate according to the formula $Kr = (M/N)1000$

N — the average number of registered live births who are at risk of dying during the period under review at the age of 1 year

Z-score according to the formula $Z = (X - \mu)/\sigma$

σ — standard deviation

X — the value of the indicator

μ — the average value of the sample

Our methodology involves analyzing regulatory documents, collecting and processing publicly accessible data, along with employing mathematical techniques for reconstructing and interpreting the information.

A review of Order No. 1687n of the Russian Ministry of Health and Social Development (dated December 27, 2011), which regulates medical birth criteria and birth certificate issuance procedures, revealed that it does not specify a formula for

1 Google Sheets: Data Sourse 2015–2024. Available at https://docs.google.com/spreadsheets/d/1r7v4cjQ6BIAUI79rUPjibg9m6xb7X99e/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

2 Federal State Statistics Service. Russia Natural Movement of the Population. Available at: http://ssl.rosstat.gov.ru/storage/mediabank/EDN_12-2024.htm (accessed 09.05.2025)

calculating the infant mortality rate. The document sets out the rules for confirming the fact of birth but does not regulate the procedure for calculating statistical indicators, which indicates a lack of legal regulation[1]. Until 2018, data on infant mortality was transmitted on the basis of statistical observation forms No. 1-U and No. 1-ROD. The forms were approved by the Federal State Statistics Service and served as the main data transmission channel, but they themselves did not contain formulas for calculating infant mortality rates (Federal State Statistics Service Order No. 339 dated July 23, 2015). Since 2018, legislative changes occurred, and Rosstat issued Resolution No. 756 on December 21, 2018, introducing a methodology for calculating infant mortality rates via the formula K = (x/A + y/B)1000[2]. In the middle of 2021, the formula for calculating the infant mortality rate is changing and is as follows: Kr = (M/N)1000[3]. Thus, the formula for calculating the infant mortality rate has two periods: until 2021 and after 2021. It is important to emphasize the existence of two fundamentally different periods in the methodology of official calculations: the period up to 2021, characterized by the use of formula K, and the period after 2021, where the formula Kr is used.

Data collection was the initial and important stage of the research, which was carried out on the basis of official information extracted from open sources, such as published reports of the Federal State Statistics Service. The initial stage involved collecting comprehensive information on all significant factors affecting infant mortality, including annual data on the number of births and deaths, as well as other related indicators [Kasasa 2021]. The entire set of collected data has been pre-prepared and checked for completeness and correctness. All available information was systematized and organized into a single table, which created a convenient format for subsequent analysis. With this approach, it was possible to identify cases of incomplete data that required additional investigation and supplementation. Application of a system of linear equations to recover missing data for a comprehensive analysis of infant mortality in the Russian Federation in the period 2015–2024.

1. The solution of finding x and y for the period 2015–2021 using the formula:

$$K = (x/A + y/B)1000$$
$$x + y = M \qquad (1)$$

The problem of restoring data on children born and died in year n; children born in year n-1 and died in year n is considered.

2. The solution of finding x, y, K for the period 2021–2024 using the formula:

$$K = (x/A + y/B)1000$$
$$K = (x/A)1000$$
$$M = x + y \qquad (2)$$

The task of restoring data on children born and died in year n, children born in year n-1 and died in year n, and the infant mortality rate according to the formula used in 2021 is being considered.

3. The solution of finding y≠0 for the period 2021–2024 using the formula:

$$(x + y)/(M/Kr) = (x/A + y/B) \qquad (3)$$

---

1 Russian Federation : Legislation of the Russian Federation : Ministry of Health and Social Development of Russia : Orders : O meditsinskikh kriteriyakh rozhdeniya, forme dokumenta o rozhdenii i poryadke yego vydachi [On medical criteria for birth, the form of the birth document and the procedure for its issuance] : dated December 27, 2011 N 1687n, as amended by Order of the Ministry of Health of the Russian Federation dated October 13, 2021 No. 987n. Available at https://normativ.kontur.ru/document?moduleId=1&documentId=418434 (accessed 01/19/2026). In Russ.

2 Russian Federation : Legislation of the Russian Federation : Rosstat : Orders : Ob utverzhdenii metodik rascheta zakreplennykh za Rosstatom pokazateley dlya monitoringa tselevykh pokazateley natsional'nykh proyektov (vmeste s «Metodikoy rascheta pokazatelya «Smertnost' ot bolezney sistemy krovoobrashcheniya (na 100 tys. naseleniya)», «Metodikoy rascheta pokazatelya «Smertnost' ot novoobrazovaniy, v tom chisle ot zlokachestvennykh (na 100 tys. naseleniya)», «Metodikoy rascheta pokazatelya «Mladencheskoy smertnost' (na 1 tys. rodivshikhsya detey)», «Metodikoy rascheta pokazatelya «Smertnost' naseleniya starshe trudosposobnogo vozrasta») [On approval of methods for calculating the indicators assigned to Rosstat for monitoring the targets of national projects (together with the «Methodology for calculating the indicator «Mortality from diseases of the system... (Appendix 3 Methodology for calculating the indicator «Infant mortality (per 1,000 born children)] : dated 21.12.2018 No. 756. Available at https://sudact.ru/law/prikaz-rosstata-ot-21122018-n-756-ob/ (accessed 01/19/2026). In Russ.

3 Russian Federation : Legislation of the Russian Federation : Rosstat : Orders : Ob utverzhdenii metodik rascheta pokazateley «Smertnost' detey v vozraste 0 - 4 goda na 1000 rodivshikhsya zhivymi» pomesyachno v godovom vyrazhenii, «Smertnost' detey v vozraste 0 - 17 let na 100 tys. chelovek sootvetstvuyushchego vozrasta» pomesyachno v godovom vyrazhenii i «Mladencheskaya smertnost' (na 1 tys. rodivshikhsya zhivymi)» pomesyachno v godovom vyrazhenii [On approval of methods for calculating the indicators «Mortality of children aged 0-4 years per 1,000 live births» monthly in annual terms, «Mortality of children aged 0-17 years per 100,000 people of the appropriate age» monthly in annual terms and «Infant mortality (per 1 thousand live births)» monthly in annual terms] : No. 375 dated 30.06.2021. Available at https://legalacts.ru/doc/prikaz-rosstata-ot-30062021-n-375-ob-utverzhdenii-metodik/ (accessed 01/19/2026). In Russ.

The problem of restoring data on children born in year n–1 and who died in year n is considered.

The use of the matrix concept for recording infant mortality data in the Russian Federation for the period 2015–2024 is a separate and independent stage of research aimed at systematization and structured presentation of information on infant mortality. The matrix concept of recording infant mortality data in the Russian Federation for the period 2015–2024 is an approach that allows a compact and efficient organization of a huge amount of information, which ensures simplicity and speed of data processing. The idea of the matrix is as follows: each cell of the matrix contains a single value corresponding to certain parameters, such as region, time interval and characteristics (birth rate, mortality, coefficients, etc.). Due to the presence of three dimensions (regions of the Russian Federation, time periods and indicators), it becomes possible to visually classify data, easily read and quickly access any value of interest. The essence of the proposed matrix model is to represent the source data as a single structure, where each cell composition is located at the intersection of a row and column, ensuring strict organization and immutability of the source data. This form of recording makes immediate access to all information available, eliminates the need for repeated access to sources, and greatly speeds up subsequent analysis and calculation operations. The Z-score method was employed to normalize and visualize infant mortality data for the Russian Federation's regions during the period 2015-2024. The procedure consisted of the following steps:

- Data preparation. The average values (μ) and standard deviations (σ) for each indicator (for example, infant mortality rate, number of births, etc.) were calculated for the entire sample.
- Calculation of z-values. For each indicator value for each subject and year, the Z-score was calculated using the formula $Z = (X - \mu)/\sigma$.
- Construction of the Z-score matrix. The calculated z-values were placed in a matrix, forming a ready-to-analyze data structure.
- Visualization. Based on the constructed data matrix, visualization was performed by filtering indicators and highlighting them with color. This process made it possible to visually assess the relative differences and dynamics of indicators between regions, years and indicators.

Leveraging the Z-score approach resulted in the normalization and graphic display of infant mortality data for the Russian Federation's territories over the period 2015–2024. Thanks to the calculation of averages and standard deviations for each indicator, as well as the construction of a matrix of Z-score, it was possible to create a structured database. Through the calculation of averages and standard deviations for each indicator, as well as the construction of a matrix of Z-score, it was possible to create a structured database. Color-coded visualization of the results provided a visual assessment of the relative differences and dynamics of indicators between regions. The result is a ready-made tool for further analysis of infant mortality data.

Results

As a result of the conducted research, important quantitative indicators were obtained in the corresponding tables and matrices. The results are presented in two-time intervals: the first covers the period from 2015 to 2021, the second — from 2021 to 2024.

During the research, the following tables and matrices were formed:

Period 2015–2021:

Source data — ABMK table 2015–2021[1]

Table of values «x,y» — xy table 2015–2021[2]

Infant mortality matrix — Infant mortality matrix 2015–2021[3]

Z-score matrix — Z-score matrix 2015–2021[4]

The period 2021–2024:

Source data — ABM table 2021–2024[5]

1 Google Sheets: ABMK table 2015–2021. Google Sheets. Available at https://docs.google.com/spreadsheets/d/1ssD4qCrEoGop_fNrY-HBA9IDmEaeNLKq/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

2 Ibid. xy table 2015–2021. Available at https://docs.google.com/spreadsheets/d/1myCc_hF6J2gZqNWRiYQTfYVlpmEzMYor/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

3 Ibid. Infant mortality matrix 2015–2021. Available at https://docs.google.com/spreadsheets/d/10rhH-iTZPQqoxjwswW3ak02V5XJOrWn1/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

4 Ibid. Z-score matrix 2015–2021. Available at https://docs.google.com/spreadsheets/d/1wfVWnmBtuUR6kMt-y--FIdb0Xo0eb-nU/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

5 Ibid. ABM table 2021–2024. Available at https://docs.google.com/spreadsheets/d/1C3pLoRq0KfdsiFR6R5MuQA3ne7QLNclp/edit?usp=sha

Table of values «xyK» — xyK table 2021–2024[1]

xNABMKr data table for finding «y≠0» — xNABMKr table 2021–2024[2]

Updated table of «y» values — y table 2021–2024[3]

Infant mortality matrix — Infant mortality matrix 2021–2024[4]

Z-score matrix — Z-score matrix 2021–2024[5]

The period 2015–2021. For the specified time period, the initial data was collected, formatted in the «ABMK table 2015–2021» table, which contains basic information for subsequent analysis. The following is a table of the found values of «x,y» necessary for a comprehensive analysis of infant mortality. Next, the initial data and the missing data were collected in the corresponding «Infant mortality matrix 2015–2021» matrix. Additionally, the «Z-score matrix 2015–2021» was built to identify the standard deviations of the indicators from the average.

The period is 2021–2024. For the specified time period, the initial data was collected, formatted in the «ABM table 2021–2024» table, which contains basic information for subsequent analysis. The following is a table of the found values of «xyK», necessary for a comprehensive analysis of infant mortality. Since the search procedure yielded results for y=0, the researchers were not satisfied with this result, and it was decided to search for values of y ≠ 0. For this purpose, a new table «xNABMKr table 2021–2024» was introduced. The results of the found values of «y» were formatted in a new table «y table 2021–2024». Next, the initial data and the missing data were collected in the corresponding «Infant mortality matrix 2021–2024» matrix. Additionally, the «Z-score matrix 2021–2024» was built to identify standard deviations of indicators from the average.

Discussion

This discussion focuses on key issues of demographic change in Russia from 2015 to 2024, covering important topics such as fertility reduction, infant mortality, statistical data generation, and future research prospects. We have identified a number of key points that allow us to better understand the current realities of the Russian demographic landscape:

No. 1 Negative dynamics in the number of live births in all constituent entities of the Russian Federation in the period from 2015–2024. The continuous decline in the number of live births in all regions of the Russian Federation from 2015 to 2024 demonstrates a steady negative demographic trend, which poses significant challenges to the long-term development of the country and the maintenance of its socio-demographic potential. The large-scale problem of reducing live births affects every region of the country, which indicates the systemic causes of this phenomenon, rooted in the socio-economic sphere. The long period of negative trends in the number of live births points to the need for an in-depth analysis of the effectiveness of existing population policy measures and the identification of factors influencing the reproductive behavior of the population. The common trend of decline in live births in all subjects of the Russian Federation indicates a commonality of reasons affecting the demographic situation regardless of regional specifics. The ten-year downward trend in the birth rate poses significant threats to the demographic well-being of the State and calls into question the sustainability of its further development. The universal reduction in birth rates across all regions of the country mirrors the depth of the demographic crisis and underscores the imperative for a nuanced exploration of its causes, tailored to the specifics of each territory.

No. 2 Specific combination of data. The conducted scientific analysis revealed the unique specifics of demographic trends characterized by a specific combination of statistical data. The results indicate a significant decrease in the average birth rate (A) by 40%, reflecting a significant weakening of the

ring&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

1 Ibid. xyK table 2021–2024. Available at https://docs.google.com/spreadsheets/d/1U7zLct_3-UJwbH5CsrltJvNcF5-QMEtI/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

2 Ibid. xNABMKr table 2021–2024. Available at https://docs.google.com/spreadsheets/d/1RZDzNq5GryrIzdJgGyQK73A3ZncTJIYS/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

3 Ibid. y table 2021–2024. Available at https://docs.google.com/spreadsheets/d/19RIhcGMSq2xopLaWGKIaGkzCKOC-1P3-/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

4 Ibid. Infant mortality matrix 2021–2024. Available at https://docs.google.com/spreadsheets/d/19Bof34lN9rlYj02KVGCrEIKbXncr9r6c/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

5 Ibid. Z-score matrix 2021–2024. Available at https://docs.google.com/spreadsheets/d/1v0sC5IbWpF1jJBdJXddTqDNTaw11ADgz/edit?usp=sharing&ouid=118281810080899529727&rtpof=true&sd=true (accessed 01/19/2026).

reproductive potential of the population. At the same time, there was a sharp decrease in the infant mortality rate (M) by 61.7%. The coefficients K and Kr were specially combined by averaging their values precisely within the framework of the specifics of this issue. This approach was based on the recognition of their equivalence and led to the formation of a single combined indicator. The final value of this indicator showed a decrease of 36% relative to the base period, which corresponded to a 2.5-point change in the complex coefficient index. Special attention is paid to the high-risk group (N), which accounts for only about 20% of the total number of deaths among children (M). This indicator is significantly lower than expected, which indicates possible anomalies in the monitoring and risk accounting system. There may be significant shortcomings in the identification of newborns belonging to the high-risk group, which reduces the accuracy of diagnosis and evaluation of the effectiveness of preventive measures. One likely factor is the underestimation of real risk due to the late detection of potentially dangerous health conditions in newborns. Another important aspect is the limited diagnostic tools, which lead to incomplete inclusion of newborns in the risk category, which creates a false impression of the state of medical care. In addition, the impact of management decisions aimed at artificially lowering the indicator N in order to improve the official statistical reports of the region or institution cannot be excluded. The technical component of the risk group calculation method also requires careful consideration, as there may be internal flaws in the methodology that contribute to a distorted perception of the scale of the problem. The negative consequences of an inaccurate risk assessment are numerous. Firstly, it creates an erroneous feeling of a prosperous situation in healthcare, which slows down the implementation of necessary improvements. The most critical effect is the risk of missing potentially life-threatening situations, which increases the number of unjustified child deaths and poses a threat to public health in general. Thus, the scientific findings underscore the urgent necessity for a comprehensive analysis of the mechanisms for registering and controlling risk among newborns, with the goal of enhancing the accuracy and sufficiency of methods for assessing children's health.

No. 3 Negative mortality: artifacts in the statistics (2015–2024). As a consequence of the research, unexpected negative mortality values (denoted as x and y) were discovered. These anomalous indicators demand particular attention from scientists and necessitate a thorough analysis and accurate interpretation. Such negative values typically arise from different sources and frequently stem from specific cases encountered during data processing:

- Data correction. Sometimes the facts of death are mistakenly entered in the subsequent reporting period, but later they are returned back to the previous period. As a result, there is a temporary discrepancy, which is recorded by the system as a negative value.
- Deleting duplicates. Due to a technical error or a human factor, the same death record may be duplicated. When duplicates are detected and deleted, the total number decreases, which creates a negative mortality effect.
- Recalculation of previous data. Regular data check includes re-evaluation and update of old records. When errors are detected in the previous count, corresponding corrections result in a decrease of the total sum and the appearance of negative numbers.
- Classification changes. In some cases, accounting standards and data classification change, which leads to recalculation of past data. New accounting rules sometimes force the exclusion of certain categories of persons from the death registry, causing short-term fluctuations, including negative values.

All these factors underline that such negative mortality values do not indicate the actual physical death, but rather are the result of complex statistical data processing and updating. Therefore, it is important to consider the specificity of such cases when analyzing demographic indicators.

No. 4 Detection of the paradox of convergence of K and Kr. In the process of analyzing statistical data on infant mortality, an unusual phenomenon was revealed, conventionally called the «paradox of convergence of indicators K and Kr». The phenomenon is characterized by the external continuity and consistency of the behavior of these indicators in the time series, despite the fundamental differences in the methods of their calculation before and after 2021. The essence of this phenomenon is that, despite the fundamental difference in methods of calculating the infant mortality rate before and after 2021, there is a seeming continuity and continuity of the dynamics of the indicator. Until 2021, the K coefficient was used, calculated using the formula K = (x/A + y/B)1000, and since 2021, a new indicator Kr has been introduced, defined as Kr = (M/N)1000. Despite completely different mathematical approaches and input parameters, visual analysis of time series creates the illusion of a smooth decrease in the indicator and maintaining the overall trend of improvement. This paradox is a statistical feature in which different mathematical models can generate comparable results with certain ratios of the initial data. There is a succession effect that can mislead researchers and analysts trying to track the dynamics of infant mortality without considering methodological changes. It is important to understand that the apparent decrease in the Kr indicator, similar to the dynamics of the previous K indicator, does not necessarily reflect real changes in the infant mortality rate. This may be the result of a specific interaction of parameters in new and old formulas, creating the illusion of continuity of the time series. Awareness of this paradox is important for the correct interpretation of statistical data and requires special attention when conducting both a retrospective and prospective analysis of infant mortality. Methodological differences should always be considered when comparing indicators from different periods and avoiding hasty conclusions based on a superficial analysis of time series.

No. 5 Practical significance of the research. The detected data anomalies are of great practical value, especially in the context of the development and implementation of effective programs to improve the health of newborns.

The main direction of practical application is in the following areas:

- Effective elimination of identified anomalies. The research findings enable the detection of data anomalies and provide specific measures for their elimination. The implementation of proposals to improve recording and processing procedures will create conditions for preventing the accumulation of systematic errors, ensure the reliability of information and lead to an improvement in the quality of statistical materials used by health authorities.
- Evaluation of the effectiveness of government programs. The improved data collection and analysis system will allow specialists to more accurately assess the effectiveness of existing newborn prevention and support programs. A complete and reliable database will become the basis for developing high-quality strategic plans, allocating resources, and developing effective solutions aimed at improving children's health.

Consequently, the practical recommendations outlined in this research will have a direct beneficial impact on the healthcare system, facilitating informed decision-making and contributing to major achievements in safeguarding the health of mothers and children.

Conclusion

Thus, the article presents a new development of an evidence-based methodology for an algorithmic approach to identify anomalies in data on infant mortality in Russia for the period 2015–2024. The research involved a comprehensive analysis, which comprised regulatory research, the use of linear equations, the construction of a matrix data structure, and the application of the Z-score method. The analysis was based on official Rosstat statistics. The developed universal methodology for detecting data anomalies in raw data combines a comprehensive approach to diagnosis, specialized linear equation methods for imputation of missing data, a matrix structure for spatiotemporal analysis and data standardization as well as a Z-score for visualizing results. This analysis identified key demographic trends: a steady decline in the number of live births in all regions of Russia, specific features of the infant mortality rate (until 2021 — coefficient K, after 2021 — coefficient Kr) and presence of anomalous values. The research offers an integrated approach to diagnosing anomalies in infant mortality data that combines original methodological techniques and analytical tools. The practical significance of the developed methodology is manifested in improving the accuracy of data defect diagnosis, optimizing the management decision-making system in healthcare and improving quality control mechanisms for medical care. This research primarily targets the identification of systemic issues in data collection and processing, alongside the development of effective methods to mitigate and eliminate these problems.

The algorithm for detecting anomalies in infant mortality data is an integral tool of the demographic monitoring system, allowing:

- to ensure the reliability of statistical indicators;

- to identify distortions in the source data;
- to prevent possible errors in the interpretation of the results;
- to form an objective picture of demographic processes.

Methodological competence of demographic statisticians and the use of modern data processing algorithms allow to minimize the risks of misinformation and ensure the validity of findings in infant mortality research. Thus, the development and implementation of effective algorithms for detecting data anomalies is not only a technical tool, but also an important analytical tool of the public administration system that facilitates informed decision-making in the field of maternal and child health. The practical significance of the research is reflected in the possibility of applying its findings to optimize the functioning of the healthcare system and implement measures aimed at improving the demographic situation in the Russian Federation. The implementation of the proposed recommendations contributes to the formation of a balanced policy in the field of protecting the health of women and children.

Based on the results obtained, the following recommendations are formulated:

- It is necessary to standardize procedures for collecting and processing demographic data
- to ensure their comparability and reliability.
- It is necessary to automate the processes of detecting anomalies in statistical materials, which will significantly improve the effectiveness of monitoring.
- It is important to strengthen quality control of statistical information at all stages of its formation.
- It is necessary to improve the skills of specialists in the field of demographic statistics to ensure the methodological correctness of research.

The prospects for further research are related to the development of a methodology for integrated monitoring of demographic processes. Special attention should be paid to improving algorithms for processing large amounts of data and integrating the defect detection system with other healthcare information platforms. Creating a common information environment for the exchange of data among population monitoring actors will improve decision-making and provide more accurate projections of demographic indicators. The development of innovative forecasting approaches based on an improved statistical database will open up new opportunities for the formation of effective state policy in the field of demographic development.

Thus, the research emphasizes that the quality of demographic statistics is a defining factor in assessing the demographic situation and developing measures to enhance it. The introduction of modern data analysis methods and the improvement of the monitoring system will significantly improve the effectiveness of management decisions in the healthcare sector.

Author contributions:

V. S. Rusova: conceptualization, algorithm development, article authorship;

M. R. Bikchenteeva: linguistic component — professional translation and formatting of the article.

Вклад авторов:

В. С. Русова: концептуализация, разработка алгоритма, авторство статьи;

М. Р. Бикчентеева: лингвистическая составляющая — профессиональный перевод и оформление.

References

Burstein, R., Henry, N. J., Collison, M. L., et al. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. DOI:10.1038/s41586-019-1545-0. *Nature,* 2019; 574(7778): 353–358.

Fisher, R. A. Statistical methods for research workers. DOI: 10.1007/978-1-4612-4380-9_6. *Breakthroughs in statistics: Methodology and distribution* (pp. 66–70). New York : Springer, 1992. ISBN: 978-0-387-94039-7.

Fundamental Principles of Official Statistics, General Assembly Resolution A/RES/68/261 (29 January 2014). *UN* : official website. Available at https://unstats.un.org/fpos/ (accessed 12/16/2025).

Gassman, J. J., Owen, W. W., Kuntz, T. E., et al. Data quality assurance, monitoring, and reporting. DOI: 10.1016/0197-2456(94)00095-k. *Controlled clinical trials*, 1995 Apr;16(2 Suppl):104S–136S.

Glasser, J. H. The quality and utility of death certificate data. DOI: 10.2105/ajph.71.3.231. *American journal of public health*, 1981 Mar;71(3):231–3.

Isermann, R. Model-based fault detection and diagnosis – status and applications. DOI:10.1016/S1474-6670(17)32149-3. *IFAC Proceedings Volumes*, 2004; 37(6), 49–60.

Johnson, S. C., Cunningham, M., Dippenaar, I. N., et al. Public health utility of cause of death data: applying empirical algorithms to improve data quality. DOI: 10.1186/s12911-021-01501-1. *BMC medical informatics and decision making*, 2021 Jun 2;21(1):175.

Kasasa, S., Natukwatsa, D., Galiwango, E., et al. Birth, stillbirth and death registration data completeness, quality and utility in population-based surveys: EN-INDEPTH study. DOI: 10.1186/s12963-020-00231-2. *Population Health Metrics*, 2021 Feb 8;19(Suppl 1):14.

Liu, L., Johnson, H. L., Cousens, S., et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. DOI: 10.1016/S0140-6736(12)60560-1 *The lancet*, 2012 Jun 9;379(9832):2151–2161.

Principles and recommendations for a vital statistics system (Rev. 3). New York : United Nations, 2014. 240 p. ISBN: 978-92-1-161572-2. Available at https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/CRVS/M19Rev3-E.pdf (accessed 12/16/2025).

Semashko, N. A. *Ocherki po teorii organizatsii sovetskogo zdravookhraneniya* [Essays on the Theory of Soviet Health Care Organization], Moscow : Izdatel'stvo Akademii Meditsinskikh Nauk SSSR Publ., 1947. 46 p. (in Russ.)

Whipple, J. C. and Novoselsky, S. A. *Osnovy demograficheskoy i sanitarnoy statistiki* [Fundamentals of demographic and sanitary statistics] : a handbook for medical universities, Moscow : State Medical Publishing House, 1929. 683 p. Available at https://www.demoscope.ru/weekly/knigi/novoselskij/novoselskij_1929.pdf (accessed 12/16/2025) (in Russ.).

Wilmoth, J. R., Andreev, K., Jdanov, D., et al. *Methods protocol for the human mortality database*. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock, 2007. 83 p. Available at https://www.academia.edu/23408295/Methods_Protocol_for_the_Human_Mortality_Database (accessed 12/16/2025).

World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals. Geneva : World Health Organization; 2023. 119 p. ISBN: 978-92-4-007433-0.

Information about the authors:

Rusova Violetta S.— independent researcher and public and municipal administration specialist, self-funded (no institutional affiliation), SPIN-code: 9539-6059, AuthorID: 1120135 (Moscow, Russia);
Bikchenteeva Mariya R. — independent researcher, English language instructor (Saint Petersburg, Russia).

Информация об авторах:

Русова Виолетта Сёмавна — независимый исследователь и специалист в области государственного и муниципального управления, самофинансирование (без институциональной принадлежности), SPIN-код: 9539-6059, AuthorID: 1120135 (Москва, Россия); **Бикчентеева Мария Рашидовна** — независимый исследователь, преподаватель английского языка. (Санкт-Петербург, Россия).

The article was submitted 02/05/2026; approved after reviewing 02/20/2026; accepted for publication 02/27/2026.

Статья поступила в редакцию 05.02.2026; одобрена после рецензирования 20.02.2026; принята к публикации 27.02.2026.