ENGLISH
SELECTION
2022

THE ELECTRONIC JOURNAL

DEMOGRAPHIC REVIEW

HSE UNIVERSITY

ISSN 2409-2274

FOUNDED BY ANATOLY G. VISHNEVSKY (1935-2021) IN 2014.

Released quarterly. Published since 2014.

All manuscripts are obligatory peer-reviewed.

The journal is registered on October 13, 2016 in the Federal Service for Supervision of Communications, Information Technology, and Mass Media.
Certificate of Mass Media Registration ЭЛ № ФС77-67362.
ISSN 2409-2274

Editorial address

Bolshoy Trekhsvyatitelskiy lane 3, office 303, Moscow, 109028, Russia
Phone: 8-495-772-95-90 * 11864 / *11824
www.demreview.hse.ru
demreview@hse.ru

Original papers

Spatial patterns of age-specific sex ratios in Russian intraregional migration

Aleksandr Gerasimov

THE ORIGINAL ARTICLE IN RUSSIAN WAS PUBLISHEDIN DEMOGRAPHIC REVIEW IN 2022, 9(1), 91-107.
https://doi.org/10.17323/demreview.v9i1.14575

Demographic features of industrial serfs in the 18th-19th centuries: parish registers of the village of Kudinovo in the Bogorodsky district, 1777-1862

Vlas Ryazanov

THE ORIGINAL ARTICLE IN RUSSIAN WAS PUBLISHEDIN DEMOGRAPHIC REVIEW IN 2022, 9(2), 4-21.
https://doi.org/10.17323/demreview.v9i2.16203

Digital traces of the population as a data source on migration flows in the Russian Arctic

Andrey Smirnov

THE ORIGINAL ARTICLE IN RUSSIAN WAS PUBLISHEDIN DEMOGRAPHIC REVIEW IN 2022, 9(2), 42-64.
https://doi.org/10.17323/demreview.v9i2.16205

Excess summer mortality at young ages: do long school holidays matter?

Polina Kuznetsova, Tatyana Maleva

THE ORIGINAL ARTICLE IN RUSSIAN WAS PUBLISHEDIN DEMOGRAPHIC REVIEW IN 2022, 9(2), 81-101.
https://doi.org/10.17323/demreview.v9i2.16207

Who needs (or doesn't) three years of parental leave?

Irina Kalabikhina, Polina Kuznetsova

THE ORIGINAL ARTICLE IN RUSSIAN WAS PUBLISHEDIN DEMOGRAPHIC REVIEW IN 2022, 9(3), 24-43.
https://doi.org/10.17323/demreview.v9i3.16468

Assessment of reliability of results of the 2019 Belarus population census based on an analysis of changes in the ethnolinguistic composition of the population

Aleksandr Sokolov

THE ORIGINAL ARTICLE IN RUSSIAN WAS PUBLISHEDIN DEMOGRAPHIC REVIEW IN 2022, 9(4), 61-103.
https://doi.org/10.17323/demreview.v9i4.16744

Демографическое обозрение / Demographic Review
2022;9(1)91-107 DOI: https://doi.org/10.17323/demreview.v9i1.14575

Spatial patterns of age-specific sex ratios in Russian intraregional migration

Aleksandr A. Gerasimov
(aagerasimov_4@edu.hse.ru), HSE University, Russia.

Abstract: *Migration is one of the key processes that shape the age and sex structure of a population. Intraregional migration, which accounted for 39% of migration turnover in Russia in 2016-2020 and follows the centripetal trend, has a significant impact on the population of many territories. The study of sex disparities in migration deepens the understanding of its impact on the age-sex population structure, fertility and marriage patterns, and local labor markets. This article aims to identify the differences in the participation in intraregional migration between men and women, using data on the age-sex population distribution and the number of arrivals and departures by age and sex in 2016-2021 in Russian municipalities. Based on these primary data, the coefficients of arrivals and departures per 1000 residents for individual age and sex groups for 2016-2020 were calculated. Also, the corresponding standardized coefficients were calculated for the ages of 15-29 and 40+ years. The case of Central Russia was considered in detail. The results revealed that women show higher migration activity at the age of up to 40 years, while men do so at the age of over 40 years and in all types of territories. The higher migration activity of women at young ages is associated with their greater involvement in educational migration and fewer employment opportunities in peripheral territories, as well as with a potentially more frequent change of registration when getting married, having children, and enrolling them in preschool institutions and schools. As for men, their higher migration activity at older ages is probably due to the fact that they less often than women take care of children alone after a divorce or of elderly relatives, which makes them less restrained in terms of migration.*

Keywords: *intraregional migration, migration structure by sex and age, core, periphery, municipal data.*

For citation: *Gerasimov, A. A. (2022). Spatial patterns of age-specific sex ratios in Russian intraregional migration. Demographic Review, 9(1), 91-107.* https://doi.org/10.17323/demreview.v9i1.14575

Introduction

The sex ratio is regarded as one of the most crucial indicators of the demographic situation (Zayonchkovskaya, 1991). Migration plays a pivotal role in shaping this ratio in numerous territories. Gender imbalance in migration alters the population composition in areas of both departure and arrival, consequently influencing birth rates, marriage rates, and local labor markets. An analysis of migration structure by gender is imperative to gain a comprehensive understanding of population dynamics and demographic processes.

The study of gender-specific migration patterns dates back to the pioneer of migration research, E. Ravenstein (Ravenstein, 1885). He observed that women exhibit higher migration activity over short distances, while men tend to do so over longer distances. In the context of Russian statistical publications, intraregional migration refers to moves over short distances, while interregional and international migration encompasses long-distance moves. Recent statistics from Rosstat (2021) confirm the prevalence of women in intraregional migration, especially among the youngest and most mobile age groups.

Intraregional migration holds particular interest not only due to its greater gender disproportions compared to other migration flows but also because it largely adheres to a center-peripheral hierarchy of territories. Accounting for 39% of migration turnover in Russia between 2016 and 2020 (Rosstat, 2021), intraregional migration significantly influences the age-sex population structure throughout vast areas of the country.

This article aims to identify differences in the migration activity of men and women regarding intraregional moves. The data were analyzed based on individual age groups within a spatial context.

Previous research

Intraregional migration in Russia

In Russian statistical publications, intraregional migration refers to moves between municipalities of a region. Such moves are recorded when individuals change their place of residence registration or when a temporary registration is obtained or has expired. The classification of moves adheres to the grid of administrative and municipal territorial divisions, following all of their peculiarities. For instance, migration spanning hundreds of kilometers within large Siberian regions is categorized as intraregional, whereas a move to the opposite side of the street in Moscow and St. Petersburg metropolitan areas, if the regional border follows it, is considered interregional migration.

Changes in migration between urban cores and the periphery are influenced by the stages of urbanization (Nefedova, Treyvish 2002), which result in varying degrees of centripetal and centrifugal flows. In Russia, in contrast to other European countries, there is a pronounced trend towards spatial concentration (Karachurina, Mkrtchyan 2016). This trend is not driven solely by internal migration, but also by a gradual increase in natural population decline observed when moving from cores to the periphery. However, this gradient is somewhat disrupted by local secondary and tertiary centers (Karachurina, Mkrtchyan 2013).

Intraregional migration leads to the concentration of population in regional cores and their adjacent municipalities (Karachurina, Mkrtchyan, Petrosyan 2021), although there are several examples of noticeable alternative centers (Cherepovets in Vologda Oblast, Surgut and Nizhnevartovsk in KhMAO, and Sochi in Krasnodar Krai).

Migration selectivity by gender

The characteristics of migration based on gender remain less explored than age-based migration studies (Mkrtchyan, 2021). Nevertheless, differences in migration patterns between men and women are evident across various aspects, including the reasons and consequences of migration, its intensity, and its spatial and temporal features.

Gender-based differentiation in migration patterns is increasingly influenced by economic causes. In developed countries, a connection has been shown between higher female mobility and gender imbalances in the labor market, such as variations in employment structures and wage levels (Faggian, McCann, Sheppard, 2007). In East Germany, where the gender ratio spatial patterns resemble those in Russia (with a disproportionately higher number of men in rural areas and women in cities), women's higher migration activity is also driven by labor market inequalities (Leibert, 2016).

Other studies also show the presence of gender selectivity of migration at certain spatial levels in developed countries. For example, in Sweden, women living in rural areas are more actively involved in educational migration than men, which enhances their chances of moving to pursue education and subsequently entering the labor market or getting married (Johansson, 2016; Karpestam, Håkansson, 2021). Indirectly, Russian data (Bessudnov, Malik, 2016) also suggest that females are more likely to enter the 10th grade (high school), thereby increasing their opportunities to access universities primarily located in regional centers.

On the contrary, at older ages, the relatively higher migration activity of men may be associated with the consequences of divorce (Clark, 2013; Mkrtchyan, 2021). In Russia, children typically stay with their mothers after divorce, which can hinder their mobility. In addition, in most cases, women are responsible for taking care of their elderly parents.

Core-peripheral spatial relations in Russia

The Russian space exhibits considerable polarization, with stark socio-economic contrasts between large cities, small towns, and rural areas (Nefedova, 2009). This spatial heterogeneity results in varying living standards, decreasing hierarchically from centers to the remote periphery (Zubarevich, 2013). Centripetal migration, which has been ongoing for decades, plays a vital role in this dynamic and includes intraregional moves (Karachurina, Mkrtchyan 2016).

Centripetal migration, along with its consequences, alters the age-sex structure of the population in the territories involved. Regional cores tend to have a younger population compared to the periphery, with a sex ratio disproportionately favoring women due to the gender selectivity of centripetal migration (Kashnitsky, 2014; Gerasimov, 2022). The most notable core-peripheral differences are observed in the Center and the North-West regions, as well as in Siberia.

Educational institutions and the availability of more diverse services, leisure activities, and labor markets attract migrants from peripheral territories to large centers. However, the scale of the increasing concentration of the country's population in major urban cores and their metropolitan areas may be underestimated. Many migrants who have left peripheral areas and have been residing elsewhere for an extended period may still be registered at their former place of residence, thus inflating the population figures for rural areas and small towns (Fomkina, 2017; Alekseev, Vorobyov, 2018). Consequently, the actual rate of out-migration from the periphery may be even higher than indicated in statistical data.

Study limitations

This study faces several limitations due to the migration registration rules in Russia and the quality and nature of the available data:

- Russian migration statistics exhibit significant distortions that complicate analysis. Migration registration practices affect data on moves differently for men and women. Events like childbirth and marriage often prompt changes in registration for women (Mkrtchyan, 2021), usually following an extended period of residence at a place different from the registered address.
- Migration registration in Russia creates the statistical artifact of "pseudo-return" migration (Mkrtchyan, 2020). This refers to the automatic "return" of migrants with expired registration from their current place of stay to their original place of residence. Consequently, the number of arrivals in donor municipalities and departures in recipient ones is inflated.
- Men's higher involvement in circular labor migration, predominantly interregional (Mkrtchyan, Florinskaya, 2019), which is not reflected in migration statistics (Mkrtchyan, 2021), leads to the underrepresentation of men in intraregional migration in peripheral municipalities where circular labor migration is most common.
- The institutional population, including the armed forces and prisoners, is registered at their place of stay (Pyankova, 2014), but their moves are recorded in migration statistics only on rare occasions[1]. As institutional groups are predominantly male, this leads to an underestimation of male migration activity indicators in their location municipalities.
- Evasion of military service by draft-age men may further hinder the analysis of statistical data, as some men, after moving, may not be registered at their actual address.
  This general limitation arises from the assumption that Russian migration statistics record changes in registration, not actual moves.
- The study excludes federal cities and adjacent regions due to the consideration of migration between them as interregional rather than intraregional, making comparisons with other regions problematic. Moreover, intraregional migration in federal cities consists of moves between city districts, which are not comparable to municipalities in other regions. In addition, due to missing data for 49 municipalities across the country, they also were excluded from the calculations, either entirely or in part.
- To achieve a more accurate spatial picture, longer statistical series are required to eliminate random fluctuations caused by a small number of observations. Using earlier data is challenging due to the short time series in the Municipal Indicators Database (BDPMO) and the data's incomparability resulting from changes in the statistical recording of migration in 2011. After this, migration flow volumes stabilized only by 2015-2016 (Mkrtchyan, 2020).

[1] According to Rosstat instructions, migration statistics take into account the moves of military personnel serving under contract. Conscripts and prisoners are counted only if they changed their place of residence after completing their service or serving their sentence respectively.

Data and methods

The primary data source used in this study is the Municipal Indicators Database (Rosstat, 2022a). Territorial cells consist of the first-level municipal entities, including municipal districts and urban and municipal okrugs. We utilized data on the population of municipalities by sex and age at the beginning of each year for the period 2016–2021, along with data on the annual number of arrivals and departures by sex and age from 2016 to 2020.

The indicators for the five-year period were averaged, and the focus of this study was not on analyzing dynamics. The fluctuating sizes of age cohorts among residents and migrants in sparsely populated municipalities and the limited time series of data made it impractical to analyze the time series of indicators . Averaging the indicators over a five-year period enabled us to observe a relatively stable picture of migratory flows. Notably, significant changes in migration activity in Russia in 2020 due to the COVID-19 pandemic did not impact the sex ratio in migration flows (Mkrtchyan, 2021), allowing us to include the data for that year on equal terms with the rest.

The analysis was conducted for 2,164 municipal entities (MEs) within their boundaries for 2016. Some municipalities underwent administrative transformations during the study period, such as transforming from municipal districts into urban okrugs. In such cases, statistical series for transformed municipalities were combined. For calculations concerning unified or divided MEs, series before their conversion were used if their length was sufficient (at least three years).

The original database contained a significant number of gaps. To address population data them, we filled them with mean values when data were available for the previous and following years. For migration data, gaps were categorized as "systemic," affecting municipalities in all age groups for specific years, and "other", concerning sparsely populated municipalities and/or older age groups (i.e., the least mobile). We replaced these gaps with zeros to enable calculations, as described below.

For the municipalities analyzed, we calculated coefficients of arrivals and departures per 1000 inhabitants of the respective sex in the following age groups: 0–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70+ years old. The age group boundaries align with the grouping of age cohorts in the population data source. We first calculated the average population for 2016–2020 as the mean between the figures for January 1 of the subsequent two years. Then, we calculated the coefficients for each year; if there was a gap in the series for at least one of the indicators, the year was excluded. The obtained indicators were averaged over a five-year period. However, if there were more than two gaps in the coefficients' series, we did not make the calculation.

For cartographic analysis, we selected two enlarged age groups of particular interest: 15–29 years old, representing the age with the highest migration activity, and 40+ years old, the age range with higher migration activity of men. To account for differences in the age composition of men and women within these age groups, we standardized the indicators using the average weight for the period under review of the initial age groups in the Russian population.

For a more detailed examination of spatial differences in migration activity, we focused on Central Russia, where well-established and pronounced center-peripheral relations of territories are observed. The municipalities were categorized into four groups based on calculations by A. Raysikh (Raysikh, 2019), who identified gravity zones of large settlements in

Russia using the gravitational method. The four groups included regional centers with suburban areas (28 MEs), large cities with a population of over 100 thousand people and their suburban areas (10 MEs), local centers with a population of less than 100 thousand and their adjacent areas (58 MEs), and other municipalities (251 MEs). The analysis excluded municipalities belonging to the Moscow agglomeration, as well as territories where institutions of the Federal Penitentiary Service and large contingents of the armed forces are located[2] (except for large cities, where their influence on the age and sex structure of the population is insignificant). Population sizes and the number of arrivals and departures were aggregated for the aforementioned four groups, enabling us to obtain coefficients of arrivals and departures similar to calculations for individual municipalities.

Results

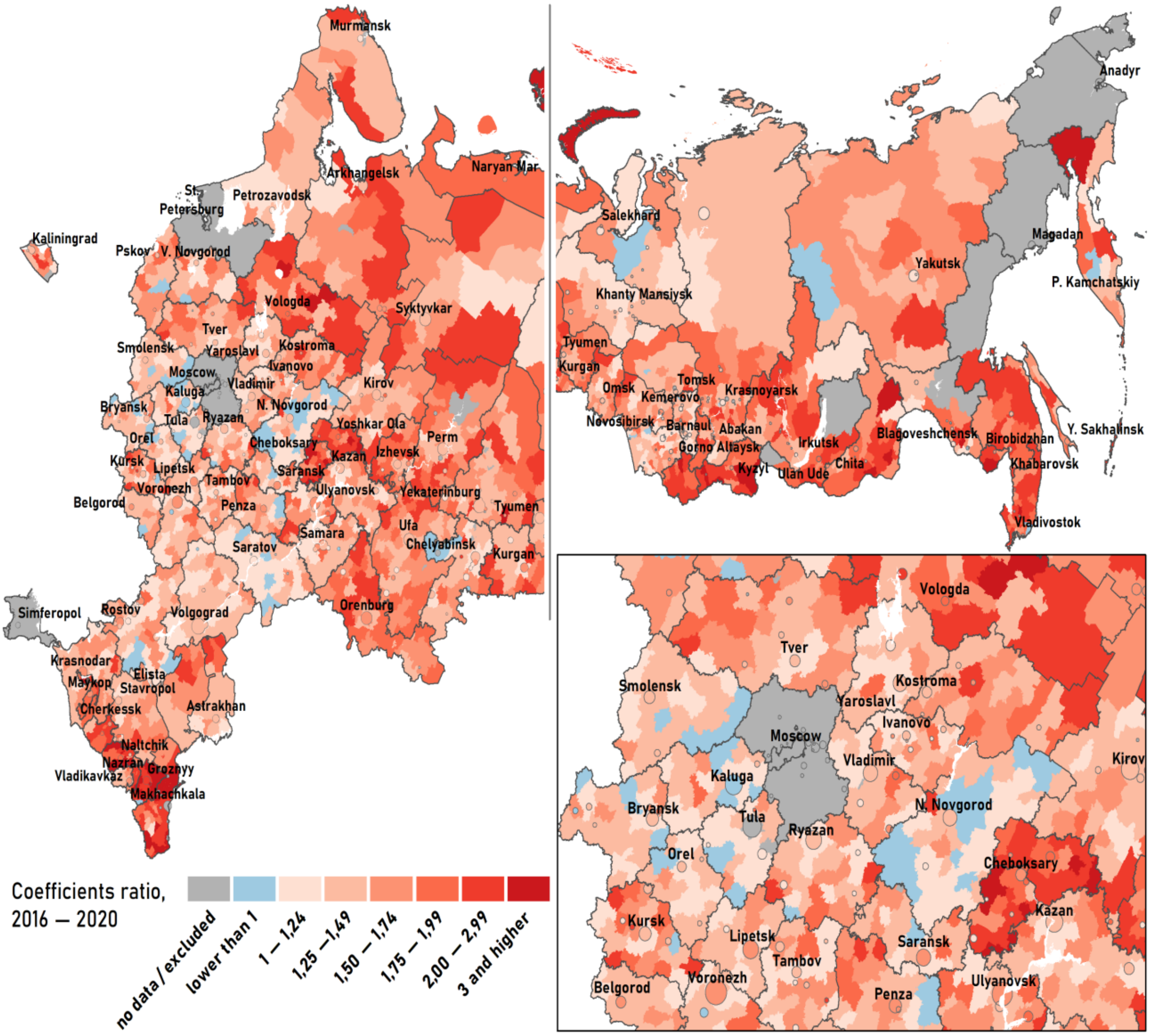
By dividing the age coefficients of female arrivals and departures by the corresponding coefficients for men, we were able to compare the migration activity by municipalities and obtain the following results.

Arrivals aged 15–29. Figure 1 displays the coefficient ratios for arrivals aged 15–29, revealing a higher migration activity of women in the majority of municipalities. In numerous peripheral territories, arrivals in this age group are a "reflection" of departures associated with educational migration, influenced by migration registration peculiarities. Large cities, attracting centripetal flows, exhibit a more intensive arrival of women. These findings align with studies on individual regions. For example, G. Leonidova and N. Vyacheslavov (2016), using data from Vologda Oblast, found that women account for up to two-thirds of all intraregional moves at young ages.

The largest disproportions in favor of women are observed in the North, Far East, and North Caucasus regions. In the first two macro-regions, and some municipalities in other parts of the country, the higher migration activity of women is linked to the presence of military contingents, artificially lowering the migration activity of local men due to the registration of the institutional population at the place of stay without recording its moves in migration statistics. For the North Caucasus, limitations arise from noted problems with demographic statistics (Kazenin, 2014; Mkrtchyan, 2019).

[2] Lists of penal colonies and open prisons are provided on the websites of the regional departments of the Federal Penitentiary Service. The location of large military units was determined according to open data, checked with municipal statistics.

Figure 1. Coefficient ratios of arrivals for persons aged 15–29 per 1,000 residents of the corresponding sex (female/male)[3]



Source: based on the author's calculations

Noteworthy are the two Volga republics, Chuvashia, and Mari El. In most municipalities within both regions, there is significantly higher migration activity of women compared to men, visible not only in arrivals but also in departures (Figure 2). Calculations for five-year age groups indicated that women's peak migration activity in most municipalities occurs in the 25–29 years age group, while for men, it is in the 15–19 years age group. Data for municipalities with the most significant disparities are presented in Table 1.

[3] Here and below, on the maps, urban okrugs are presented with generalized boundaries for better representation at scale.

Table 1. Population and arrivals per year in the age group 25-29 in several districts of Mari El and Chuvashia republics

	District	Population at start of year						Arrivals over year				
		2016	2017	2018	2019	2020	2021	2016	2017	2018	2019	2020
Women	Kuzhenerskiy [4]	283	208	126	87	91	143	35	35	29	27	25
	Alikovskiy	260	134	53	25	81	157	42	33	23	27	19
	Krasnochetayskiy	195	76	27	19	31	82	24	28	41	21	25
	Shemurshinskiy	188	99	13	30	82	142	25	20	24	16	17
	Shumerlinskiy	103	90	31	17	41	69	21	15	11	10	9
	Yal'chikskiy	79	17		42	95	205	40	34	39	39	39
Men	Kuzhenerskiy	492	405	351	308	286	306	15	11	7	8	13
	Alikovskiy	575	485	338	314	319	392	22	16	28	20	10
	Krasnochetayskiy	538	462	376	314	277	336	32	22	12	21	12
	Shemurshinskiy	425	359	245	185	183	180	13	10	12	8	6
	Shumerlinskiy	278	199	185	169	179	215	8	8	3	8	
	Yal'chikskiy	382	321	220	244	291	389	27	22	25	12	10
Sex ratio, men per 100 women	Kuzhenerskiy	174	195	279	354	314	214	43	31	24	30	52
	Alikovskiy	221	362	638	1256	394	250	52	48	122	74	53
	Krasnochetayskiy	276	608	1393	1653	894	410	133	79	29	100	48
	Shemurshinskiy	226	363	1885	617	223	127	52	50	50	50	35
	Shumerlinskiy	270	221	597	994	437	312	38	53	27	80	
	Yal'chikskiy	484	1888		581	306	190	68	65	64	31	26

Source: Compilation based on BDPMO data (Rosstat 2022a).

In the districts presented in the table, as well as in other municipalities of the two republics, there is a significant disproportion in the sex ratio in favor of men, exceeding the values found in peripheral municipalities of Central Russia (120–130 men per 100 women (Gerasimov, 2022)). The reason for this anomaly is unclear, but its clear localization within the regions suggests problems with the statistical recording of migration events at the municipal level.

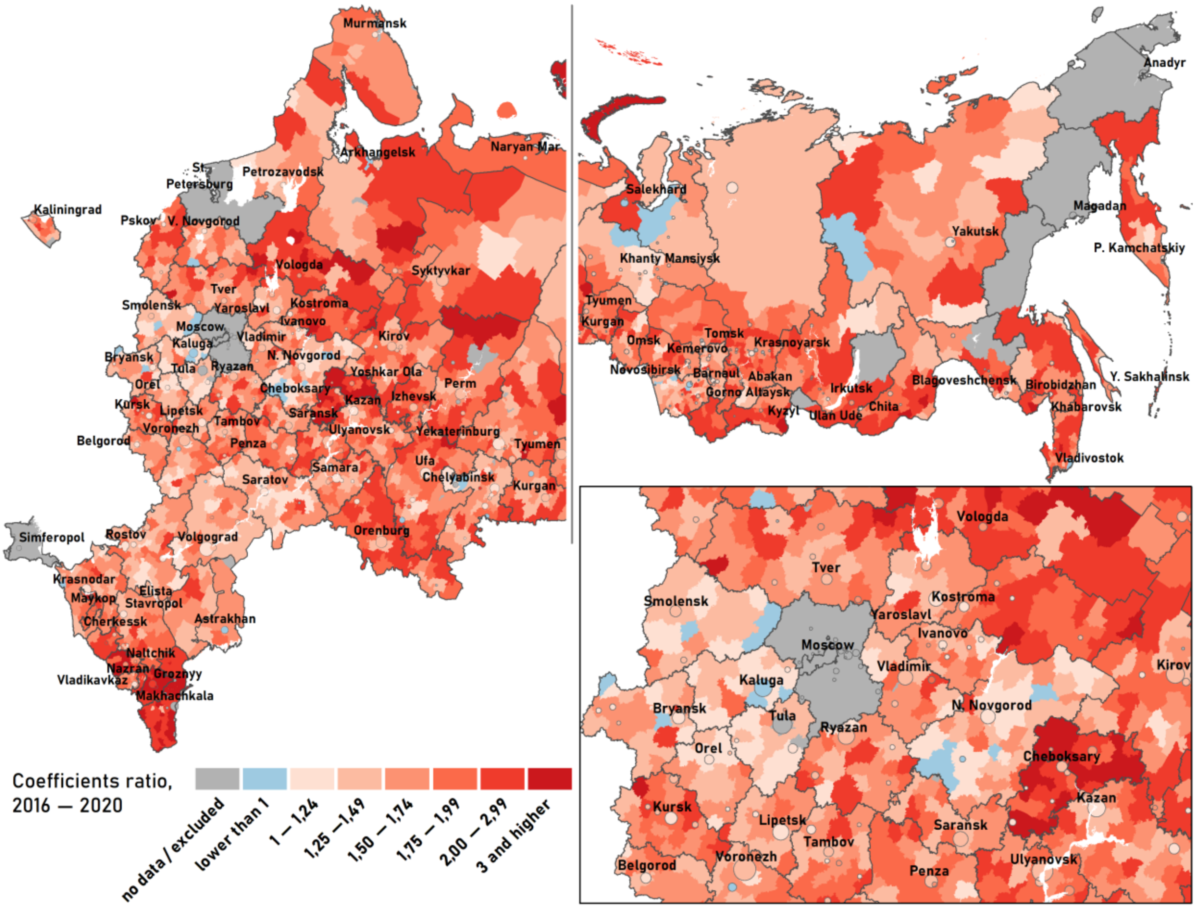
Departures aged 15–29. Compared to arrivals throughout the country, there is an even higher migration activity of women in out-migration (Figure 2). For certain regions (the Republics of Altai, Tyva, and Buryatia), the intensity of migration at young ages may also be influenced by the popularity of contract service in the armed forces, which involves interregional migration of some men. Service in the armed forces acts as a social lift and one of the few reliable sources of income for many peripheral territories.

Arrivals aged 40 and over. For both arrivals (Figure 3) and departures (Figure 4), there is a more mosaic spatial pattern compared to ages 15–29. This is due to lower migration activity at older ages and, consequently, the influence of small numbers on the results, despite attempts to mitigate it in the calculations described above (averaging indicators over 5 years and aggregating

[4] The Kuzhenerskiy municipal district is located in the Republic of Mari El, while the rest are in the Chuvash Republic.

age groups). As a result, more intensive arrivals among women of these ages, as well as a significant disproportion in favor of men, are typically found in sparsely populated municipalities.

Figure 2. Coefficient ratios of departures aged 15–29 per 1,000 residents of the corresponding sex(female/male)



Source: based on the author's calculations

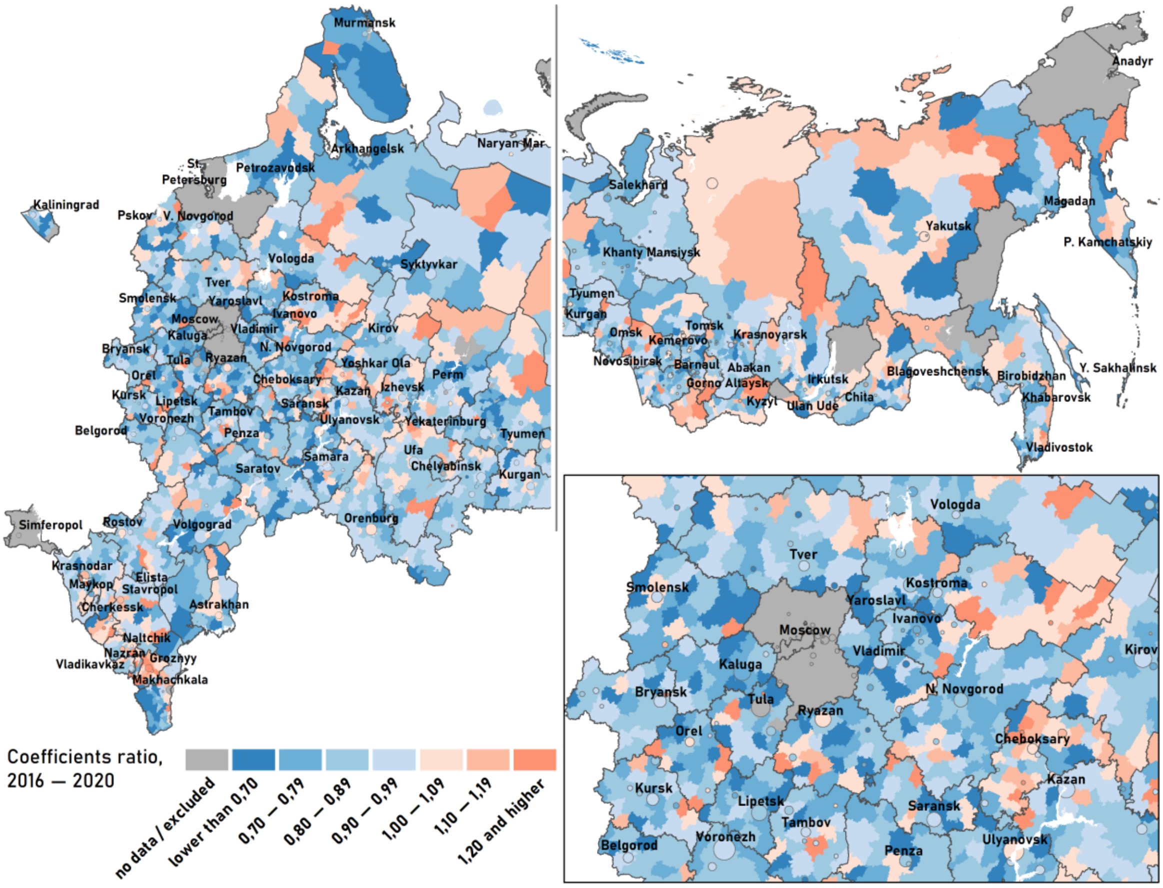
In most municipalities, men exhibit higher migration activity, both in large cities and peripheral territories. However, in some cases, the migration activity of men is distorted by the predominantly male contingents of the Federal Penitentiary Service, particularly in the north of the Perm Krai, northeast of the Kirov Oblast, and north of the Nizhny Novgorod Oblast.

Departures aged 40 and over. A higher intensity of departures is observed among men, but this phenomenon is less common across municipalities. Nevertheless, where women show higher migration activity, it tends to occur in sparsely populated municipalities. Large cities, on the other hand, consistently exhibit an imbalance in departures, favoring men.

Younger ages display a higher migration activity of women throughout the country, both in urban centers and peripheral regions. Significant sex disparities in favor of women are noted in macro-regions with a substantial institutional population, such as the North and the Far East, as well as regions where military service is more popular among men.

At older ages, the spatial picture is less clear due to the influence of small numbers. However, densely populated municipalities, especially large cities, tend to have a higher migration activity of men.

Figure 3. Coefficient ratios of arrivals aged 40 and over per 1,000 residents of the corresponding sex(female/male)

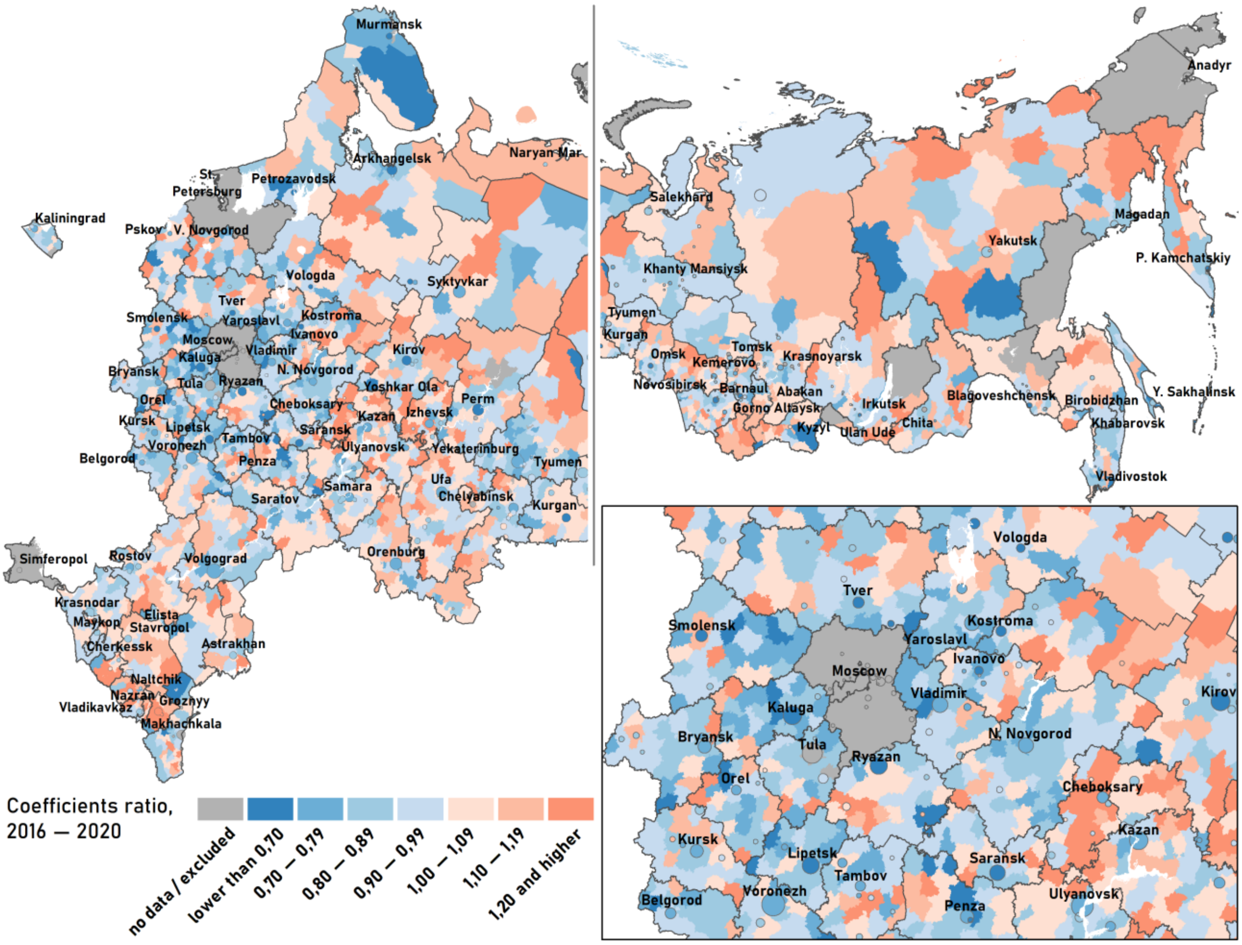


Source: based on the author's calculations

Intraregional migration in Central Russia. Detailed calculations for Central Russia confirmed the results of the cartographic analysis. Up to the age of 40, women display higher migration activity across all territorial types, while over 40, men become more active migrants (Figures 5-8). In this case, we managed to get rid of the influence of small numbers. Large cities receiving migrants show a peak disproportion of 1.5 times in favor of women at the age of 20–29 years, while men aged 40 and above tend to have 10–30% higher migration activity compared to women. Similar disproportion profiles are observed in medium and small towns with peripheral areas, with the peak disproportion by 2 times in favor of women occurring at the age of 20–29 years, and differences at ages 40+ being similar to those in large cities.

Graphs for all types of territories indicate that women are more active in migration flows directed toward large cities (arrivals in regional centers and departures from the periphery) compared to "pseudo-return" flows (departure from large cities and arrival in the periphery). Events such as childbirth, admission of children in kindergartens or schools, and marriage may lead women to change their registration to their actual residence, contributing to their higher migration activity, as pointed out by N. Mkrtchyan (2021). For large cities that are not regional centers, arrivals and departures show similar values, likely due to their smaller role as recipients of migrants at the regional level compared to regional centers.

Figure 4. Coefficient ratios of departures aged 40 and over per 1,000 residents of the corresponding sex(female/male)



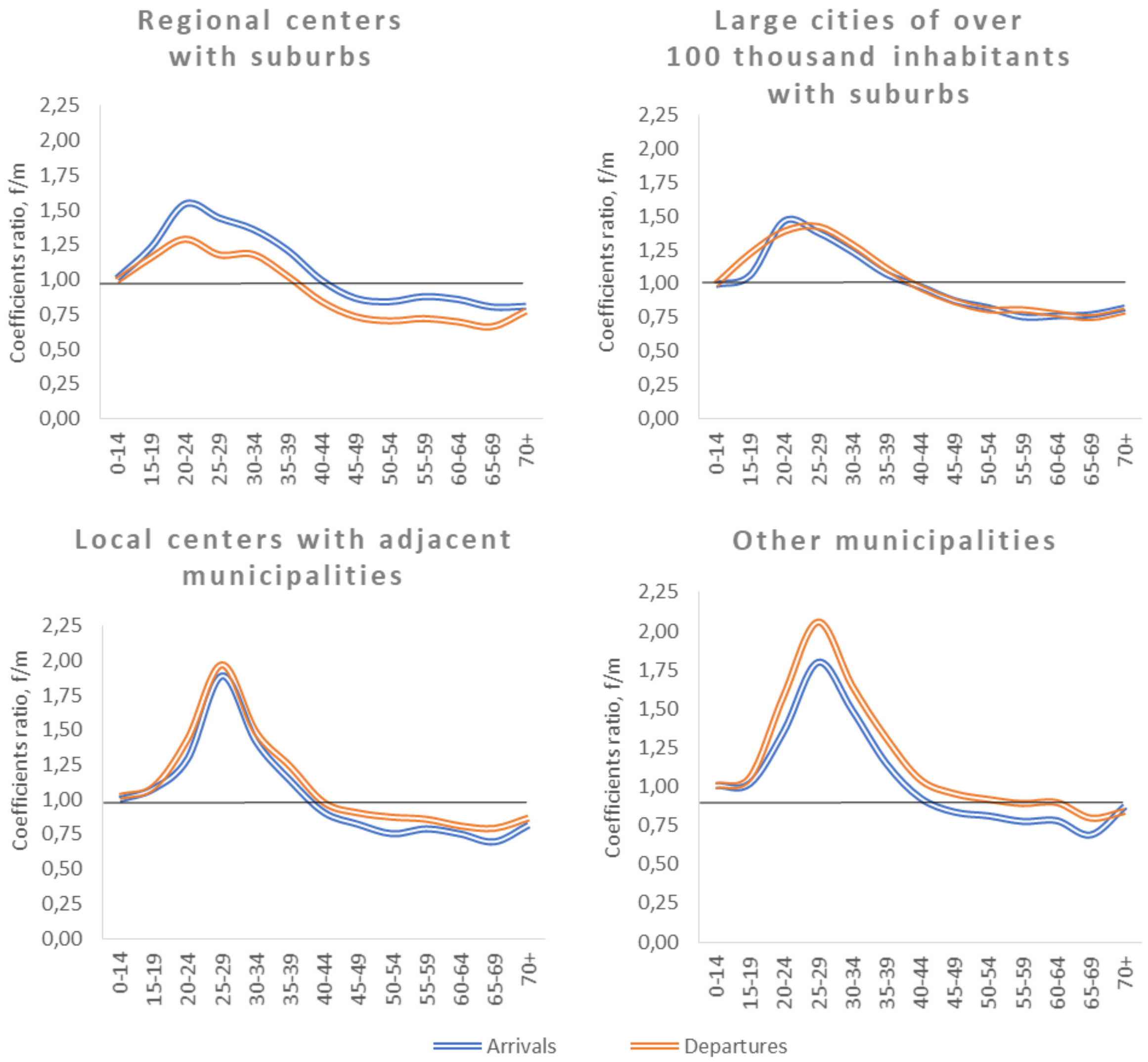
Source: based on the author's calculations

In large cities, in contrast to local centers and peripheries, the largest gender disparities in favor of women occur in the age group of 20–29 years, while disparities tend to be smaller due to larger population cohorts. At ages 40 and over, differences in migration intensity between men and women are similar for all types of municipalities, with a disproportion in favor of men.

The age profiles of local centers and peripheral municipalities are similar, primarily because small towns have lost their attractiveness to young residents and those from surrounding territories (Mkrtchyan 2017). Consequently, the characteristics of migration flows in local centers have become similar to "peripheral" ones. The sharp peak of the disproportion in favor of women in these areas occurs in the 25-29 years age group, which corresponds to the period of high birth (Rosstat 2022b) and marriage rates[5], likely contributing to the greater migration activity of women.

[5] The second highest age-specific marriage rate after the age group of 20-24 years. Own elaboration for five-year age groups according to data from the Demographic Yearbook of Rosstat (Rosstat 2022b).

Figure 5-8. Coefficient ratios of arrivals and departures (female/male) in municipalities of Central Russia of different types



Source: based on the author's calculations

Conclusion

At young ages, intraregional migration is more intense among women, with variations depending on the type of area. Factors such as higher school performance leading to better university opportunities in regional centers and limited employment prospects in peripheral territories contribute to this trend. The peak disproportion in favor of women at 25–29 years is likely influenced by changes in registration due to marriage, childbirth, or enrolling children in preschool or school.

However, the migration activity of men is potentially underestimated due to the way military personnel and contingents of the Federal Penitentiary Service in demographic statistics are recorded. Service in the armed forces, involving men in interregional migration, is more widespread in some regions, which entails an artificial underestimation of the intensity of intraregional moves of men. Moreover, the presence of institutional populations artificially increases the "denominator" and, thus, underestimates the migration activity of "local" men.

Men over the age of 40 exhibit higher migration activity in all types of territories, possibly due to fewer responsibilities for childcare after divorce and elderly care, making them more mobile compared to women.

As a result of intraregional migration at young ages, an imbalance in the sex ratio arises in the population of large cities, with a higher proportion of women, and in peripheral territories, with a higher proportion of men. Despite the higher intensity of moves of men relative to women in older age groups, it has a much smaller impact on the age-sex structure of the population due to the low overall level of migration mobility at these ages.

References

Alekseev A.I., Vorobiev M.I. (2018). How many rural residents are there in Russia in winter? *Vestnik Moskovskogo universiteta. Seriya 5, Geografiya*, (5), 104-107. (In Russian).

Bessudnov A. R., Malik V. M. (2016). Socio-economic and gender Inequalities in educational trajectories upon completion of lower secondary education in Russia. *Educational Studies. Moscow,* (1), 135-167. https://doi.org/10.17323/1814-9545-2016-1-135-167 (In Russian).

Clark W.A.V. (2013). Life course events and residential change: Unpacking age effects on the probability of moving. *Journal of Population Research*, 30(4), 319-334. https://doi.org/10.1007/s12546-013-9116-y

Faggian A., McCann P., Sheppard S. (2007). Some evidence that women are more mobile than men: Gender differences in UK graduate migration behavior. *Journal of Regional Science*, 47(3), 517-539.

Fomkina A. A. (2017). Discrepancies in Estimates of the Rural Population Size (Case Study of the Tver Oblast). *Vestnik Moskovskogo universiteta. Seriya 5, Geografiya*, (6), 88-91. (In Russian).

Gerasimov A. A. (2022). Sex ratio by age in Russia: spatial aspects. *Vestnik Moskovskogo universiteta. Seriya 5, Geografiya*, (4), in print. (In Russian)

Johansson, M. (2016). Young women and rural exodus–Swedish experiences. *Journal of Rural Studies*, 43, 291-300.

Karachurina L. B., Mkrtchyan N. V. (2013). Change of Population Numbers in Administrative Units and Cities of Russia (1989–2010): Centre-Periphery Relationships. *Voprosy Geografii* [Geography Studies], 135, 82-107. (In Russian).

Karachurina L. B., Mkrtchyan N. V. (2016). The Role of Migration in Enhancing Contrasts of Settlement Patterns at the Municipal Level in Russia. *Izvestiya Rossiiskoi akademii nauk. Seriya geograficheskaya*, (5), 46-59. http://dx.doi.org/10.15356/0373-2444-2016-5-46-59 (In Russian).

Karachurina L.B., Mkrtchyan N.V., Petrosian A.N. (2021). Spatial Patterns of Net Migration in the Suburbs of Russian Regional Centers. *Vestnik Moskovskogo universiteta. Seriya 5, Geografiya*, (6), 123-134. (In Russian).

Karpestam, P., Håkansson, P. G. (2021). Rural boys, urban girls? The mystery of the diminishing urban-rural gender gap in Sweden. *Journal of Rural Studies*, 86, 282-297.

Kashnitskii I. (2014) Chto v vozraste tebe moem? *Demoscope Weekly*, 581-582, 1-21. http://www.demoscope.ru/weekly/2014/0581/demoscope581.pdf (In Russian).

Kazenin K. I. (2014). Migratsiya dagestantsev vnutri regiona: nekotorye demograficheskie zagadki. Etnokul'turnye landshafty na postsovetskom prostranstve: problemy i osobennosti formirovaniya dagestanskogo komponenta (k 90-letiyu IIAE DNTs RAN). Makhachkala: IIAE DNTs RAN, 2014, 30-40. (In Russian).

Leibert, T. (2016). She leaves, he stays? Sex-selective migration in rural East Germany. *Journal of Rural Studies*, 43, 267-279. http://dx.doi.org/10.1016/j.jrurstud.2015.06.004

Leonidova G. V., Vyacheslavov V. N. (2016). Gender Specifics of Intraregional Migration. *Territorial Development Issues*, 2 (32), 1-12. (In Russian).

Mkrtchyan N. V. (2017). Youth migration from small towns in Russia. Monitoring of Public Opinion: Economic and Social Changes, 1, 225—242. https://doi.org/10.14515/monitoring.2017.1.15 (in Russian).

Mkrtchyan N. V. (2019). Migration in the North Caucasus and the Accuracy of Statistics. *The Journal of Social Policy Studies*, *17*(1), 7-22. https://doi.org/10.17323/727-0634-2019-17-1-7-22 (In Russian).

Mkrtchyan N. V. (2020). Problems in the Statistics of Internal Russian Migration Caused by Changes in Accounting Methods in 2011. *Demographic Review*, *7*(1), 83-99. https://doi.org/10.17323/demreview.v7i1.10821 (In Russian).

Mkrtchyan N. V. (2021). Sex Imbalances in Long-Term Migration Flows in Russia. *Demographic Review*, 8(3), 6-19. https://doi.org/10.17323/demreview.v8i3.13264 (In R Russian).

Mkrtchyan N. V., Florinskaya Y. F. (2019). Residents of Small and Mid-Size Towns of Russia: Labor Migration as an Alternative to Permanent Transfer. *Journal of the New Economic Association*, 3(43), 78-94. (In Russian).

Nefedova T. G. (2009). Polarization of Russian Space: Areas of Growth and "Black Holes". *Economics of Contemporary Russia*, 44(1), 62-77. (In Russian).

Nefedova T. G., Treivish A. I. (2002). The Theory of "Differential Urbanisation" and Hierarchy of Cities in Russia. *Problemy urbanizatsii na rubezhe vekov* [Issues of Urbanisation at the turn of the Century], Smolensk: Oikumena, 71-86. (In Russian).

Pyankova A. I. (2014). Problems of Comparability of Census Data for 2002 and 2010 (the Case of Moscow Region)], *Regional'nye issledovaniya* [Regional study], 2014, (1), 109-121. (In Russian).

Ravenstein E.G. (1885). The laws of migration. *Journal of the statistical society of London*, *48*(2), 167-235.

Raysikh A. E. (2019). 434 ekonomicheskikh rayona, 29 millionnykh aglomeratsiy i drugoe. https://reissig.livejournal.com/11959.html (In Russian).

Rosstat (2021). Migratsionnyy prirost naseleniya po polu, vozrastu i potokam peredvizheniya. https://showdata.gks.ru/report/278004/ (In Russian).

Rosstat (2022a). Baza dannykh pokazateley munitsipal'nykh obrazovaniy. https://www.gks.ru/dbscripts/munst/ (In Russian).

Rosstat (2022b). Demographic Yearbook of Russia 2021. https://gks.ru/bgd/regl/B21_16/Main.htm (In Russian).

Zayonchkovskaya Zh. A. (1991). *Demograficheskaya situatsiya i rasselenie*. Moscow: Nauka. (In Russian).

Zubarevich N. (2013). Four Russias: Human Potential and Social Differentiation of Russian Regions and Cities. In: Lipman M., Petrov N. (eds) Russia 2025. Palgrave Macmillan, London. https://doi.org/10.1057/9781137336910_4

Демографическое обозрение / Demographic Review
2022;9(2)4-21 DOI: https://doi.org/10.17323/demreview.v9i2.16203

Demographic features of industrial serfs in the 18th-19th centuries: parish registers of the village of Kudinovo in the Bogorodsky district, 1777-1862

Vlas A. Ryazanov
(vlas.ryazanov@gmail.com), Institute of Economics of the Russian Academy of Sciences, Russia.

Abstract: *The author uses parish registers of Kudinovo in the Moscow region as a source of information about the demographics of serfs employed in industrial production of textiles and bricks. The neighboring parish of Biserovo, with state peasants and no industrial facilities, is used for comparison. Nearly all demographics of industrial serfs are higher than those of their free neighbors. Industrial serfs married earlier, wedded younger women, had more sons surviving to adulthood, and their population grew by a third faster despite lower adult life expectancy. Economic protection by the nobility, most notably their exemption from income tax, allowed many peasants to run their own business and then move into the merchant class.*

Keywords: *Russia, historical demography, marriages, births, deaths, parish registers, eighteenth century, nineteenth century, industrialization, peasants.*

Acknowledgments: *To my teacher Vladimir Nikolaevich Gorlov (1939-2016), an expert on the economics and geography of Russian industry, associate professor at Moscow State University, grandson of an Orekhovo-Zuevo weaver.*

For citation: *Ryazanov V. (2022). Demographic features of industrial serfs in the 18th-19th centuries: parish registers of the village of Kudinovo in the Bogorodsky district, 1777-1862. Demographic Review, 9(2), 4-21. https://doi.org/10.17323/demreview.v9i2.16203*

Introduction

Analyzing the demographic processes of the past is important for understanding the driving mechanisms of population dynamics, testing and correcting relevant theories and hypotheses, and shaping effective demographic policy that takes into account the historical and cultural characteristics of the population. By using long time series of demographic indicators, we can draw unbiased conclusions about the general socio-economic and cultural trajectory of individual communities, territories and countries.

Modern sources of demographic statistics are data from registry offices, population censuses and targeted research. Works on historical demography, as a rule, are based on individual lists of the population - a description of the sex, age and family structure at a certain date. The sources of continuous time series are parish registers and databases created on their basis, which can additionally rely on any other references to people and events in their lives. In foreign, primarily European, practice, such databases have long been extensively used in historical and demographic research.

In Russia, for a long time, only highly aggregated indicators obtained from parish registers, collected by statistical authorities since the middle of the 19th century, were available for scientific use. Work with primary sources of information was hampered by the restrictive nature of Soviet and later Russian archives, the unreadiness to use automatic data processing methods, and the almost complete lack of state funding for this important area of science (Vladimirov, Sarafanov, Shchetinina 2019). Among the few completed and published Russian studies based on a continuous array of parish registers it is worth noting works on the Vykhino estate of the Sheremetevs in the Moscow district in 1815-1917 (Avdeev, Troitskaia, Blum 2004) and on the Intercession parish of Barnaul in 1877-1886 (Vinnik 2012), as well as the author's work on three parishes in geographically different parts of European Russia in 1758-1862 (Ryazanov 2021). The latter revealed a large spatial difference of demographic indicators depending on local socio-cultural characteristics, with common features of long-term dynamics from the mid-18th to the mid-19th centuries: the postponement of marriage and the decrease in fertility amid an increase in mortality of both the child and adult population.

In a detailed review of long-term demographic studies of the Skåne region in southern Sweden, Lund University researchers T. Bengtsson and M. Dribe (Bengtsson, Dribe 2021: 80) came to similar conclusions. The population of this territory in the 18th and 19th centuries experienced a mass postponement of marriages, the impossibility for many of creating a family, and a decrease in fertility with high mortality. An article about the sad pages of the demographic history of their country, which don't really fit the classical theory of demographic transition, which assumes a decrease in fertility as a reaction to a decrease in child mortality (Vishnevsky 2006: 9), was called by the authors "The Long Road to Health and Prosperity". The reason for the crisis was the "downward" social mobility of the bulk of the rural population as the grain export economy progressed.

The demographic problems of traditional rural communities as a result of the development of a market economy were also noted in Russia, where works in this area of study mainly used the figures of the second half of the 19th century (Vishnevsky 2006: 18). The keen interest in this period in Soviet and post-Soviet historiography and demography is related to the fact that it preceded the revolution of 1917. Attention to the demography of an earlier era, before the abolition of serfdom, was associated with attempts to estimate the population dynamics of the serfs. The large body of works in this area relied only on the data of revision lists and led to a consensus on the demographic decline of the serf population before 1861 due to "landlord exploitation" (Perkovsky 1977). This conclusion correlated well with the official political guidelines of the pre-revolutionary, Soviet and early post-Soviet years (personal freedom as an undoubted good), but ignored, for example, the ultra-high (2-3% per year) natural increase of slaves in the United States in the first half of the 19th century (U.S. Census ... 1870: 7).

One of the key components of the classical theory of demographic transition is urbanization, which in the 18th and 19th centuries was usually associated with the development of industry and trade. In Russia at that time, the share of the urban population, according to revision data, was small, 7-8% in 1811 and 1856 (Rashin 1956). As we will see below, even these small figures are most likely overstated. Industrial activity during that period was concentrated mainly in rural areas (Stolbov 2013). So, this work is devoted to the study of the demographic indicators of the rural industrial serf population.

Information sources

The basis of the present research is parish registers of two neighboring churches of the Bogorodsky district of the Moscow province in 1777-1862: one of Biserovo and one of Kudinovo. The focus of this work is the parish in Kudinovo, while the adjacent parish of Biserovo, discussed in detail in an earlier article by the author (Ryazanov 2021), is used for comparison. The fundamental difference between the populations of the two adjacent parishes was the class of the inhabitants (state peasants in Biserovo, serfs in Kudinovo), as well as the presence of industrial enterprises in the Kudinovo parish that were significant for that time, and not just home handicraft production (table).

Industry in the infertile eastern part of the Moscow outskirts emerged as an additional source of income for the local population, which became especially important starting in the mid-1760s[1]. Thanks to easily accessible clay layers, brick and then faience production was developed in Kudinovo (Lyubavin 2010). In 1770, in the village of Kamenka of the same parish, landowner M.G. Okulov built a textile factory (Lyubavin 2004). The progress of local production was facilitated by the renaissance of construction and trade in Moscow with an increased attention to the city after the plague and riots of 1770-1772[2]. The industrial development of Kudinovo continued well beyond the period under review. In 1908 the first carbon black plant in the country was built nearby, and in 1956 most of the former villages of the parish became the city of Elektrougli. Now the entire area of interest is a part of the Bogorodsky urban district (formerly the Noginsk district).

[1] In 1766, the government of Catherine II reduced export duties on grain, which led to a doubling of bread prices in 10 years (Mironov 1985:48). In regions with low yields, this caused a sharp drop in living standards, leading, already in 1767, to the legalization of cottage weaving, with which peasants could support themselves.

[2] Major construction projects of that time were the Tsaritsyno estate, the Mytishchi water pipeline, the University on Mokhovaya, the Petrovsky Travel Palace, Demidov's house, Pashkov's house and the house of Yushkov, the owner of two Kudinovo villages. At the same time, the Great Broadcloth factory near the Stone Bridge, from which the plague epidemic began, fell into decay, giving the market to private manufactories.

Table. Information about the parishes studied

	Biserovo	Kudinovo
Parish church	Church of the Epiphany	Church of the Intercession of the Mother of God
Years of skipped parish registers	1780, 1782, 1799, 1801, 1803, 1805, 1809, 1812, 1814, 1820, 1826, 1828, 1829, 1857-1859	1780, 1782, 1801, 1803, 1805
Birth records, total	3952	7859
Marriage records, total	786	1560
Death records, total	2878	5217
Parish composition	State peasants	Landlord peasants
Reference lists of the population	RL 1773, 1811-1858, HC 1869 + CR 1786 и 1799 For Biserovo, Novaya, Chernaya and Vishnyakovo	RL 1816 (all), 1834 (part), 1850 (all) + CR 1786, 1804, 1828, 1833, 1860 For Kudinovo, Belaya, Cherepkovo, Sokolovo (Bykovo), Isakovo, Safonovo, Vasilievo, Kamenka with offshoot settlements
Local geography of the parish	The village on Biserovo Lake, villages on Vladimir and Nosovikhinsky tracts	Villages around the Nosovikhinsky tract, the village to the north of it
Population growth rate per year in 1786-1860, %	0.88	1.17

Note: RL - revision list, HC - household census, CR - confession register.

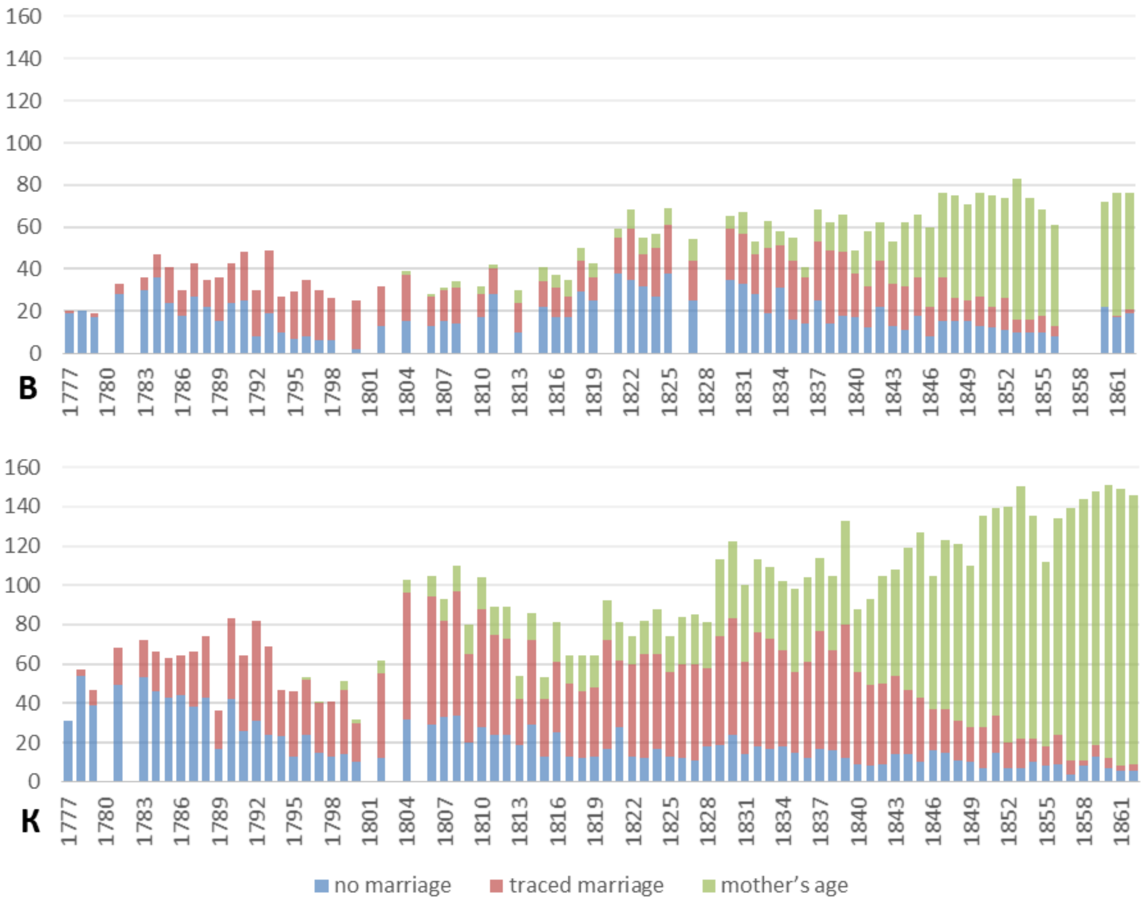
The parish of Biserovo included the former possessions of the Moscow Spaso-Andronikov Monastery, one of whose villages (Vishnyakovo) was surrounded by the twice as large parish of Kudinovo. Despite this, before the provincial reform of 1775, they belonged to different *stany* of the Moscow district (the parish of Kudinovo to Kamensky, and the entire estate of Biserovo to Pochernev). At the end of the 1810s, the northern part was separated from the Kudinovo parish (not considered in this study), and by the middle of the century new settlements were founded within it. The large number and frequent change of owners of villages in the parish of Kudinovo make it very difficult to form a reference set of lists of the population according to revision lists; for this, confessional registers were used along with revisions.

The methods of studying parish registers were described in the author's previous work (Ryazanov 2021) and remained unchanged for this study. Automatic data processing (the search for links between the three parts of parish registers using Excel formulas) with manual verification through population lists was applied[3]. This made it possible to correct errors and determine the exact age of newlyweds (Figure 1) and the names of mothers in years when these details were not yet mandatory for Bogorodsk parish registers (until 1839 and 1830, respectively), as well as to prove the reliability of marriage records as early as the end of the 18th century (according to the share of children whose parents' marriage was traced in the registers).

[3] So, to determine the exact age of newlyweds, the formulas look for the number of newborns in part 1 of the register of the appropriate age, place of birth, first name and patronymic, for which there were no records of death in childhood in part 3 and there is no marriage mark yet. The author manually checked the records found for compliance with information about the parents of the newlyweds, if any, with simultaneous verification through revisions to exclude persons of the same name, and only in the case of an unambiguous correspondence made a marriage mark.

All source materials (databases of parish registers and lists of the population) are posted by the author in the public domain on the All-Russian Genealogical Tree portal[4].

Figure 1. The number of newborns by years and information about the marriage of their parents (including those where the age of the bride is known) in Biserovo (B) and Kudinovo (K), persons



Source: Author's calculations.

Marriages

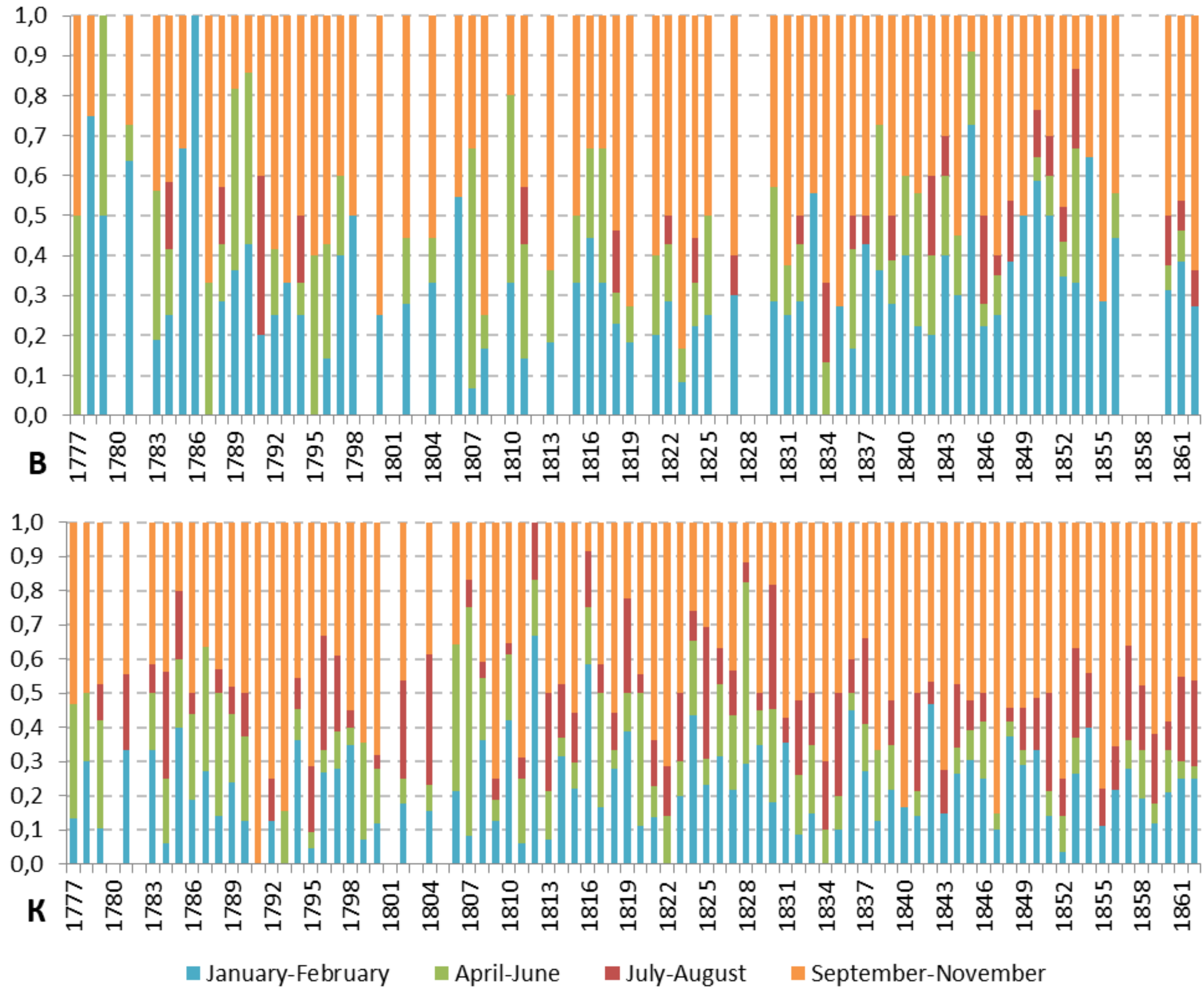
An analysis of the seasonality of weddings (Figure 2) already shows clear differences between the two parishes. Each of the marriage seasons between multi-day Orthodox fasts[5] had advantages and disadvantages for the rural population, which were determined by the proximity of the fasts and the strictness of their observance, the cycle of labor-consuming agricultural work and the related amount of food stocks in households. In Kudinovo, autumn and even summer marriages were popular, while in Biserovo, over time, weddings increasingly took place at the beginning of the year. The seasonality of the brick industry was the opposite to that of agriculture – the inhabitants of Kudinovo dug clay in the winter, while bricks were produced in the summer (Lyubavin 2010). This explains the low popularity of winter marriages and the relatively large

[4] https://vgd.ru/

[5] According to the rules of the Church, weddings are performed in four periods: in the winter from Epiphany to the beginning of Shrovetide, in the spring from the end of Holy Week to the beginning of St Peter's Fast, in the summer from the day of Peter and Paul to the beginning of the Dormition Fast, in the fall from the Dormition of the Mother of God to the beginning of the Nativity Fast.

number of weddings in summer, when the agrarian population was reluctant to marry. Winter marriages added working women to the families before the agricultural season, but for the industrial population this was not so important, so weddings in Kudinovo tended towards the well-stocked autumn off-season.

Figure 2. Seasonal distribution of weddings in Biserovo (B) and Kudinovo (K) by years, shares of 1

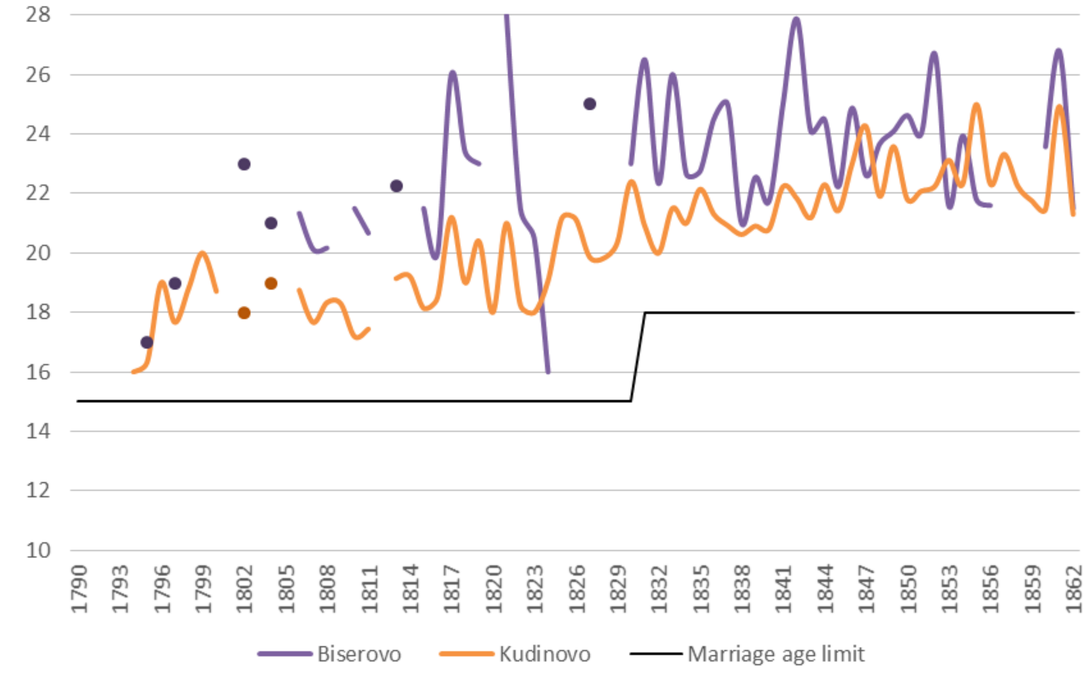


Source: Author's calculations.

The age of men at first marriage (Figure 3) in Kudinovo was 2-3 years lower than in Biserovo for almost the entire period reviewed, excluding the end of the 18th century. However, even at that time, the average age of the first wedding of young men in both parishes was noticeably higher than the minimum allowed (15 years). By the middle of the 19th century, first-time grooms in Kudinovo had become 4.5 years older, while in Biserovo they were 6.5 years older, and in some years the average age of the grooms there was close to 30.

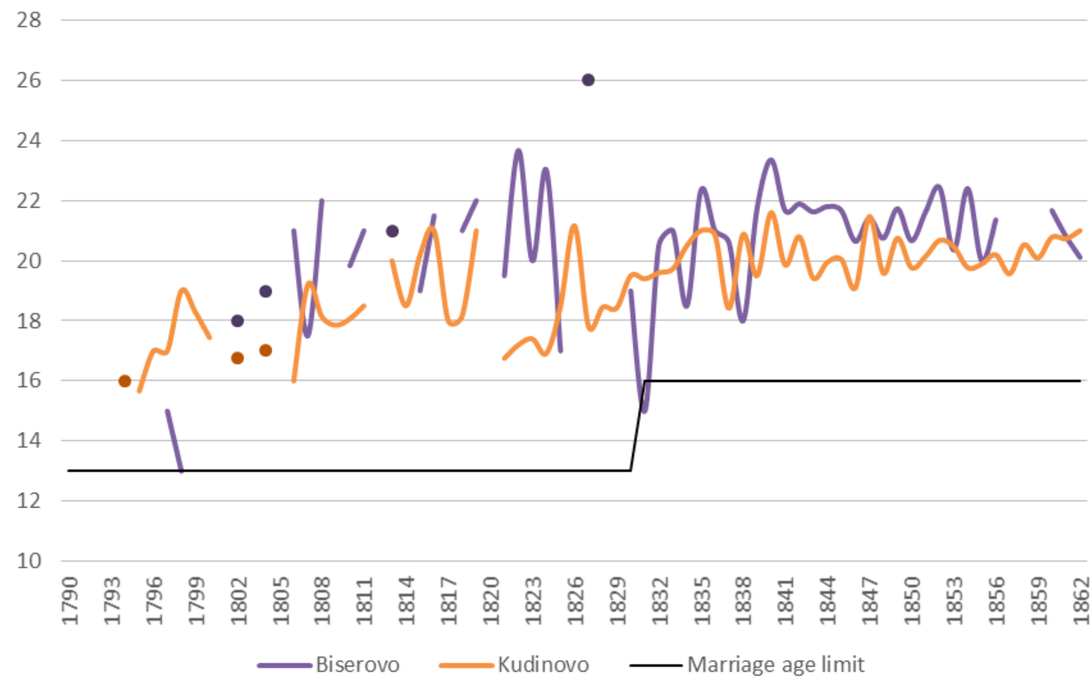
The increasing postponement of first marriages occurred in Kudinovo only in the 1820s-1830s, while in Biserovo it began much earlier and lasted several decades. Kudinovo's indicators are in good relation with figures of the Vykhino estate (Avdeev, Troitskaia, Blum 2004).

Figure 3. Average age at first marriage for men and minimum age of marriage by years, years



Source: Author's calculations.

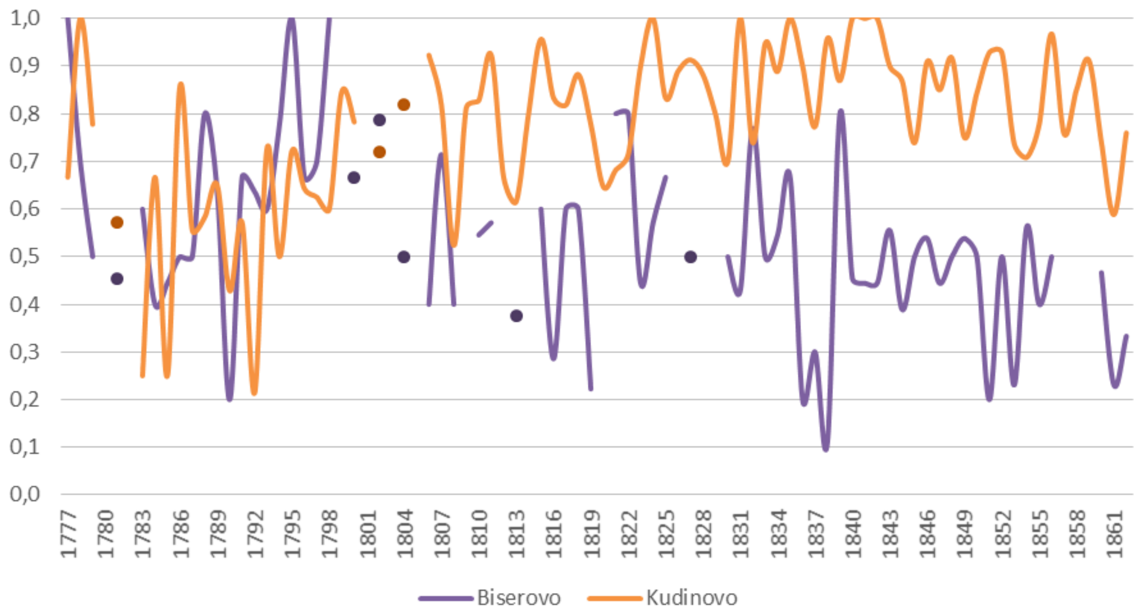
Figure 4. Average age at first marriage for women and minimum age of marriage by years, years



Source: Author's calculations.

In Kudinovo, the age of brides (Figure 4) at their first marriage increased by about 3 years in the first quarter of the 19th century and starting in the 1830s was about 20 years, while in Biserovo women married an average of 2 years later. In both parishes, marriage among the female population was not universal, but while in Biserovo this trend, associated with the postponement of marriage by men and their higher mortality, manifested itself already at the end of the 18th century, in Kudinovo it became the reality only by the middle of the 19th century. The age of marriage in both parishes was much higher than the minimum allowed; breaches of this, when the bride or groom was underage, were extremely rare. The situation when the groom was younger than the bride, which is common in multigenerational families, was quite common in both parishes. In Kudinovo in the first third of the 19th century, 39% of fiancés were younger than their future wives; in subsequent decades this proportion dropped to 21%. The picture in Biserovo was similar.

Figure 5. Share of brides from the same parish by years, shares of 1



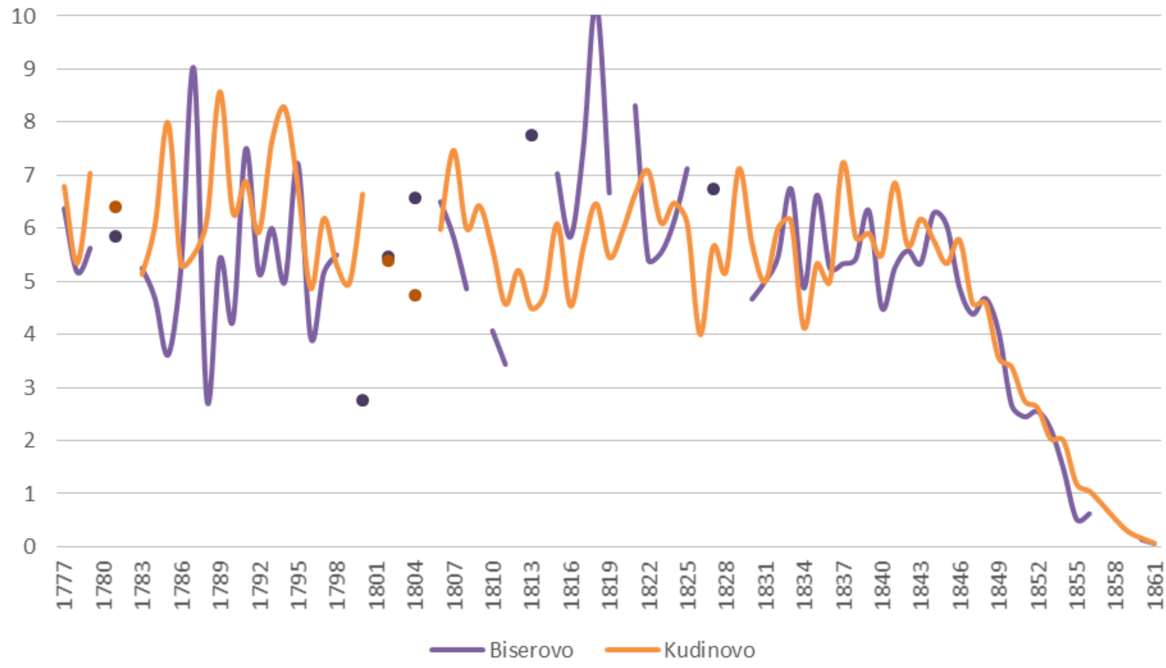
Source: Author's calculations.

In Kudinovo, in contrast to Biserovo, there was a prominent local-territorial endogamy, and while the share of local brides in the first parish grew, in the second it steadily decreased (Figure 5). The extremely small exchange of brides between the two neighboring parishes is noteworthy: only 18 from the Biserovo parish to Kudinovo and 21 in the opposite direction. Moreover, most of these wedding trains from Biserovo set off in the 1840-1850s, while the grooms of the Biserovo estate took brides from Kudinovo mainly before the 1820s.

Births

Birth registration in the parish of Kudinovo, as well as in Biserovo, was reliable already in the first surviving parish registers, if we evaluate it by the ratio of girls and boys, which is close to 0.95 for natural fertility. However, later in Kudinovo, the quality of record-keeping dropped dramatically, and already in the 1790s this figure fell to 0.6, i.e., a third of the births of girls were not recorded. In the 1800s-1810s, this indicator averaged 0.8. Unsurprisingly for an industrial rural population, the appearance of additional females was less important than of men. Brickmaking required hard physical labor, and at Okulov's Kamenka manufactory the confession register of 1786 shows a lot of "single" boys aged 13–19, but not girls. To correct birth rates for "problem" years, we should multiply the known figures by (1 + 0.95) / (1 + k), where k is the observed sex ratio, and to adjust for the years of absence of part of the parish registers, we should additionally multiply by 1/(1-x), where x is the share of missing registers in female reproductive years.

Figure 6. Average number of known children of spouses by years of wedding (if there were children) adjusted for the absence of part of the entries in the registers and of the register books themselves, pers.



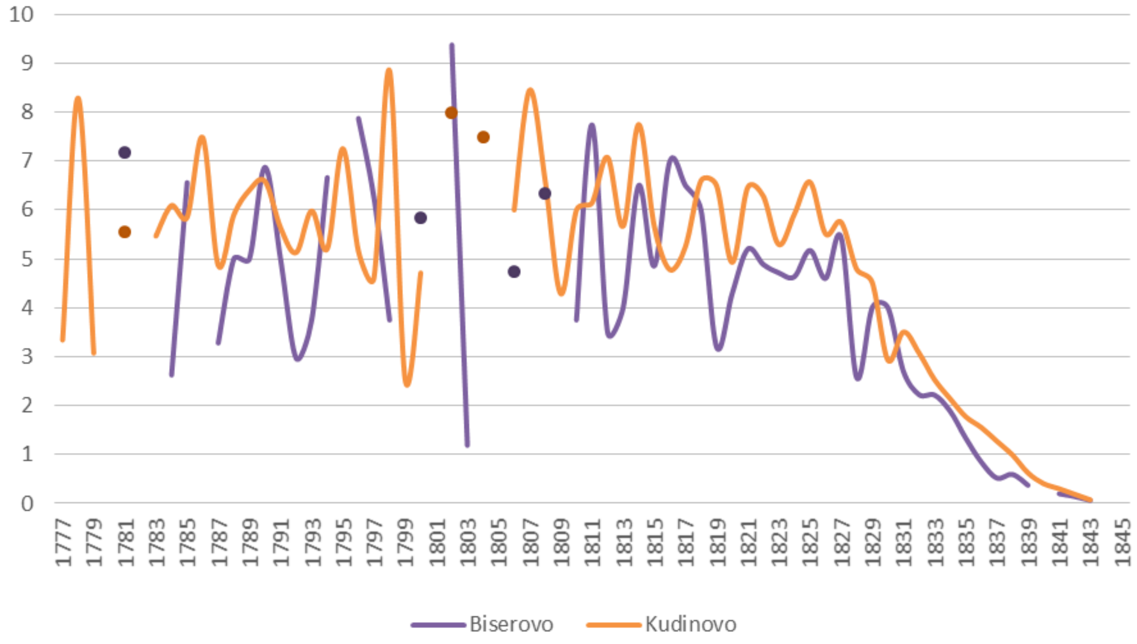
Source: Author's calculations.

The average number of children born in marriages (Figure 6) in both parishes during the reviewed period fluctuated around 6, and in Kudinovo it was usually higher, with the exception of the mid-1800s - mid-1830s. It is noteworthy that after the growth in the 1810s, the birth rate in Kudinovo was stable, while in Biserovo it began to decline already in 1830-1840, and this is not due to the limits of time series[6]. The calculation of the average number of children by year of birth of mothers (with similar adjustments) gives slightly different results (Figure 7). In Kudinovo, this birth rate is consistently higher than in Biserovo, and for the generations of the 1810s, the difference exceeds 1 on average. It should be noted, however, that the sizes of the samples, already different due to the twice as large parish of Kudinovo, are very different due to the spread of exogamy in Biserovo, hence the difficulty of determining the exact age of the brides from outsider parishes (figure 1).

Despite the fact that the brides in Biserovo were on average 2 years older, the age structure of mothers of newborns in the two parishes is almost identical - in 1830-1850 about 70% of children with a known mother's age were born to women 20-30 years old. This means a higher intensity of births at these ages in Biserovo. The parish of Kudinovo leads in two non-demographic indicators: the highest verified number of children for a married couple (15 versus 14) and the highest verified age of a mother at childbirth (47 versus 44).

[6] With an average age of first-married women of 20-22 years and an active reproductive age of up to 45, the time series limit of 1862 for parish registers affects the traced number of children born in marriages of the early 1840s and later.

Figure 7. The average number of children by years of birth of mothers (if there were children), adjusted for the absence of part of the entries in the registers and of the register books themselves, pers.



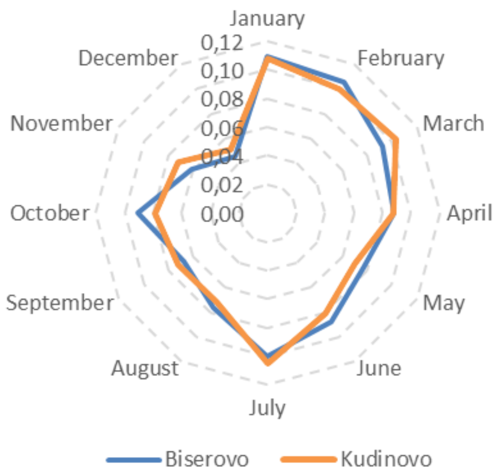
Source: Author's calculations.

The seasonality of fertility in both parishes is similar (Figure 8) and is typical for an agricultural population.

The birth of children in January-February had obvious advantages (the final months of pregnancy and the first months of infancy did not fall on the summer peak of agricultural work, which was also the peak of enteric infections). The seasonality of births which would be determined by the strict observance of Orthodox fasts (Avdeev, Blum, Troitskaya 2002; Mironov 2005; Vinnik 2012) is poorly represented in both parishes. From as early as the 1790s, December, May and September births (meaning probable conceptions during the fast of Great Lent with two adjacent weeks, and the Dormition and Nativity fasts) represented approximately 20% of all births. Abstinence was observed a little more strictly in the 1800s and 1840s.

Indicators of out-of-wedlock births differed significantly in the two parishes. In Kudinovo, this phenomenon became a stable part of social reality as early as the 1790s, while in Biserovo it did so only in the 1840s. On average, in the 1820s almost 5% of children were born out of wedlock in Kudinovo; in Biserovo, it was half that at its peak. In the following decades, the proportion of out-of-wedlock births dropped to 1-2% in both parishes. It is noteworthy that in Kudinovo, the percentage of out-of-wedlock births aggregated over decades (see Appendix) noticeably correlates with the share of probable conceptions of children during Orthodox fasts (coefficient 0.62).

Figure 8. Birth seasonality index, 1830-1850



Source: Author's calculations.

Deaths

The very first surviving registers of Kudinovo already contain records of infant and child mortality (in Biserovo they appear only in 1783). However, with a decline in the quality of recordkeeping in this parish, such information becomes rare in the 1790s-1810s. If we evaluate child mortality by the ratio of deaths under 10 years of age[7] to the number of births, then in both parishes it becomes plausible (about 40%) in the 1830s and grows steadily starting in the mid-1840s.

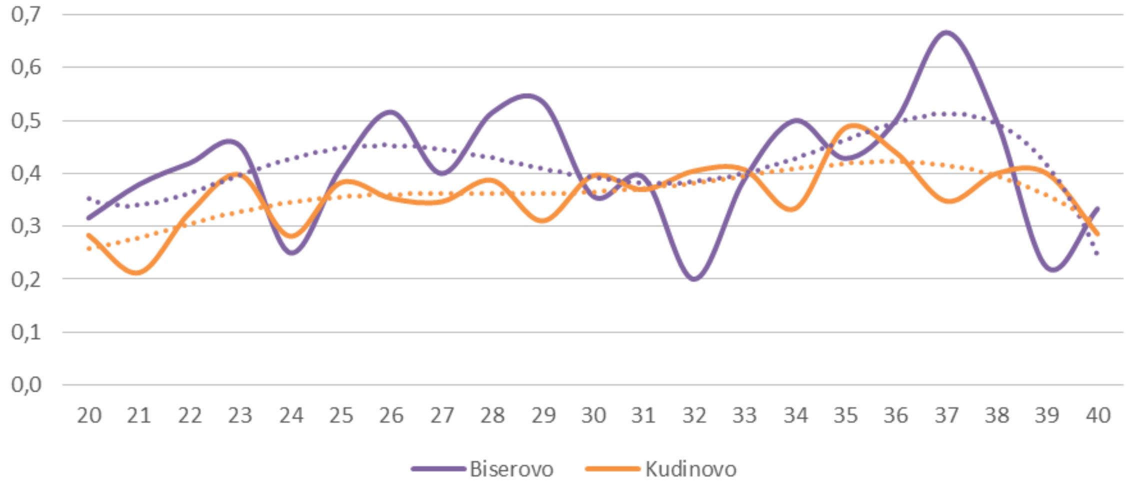
A comparison of the two parishes in terms of the proportion of those who died before the age of 10 depending on the age of the mother (Figure 9) shows a lower infant mortality rate in Kudinovo in most maternal age groups. However, with the increase of age of the mother from 20-25 to 35-40, mortality increases by an average of 10 p.p. in both parishes.

A relatively accurate assessment of demographic well-being in the case of obviously incomplete records of child mortality, is the opposite indicator, survival to a marriage (which means at least reaching adulthood). This can be calculated by automatically processing and linking the data of three parts of parish registers. Due to Biserovo's trend towards exogamy and a significant proportion of outsider brides in the parish of Kudinovo, it is possible to estimate in this way only the survival of boys.

The results of the calculation (Figure 10), above all, clearly demonstrate the limitations of the method for Biserovo, where part of the parish registers is not available. However, for generations whose probable marriages occurred during a period with a continuous array of register books, a comparison of the two parishes seems appropriate, and it is almost always in favor of Kudinovo.

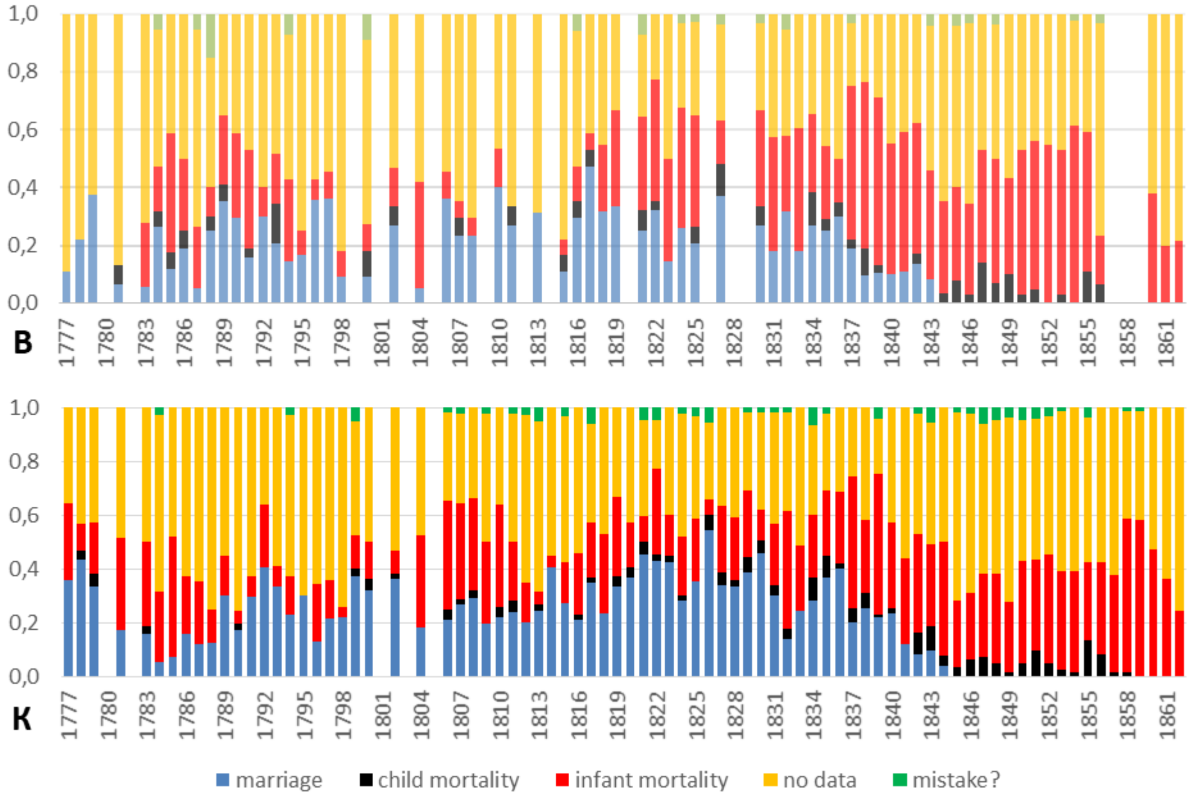
[7] For that era, the periods of childhood with an increased risk of death were significantly longer than after the "epidemiological revolution". Currently, child mortality is usually calculated up to 5 years, and infant mortality up to 1 year. This study used the ages of 10 and 3, respectively.

Figure 9. Proportion of children dying before age 10, by age of mothers, 1830-1850, shares of 1



Source: Author's calculations.

Figure 10. Demographic events of the boys of Biserovo (B) and Kudinovo (K) according to parish registers by year of birth, shares of 1

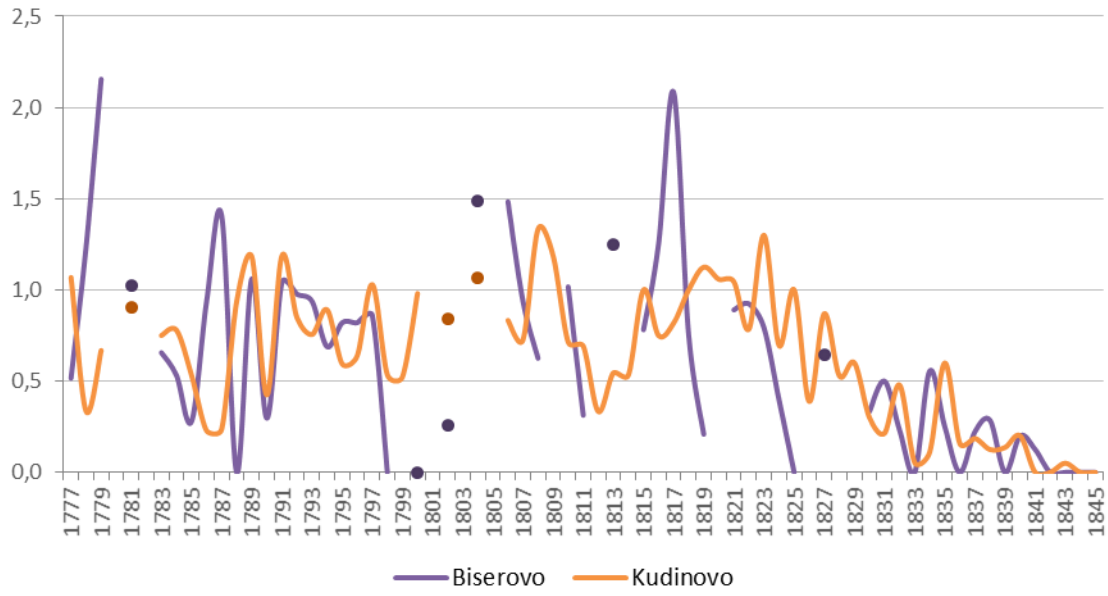


Source: Author's calculations.

Note: The algorithm searches for records by calendar years and does not allow for absolutely accurate age calculation; "mistake?" – age of death mismatch while other search criteria match

Calculation of the average number of sons who survived to their marriage based on data from parish registers (a substitute indicator for the net reproduction rate of the male population) can be relatively correct only for Kudinovo, while in Biserovo, due to the lack of part of the registers, the sample size is greatly narrowed[8]. Moreover, it is obvious that infant mortality is much more variable over the years than fertility. In the 19th century, in Kudinovo this indicator was temporary halved in the 1810s, as a result of a drop in fertility and a surge in mortality due to the war of 1812 (Figure 11)[9] .

Figure 11. The average number of sons who survived to their marriage born in marriages where there were children, by the years of their parents' wedding, adjusted for the absence of part of the parish registers, pers.



Source: Author's calculations.

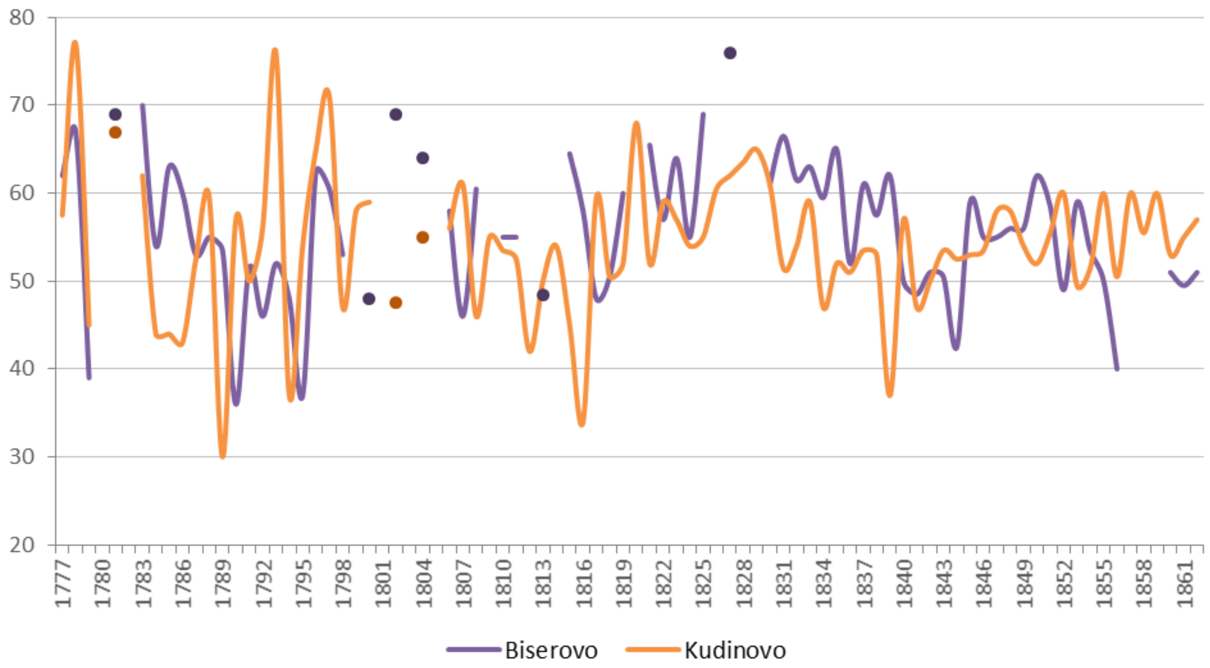
The adult mortality rate in Kudinovo was higher than in Biserovo for most of the reviewed period (see Figure 12 and Appendix).

With the onset of cholera epidemics in the 1830s, Kudinovo's indicators deteriorated faster and more dramatically. However, unlike Biserovo's, the decline in life expectancy here was not a progressive one. One of the components of Biserovo's demographic troubles was high mortality from external causes (in the Bogorodsk registers, causes of death were indicated starting in 1823, with external causes being indicated as an exceptional phenomenon even earlier), which is responsible for 2-3% of all entries in part three (deaths) of the parish registers of the middle of the century. In Kudinovo, this figure did not exceed 1%.

[8] It is necessary to have three correct entries in the registers: the marriage of parents, the birth of their son and his own marriage.

[9] The registers reflect the influx of refugees (and, obviously, infectious diseases) into the Bogorodsk district, which was then also an area of fighting. Local militia took part in the war and among them, according to revisions, casualties were heavy. So, out of 32 militants of the Kudinovo villages of Safonovo and Isakovo of P.I. Yushkov, 12 people “did not return”, and for M.G. Okulov’s Kamenka, 14 out of 25.

Figure 12. Median age at death of those aged 15 and over, years



Source: Author's calculations.

Results

The metrics of demographic processes described above allow us to conclude that the indicators of the industrial peasants of the Kudinovo parish differed significantly from the neighboring parish. Thus, in Kudinovo, marriages were frequent in autumn and even in summer, while in Biserovo they began to lean towards the winter months. The postponement of marriage with a stable increase of age at first marriage in Kudinovo occurred only in the 1820-1830s, while in Biserovo this process began half a century earlier. At the same time, both grooms and brides in Kudinovo were 2-3 years younger. Over time, the industrial parish of Kudinovo became more and more endogamous, while in the neighboring parish there was an increasing trend towards outsider brides.

The number of children born in families in Kudinovo was usually higher, with the exception of the mid-1800s to the mid-1830s. Kudinovo did not experience the decline in fertility which occurred in Biserovo in 1830-1840. Kudinovo, with earlier marriage, is characterized by a lower intensity of births among women at the beginning of their family life, while Biserovo's older brides after weddings gave birth more often, although as a result they had fewer children. Extramarital births in Kudinovo became common as early as the end of the 18th century, and their peak of 5% in the 1820s was twice as high as in the parish of Biserovo, where this phenomenon was extremely rare until the middle of the 19th century.

In industrial Kudinovo, the mortality of the child population was lower in most maternal age groups. Also, in Kudinovo the survival of the male population to marriage (so, at least to adulthood) was almost always greater. For those born in the early 1830s, the advantage of Kudinovo over Biserovo in this indicator was 10 p.p., as was its advantage in registered infant mortality in the 1850s (for Biserovo it was 10 percentage points higher). However, not childhood but adult mortality in industrial Kudinovo was almost always higher. Starting in the early 1830s, the median age of adult deaths here fell faster and more dramatically than in Biserovo, but unlike the neighboring parish, this decline did not become stable.

Some demographic characteristics of Kudinovo's industrial serfs have inherited features of the agricultural population. Above all, this concerns the tendency towards multigenerational families and the seasonality of births.

Integration of the indicators listed above leads to an unambiguous conclusion: the demographic situation in the industrial serf Kudinovo was better than in the state *votchina* of Biserovo. The men of Kudinovo delayed marriages less, took brides after the harvest with full granaries and not before sowing, and had younger wives, who could rely on the help of relatives from the parish and had more children at a more relaxed pace who more often survived to adulthood. Kudinovo, like Biserovo, experienced demographic crises, but got out of them faster and more surely.

The explanation for this is that the well-being of industrial serfs was obviously greater than that of their state-owned neighbors. Starting in the 1830s, many Kudinovites were given freedom and entered the ranks of the Moscow and Bogorodsk merchants and petty bourgeois, while continuing to live in their native area[10]. In the 1850s and 1860s, these classes already accounted for 7.5% of all birth records in the parish. Industry also passed into the hands of the former serfs from the heirs of noble entrepreneurs: merchants from Belaya, the Treshchalins, owned a brick production, and the Sorokins, merchants originally from Isakovo, bought out the Kamenka manufactory in 1835.

It is obvious that under "landowner exploitation" such a situation would not have been possible. The Kudinovo noble landowners were personally interested in the demographic well-being and income of their own serfs, from whom they collected rent. Peasants, meanwhile, received from the landowner tax immunity (no income tax was imposed on noble estates, except for 1812-1819), a community that collectively paid all obligations to the state (Avdeev, Troitskaya, Ulyanova 2015), and industrial assets with which they could develop their own business. The mechanisms of the formation of Russian industry through the entrepreneurship of serfs are well studied in the archives of the Sheremetevs' estates (Stolbov 2013)[11]. Free state peasants themselves bore the risks of non-agricultural trades, which, despite a nominally higher social status, led to worse economic and demographic characteristics. In our case, this is especially noticeable in the 19th century.

Conclusion

The integration of the Russian agrarian economy into the world market starting in the middle of the 18th century and the associated "price revolution", observed earlier in other countries, hit the "grain-consuming" territories near Moscow hard. The industrial activity of enterprising nobles allowed their serfs to exist in the new conditions much better than state peasants, which is demonstrated by almost all demographic indicators.

---

[10] Therefore, it is impossible to correctly estimate the number of urban residents according to the data of revision lists; they did not reflect the real place of residence of many merchants and petty bourgeois, all of whose demographic events took place in their native villages.

[11] It is noteworthy that Dmitry Spiridonovich (1778-1855), the founder of the Sorokin merchant dynasty, married one of his sons, Dmitry, to a native of the Sheremetev village of Vyazovka in the Vykhino *votchina*, and it was he who was made the main heir.

Of course, the features of the Kudinovo parish cannot be extrapolated to all the industrial villages of that era, not to mention to all the Russian serfs, many of whom lived in a completely different reality. For example, the indicators of the enclosed agricultural estate of Podol in the Kalyazinsky district from the author's above-mentioned article are strikingly better: the net reproduction rate of men there in the 1810s was 1.5, and the population between 1816 and 1834 grew by 1.7% per year. In many other agrarian estates, we can assume that landlords saw most of the serfs as merely surplus mouths reducing their marketable output.

The growing demographic problems of state peasants in the pre-reform period became a prototype of the massive changes that occurred after the abolition of serfdom and led to the marginalization of a significant part of the entire population of the country. The long-term experience of other states shows that it is possible to overcome the negative consequences of the formation of a market economy through a system of basic social and economic guarantees (Bengtsson, Dribe 2021: 87). In Russia in the 18th and 19th centuries, these were provided only by the guided community.

The demographic dimension of Russia's early industrialization, about which birth registers have much to tell, has been very little studied. The author hopes that this work will be another step towards understanding the relevant processes and serve as motivation for other researchers.

References

Avdeev A., Blum A., Troitskaia I. (2002). The seasonal factor of demographics of Russian peasants of the first half of the 19th century: marriages, births, infant mortality. *Russian Demographic Journal*, 1, 35. http://www.demoscope.ru/weekly/2003/0135/analit04.php

Avdeev A.A., Troitskaya I.A., Ulyanova G.N. (2015). Social differences in household structures in the 19th century: Moscow and its outskirts. *Demographic review, 2(2), 74-91.* https://doi.org/10.17323/demreview.v2i2.1783

Avdeev A., Troitskaia I., Blum A. (2004). Peasant marriage in nineteenth-century Russia. *Population*, 59(6), 721-764. https://doi.org/10.2307/3654894

Bengtsson T., Dribe M. (2021). The Long Road to Health and Prosperity, Southern Sweden, 1765-2015. Research Contributions From the Scanian Economic-Demographic Database (SEDD). *Historical Life Course Studies*, 11, 74–96. https://doi.org/10.51964/hlcs10941

Lubavin A.N. (2004). Kamenka, aka Lyubimovskoe. *Moscow Journal*, (6), 60-64. https://www.bogorodsk-noginsk.ru/atlas/4kamenka.html

Lubavin A.N. (2010). From the history of the Kudinovo brick industry. *Moscow Region Chronicle*, 1. https://www.bogorodsk-noginsk.ru/narodnoe/41_kudinovo.html

Mironov B.N. (1985). Grain prices in Russia in two centuries (XVIII-XIX centuries). L.: Science

Mironov B.N. (2005). Is it possible to see the whole of Russia from Malyye Pupki? Into the circle of ideas: algorithms and technologies of historical data: Papers of the 9th Conference of the Association "History and the Computer" (pp. 528–543). Barnaul: AltGU

Perkovsky A.L. (1977). The crisis of the demographic reproduction of the serfs in Russia in the first half of the 19th century. In A.G. Vishnevsky (Ed.), *Marriage, fertility, mortality in Russia and the USSR*. *Collected articles* (pp. 167-190), Demoscope Weekly, 493-494. http://www.demoscope.ru/weekly/2012/0493/nauka03.php

Rashin A.G. (1956). The Population of Russia over 100 years (1811-1913), Demoscope Weekly. http://www.demoscope.ru/weekly/knigi/rashyn/rashyn.html

Ryazanov V.A. (2021). Peculiarities of Demographic Processes in European Russia in the 18th-19th Centuries: Experience in analyzing the parish books of three Orthodox parishes. *Demographic Review*, 8(3), 20-41.https://doi.org/10.17323/demreview.v8i3.13265

Stolbov V.P. (2013). Manorial peasants of the Sheremetev family. *Historical and Economic Research*, 14(3).

Vinnik M.V. (2012). Seasonality of demographic processes (on the example of registers of the Intercession parish of Barnaul city, 1877-1886). In M.B. Denisenko (Ed.), *Demographic features of socio-economic development* (No. 22, Ch. 10, pp. 251-267). M.: MAX Press. https://publications.hse.ru/mirror/pubs/share/folder/bzom5qu80v/direct/77847147.pdf

Vishnevsky A.G. (Ed.) (2006). *Demographic modernization of Russia*, 1900-2000. Moscow: New Publishing House. https://wwww.hse.ru/data/307/183/1235/modernizacija_vishnevski.pdf

Vladimirov V.N., Sarafanov D.E., Shchetinina A.S. (2019). Traditional and new historical demography: a professional view. *Bulletin of Tomsk State University*, 443, 99-105. https://doi.org/10.17223/15617793/443/13

U.S. Census of Population (1870). Vol. I

Appendix

Demographic indicators of parishes and reference votchinas, aggregated by decade

	Village	1770s	1780s	1790s	1800s	1810s	1820s	1830s	1840s	1850s	1860s
Average age of groom (first marriage), years	Biserovo			19.0	20.9	22.9	22.2	24.6	25.5	25.2	25.4
	Kudinovo			18.6	18.3	18.5	19.9	21.0	22.2	22.6	23.1
Average age of bride (first marriage), years	Biserovo			14.7	20.0	20.5	22.0	20.9	22.4	22.3	20.9
	Kudinovo			17.7	17.5	18.7	18.0	19.7	20.1	20.2	20.8
Conceptions during fasts, %	Biserovo	14,3	19.5	21.1	17.9	17.5	19.3	20.6	19.0	19.6	21.9
	Kudinovo	16.1	15.1	22.2	18.9	19.2	18.4	21.5	18.0	19.4	18.9
Births of children conceived out of wedlock, %	Biserovo	0.0	0.0	0.0	0.9	0.0	1.0	0.7	2.4	1.8	1.2
	Kudinovo	0.0	0.8	3.9	3.0	2.9	4.6	2.9	3.1	1.2	0.2
Children per marriage (if any, without adjustment, by wedding year), pers	Biserovo	4.91	4.72	4.13	4.42	5.27	5.35	5.73	4.66	2.68	1.04
	Kudinovo	5.67	5.29	4.54	5.30	5.27	6.08	5.96	5.56	3.21	1.16
The ratio of the number of newborn girls and boys	Biserovo	1.25	1.18	0.92	0.91	1.12	0.87	1.02	0.99	0.91	0.86
	Kudinovo	0.89	1.01	0.62	0.76	0.78	0.97	0.93	1.06	0.98	1.04
Median age of deaths over 15 years, years	Biserovo	58.5	60	52	61	57.5	62.5	60	54	54	51
	Kudinovo	65.5	55	57	52	47	61.5	52	54	55.5	55
Survived to adulthood and married (boys), by year of birth, %	Biserovo	23.1	17.0	22.5	20.0	31.2	25.7	20.7	3.9		
	Kudinovo	38.5	13.5	27.3	25.4	27.1	38.7	28.4	4.9		

Демографическое обозрение / Demographic Review
2022;9(2) 42-64 DOI: https://doi.org/10.17323/demreview.v9i2.16205

Digital traces of the population as a data source on migration flows in the Russian Arctic

Andrey V. Smirnov
(av.smirnov.ru@gmail.com), Institute for Socio-Economic and Energy Problems of the North, Komi Science Centre of the Ural Branch of the Russian Academy of Sciences, Russia.

Abstract: *The digitalization of the economy and public life has expanded the possibilities of studying the population using digital traces – information that accumulates in the digital environment. Using digital traces, the article explores the migration of the population of the Russian Arctic, a huge macro-region that has experienced a significant outflow of population over the past decades. The text summarizes the experience of using digital traces in demographic research and formulates their strengths and limitations. Data from several digital platforms were used to study the population of the Russian Arctic. An analysis of the profiles of users of the social network VK.com made it possible to study the migration movements of the population of the Russian Arctic, and the data of the ticket service Tutu.ru provided information on air and rail movements. Using network analysis methods, the author studied migration and transport flows in the Russian Arctic at the municipal level. The article defines the features of migration and transport networks in the Arctic: low density, large distances between nodes, high relative mobility with small volumes of movements in absolute terms, a high proportion of hubs in migration exchange. The author identifies migration hubs and clusters, and migration flows are classified according to the directions of movement and types of municipalities. The text shows that the connectivity of the Arctic territories among themselves remains low, and the positive migration balance is mainly in regional capitals or cities outside the Arctic. The results of the study will improve the understanding of migration processes in the North and the Arctic, as well as the quality of demographic forecasts through more accurate modeling of migration flows.*

Keywords: *digitalization, digital traces, social networks, migration, transport network, migration flows, the Russian Arctic.*

Funding: *The article has been prepared with the support of the Russian Science Foundation, the project No 21-78-00081 "Development of tools for studying demographic processes in the context of the digitalization of society (in the case of the Russian Arctic)".*

For citation: *Smirnov A. (2022). Digital traces of the population as a data source on migration flows In the Russian Arctic. Demographic Review, 9(2), 42-64.* https://doi.org/10.17323/demreview.v9i2.16205

Introduction

In the modern world, almost all types of human activity are reflected in the Internet space. Therefore, digital data sources that preserve the history of human interaction with the Internet environment are increasingly becoming the subject of social science reflection (Katzenbach, Bächle 2019). Scientists analyze digital traces – "footprints" of human activity in the digital space, such as search queries, profiles and messages on social media, purchase information and data from global positioning systems (Dudina 2021: 5). New data sources provide information about society on a "massive and microscopic" scale at the same time (Golder, Macy 2014: 131). When studying migration, they make it possible to examine an enormous number of migratory flows with a high degree of detail and to reveal underlying patterns (Smirnov 2022).

In this article, the object of study is the spatial mobility of the population of the Arctic Zone of the Russian Federation, a macro-region in the north of the country which has already lost a third of its population over the past 3 decades, primarily as a result of interregional migration outflow (Fauzer, Smirnov 2020). Population decline persists today, but its study is hampered by the extreme unevenness and mosaic nature of demographic phenomena in the Arctic. The purpose of the study is to identify the spatial patterns of migration of the population of the Russian Arctic using digital traces. Along with migration, it will look at transport passenger flows, which characterize the connectivity of the Arctic territories with each other and with cities outside the Arctic zone. The information base of the study, in addition to official statistics, is made up of data from digital platforms: search engines, social media and ticketing services. To process these data, we used the methods of network analysis, a tool for studying the relationship between objects of any nature (Danchev, Porter 2021) - in the case of migration, between the territories of departure and arrival of people.

The article starts by systematizing the experience of using digital traces in demographic research and summarizing their advantages and disadvantages. Then, using new digital data sources, it looks at migration in the Russian Arctic at the regional and municipal levels. By analyzing migration and transport networks it reveals patterns in the movements of the population. At the end of the article, conclusions are drawn about the redistribution of the human potential of the Russian Arctic and some unresolved scientific problems are formulated.

Digital traces as a source of demographic data

Digital traces are the results of social interaction via digital tools and spaces as well as digital records of other culturally relevant materials (Cesare et al. 2018: 1980). The revolution in the use of digital traces by science has occurred thanks to the transition from small to large data. As a result of this transition, "the generation of data is continuous, exhaustive to a system, fine-grained, relational, and flexible across a range of domains" (Kitchin 2021: 61). Big data is being accumulated both in government information systems and on private digital platforms. The state collects data related to registration at the place of residence, employment, payment of taxes, visits to medical institutions, and receipt of various public services. Transport and utilities have become a digital network, equipped with numerous digital sensors that record people's behavior. Data from various sources is accumulated within large digital ecosystems covering many areas of life. A striking example is the Unified Portal of Public Services of the Russian Federation, which already has over 100 million users (Smirnov 2021: 148). As a result of the implementation of the national project "The Digital Economy" in Russia, we can expect an even greater deepening of the digitalization of society and public administration. In particular, by 2024 it is planned to transfer 70% of information systems and resources of federal authorities to a single cloud platform[1].

In addition to government information systems, big data is generated by commercial companies such as mobile phone operators (location, app use), travel and accommodation sites (reviews), social media (opinions, photos, personal info, location), transport providers (routes, traffic flows), website owners (clickstreams), financial institutions and retail chains (purchases), private surveillance and security firms (location, behavior) (Kitchin 2021: 62). Accumulation of data allows technology companies to extract greater profits through the vertical and horizontal integration of digital platforms. “In the twenty-first century advanced capitalism came to be centred upon extracting and using a particular kind of raw material: data. <...> Like oil, data are a material to be extracted, refined, and used in a variety of ways. The more data one has, the more uses one can make of them.” (Srnicek 2020: 37).

Big data is also generated through crowdsourcing and research projects. In Russia, the “Research Data Infrastructure” (RDI, data-in.ru) project should be noted, within the framework of which many valuable datasets have been prepared, for example, on the number of voters by precinct election commissions with reference to geographic coordinates in 2020. We will demonstrate this data set using the example of the Russian Arctic (Figure 1). The maps show that the data accumulated from the Internet pages of the election commissions, on the whole, quite accurately reflect the settlement system. Moreover, they evaluate the distribution of the population within cities in more detail - by polling stations.

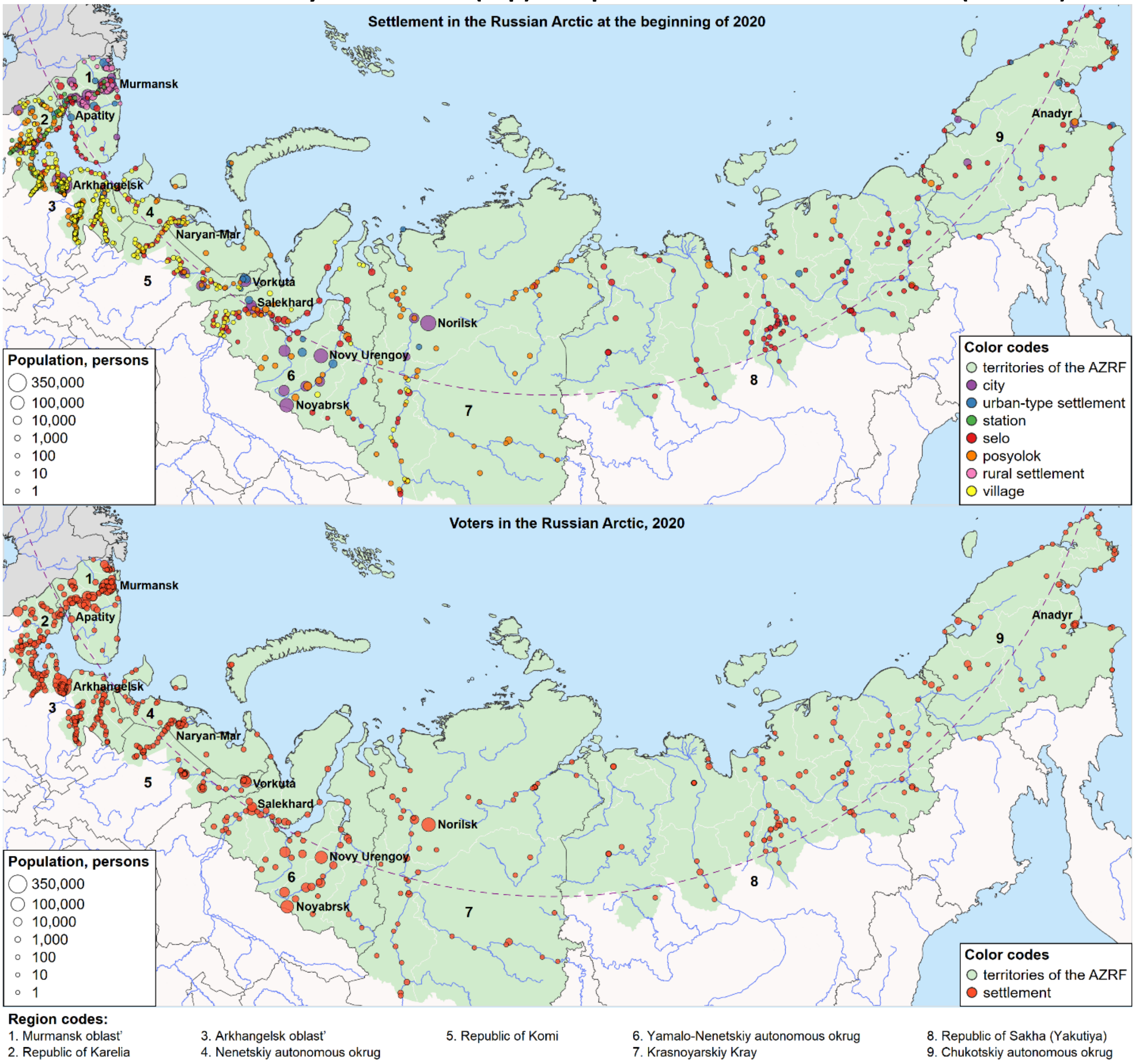
The advantages of digital data sources include large geographical coverage, continuous generation, and speed of collection and processing. They provide data on controversial topics because they are less biased by respondents' choice of socially acceptable responses than traditional surveys (Cesare et al. 2018: 1981). D. Leiser and J. Radford identify 3 types of digital data according to their sources and how they are acquired: digital life (the capturing of digitally mediated social behaviors), digital traces (only a record of the action, not the action itself) and digitized life (the movement of intrinsically analog behavior into digital form). Thanks to the development of analysis methods, research is increasingly using not only numerical and textual data obtained from the digital space, but also images, audio and video (Lazer, Radford 2017: 21-22, 33).

The disadvantages of new digital data sources include low representativity, fragmentation, vulnerability to changes, the possibility of algorithm errors, the presence of false information, bots and spam accounts, low reliability, duplication of information and limited access to data (Golder, Macy 2014; Lazer, Radford 2017). Digital data is difficult to interpret when not generated for research purposes, and digital concepts and terms may differ from theoretical ones. The problem of underrepresentation can be partly addressed by post-stratification or other bias correction techniques (Hughes et al. 2016). Often research questions are formulated in such a way that data correction is not required at all, for example, when the object of study is a virtual rather than a real population. Methods are being developed to reconcile digital trace data with other data sources, including census microdata (Alburez-Gutierrez et al. 2019). Digital research also faces many ethical challenges (Taylor, Floridi, van der Sloot 2017). To ensure data confidentiality, protect users from possible discrimination, and achieve reproducibility of studies,

[1] Passport of the federal project "Information infrastructure". https://files.data-economy.ru/Docs/Pass_Infrastructure.pdf

special open algorithms for collecting and processing digital data are being developed (Cesare et al. 2018: 1985).

Figure 1. Settlement in the Russian Arctic according to Rosstat, medical information and analytical centers (top) and precinct election commissions (bottom)



Source: Compiled by the author based on RDI datasets (http://data-in.ru/data-catalog/datasets/160; http://data-in.ru/data-catalog/datasets/203) using Natural Earth geodata (https://www.naturalearthdata.com/).

The development of methodological possibilities of research in connection with the introduction of digital data leads to attempts to develop theoretical optics suitable for new tools for the social sciences. It is proposed to turn digital traces into an independent object of study, and to rethink the problems of connection between micro- and macrolevels on the basis of D. Boullier's replication theory, which goes back to the works of G. Tarde (Dudina 2021). Applying the monadology of Leibniz to social phenomena, G. Tarde argues that “all phenomena are nebulous clouds resolvable into the actions emanating from a multitude of agents” (Tarde 2016: 32). According to Tarde, these figures do not have a coordinating center. D. Boullier distinguishes 3 stages in the development of sociological methods: the first source of data was statistics and censuses, then public opinion polls, and now digital traces. According to Boullier, digital traces reflect replications (repetitions, copying) of actions, ideas and practices (Boullier 2017). At the same time, digital platforms are perceived as a kind of “replicating machines”, “allowing the dissemination of digital traces and making them available for research” (Dudina 2021: 5). Thus, the independent research value of digital traces is postulated. G. Igntatow comes to similar valuable conclusions from a practical point of view, reflecting on the theoretical foundations of the analysis of digital texts. He proposes to consider discourses as real emerging social phenomena, which allows them to be analyzed by rigorous methods (Igntatow 2016: 108).

The impact of new digital data sources on the social sciences is seen as revolutionary by many scholars (Kitchin 2014; Ledford 2020). In demography, digital traces began being used relatively recently, but they are already being used to solve a wide range of problems. Thus, images of cars from street panoramas are used to assess the socio-demographic characteristics of areas (Gebru et al. 2017). Valuable data is extracted from search engines and social media (McCormick et al. 2017; Zagheni, Weber, Gummadi 2017). Texts published by users on the web are used to analyze reproductive, self-preservation, matrimonial and migratory behavior. They can be studied both by frequency methods, using keywords, and by machine learning methods that can classify texts and highlight their semantic content and emotional coloring. For example, using automatic extraction and analysis of the opinions of social media users, it is possible to study various aspects of the reproductive behavior of the population (Kalabikhina et al. 2021).

Digital platforms are particularly useful when national statistics are unreliable for some processes (Cesare et al. 2018) or populations are studied to which access is difficult or costly (Edelmann et al. 2020). Mobile phone data and message geotagging are used to track the spatial mobility of the population (Hughes et al. 2016). For example, they were used to study compliance with self-isolation measures in various regions during the COVID-19 pandemic (Petrov et al. 2021: 9). The movements of specialists and scientists can be analyzed by the CVs and affiliations of scientific publications (Sudakova 2020). It has been shown that search query data can be used for short-term forecasts of fertility trends (Billari, D’Amuri, Marcucci 2013) and outbreaks of morbidity and mortality during a pandemic (Ahmad, Flanagan, Staller 2020). Digital traces are often used in conjunction with official statistics and the results of sociological research, complementing each other.

Characteristics of the Russian Arctic. Methods and data

The Arctic zone of the Russian Federation for 2022 includes 75 city okrugs and municipal districts[2] in 9 northern regions of Russia (Figure 1). At the beginning of 2022, 2592.9 thousand people lived in the Russian Arctic (excluding the results of the 2021 population census). The Arctic zone accounts for about 30% of the area, 1.8% of the population and 6% of Russia's gross regional product. The key sector of the economy is resource extraction. About 90% of Russia's natural gas, a significant share of oil, coking coal, and non-ferrous metals are produced in the Arctic. Economic specialization, spatial remoteness, and uncomfortable climatic conditions have an impact on demographic structures (Heleniak, Bogoyavlenskiy 2014; Fauzer, Lytkina 2017; Zamyatina, Yashunsky 2017). Since the population of the Arctic is relatively young and lives mainly in cities, it is characterized by high values of digitalization indicators. More than 90% of

[2] Federal Act of 13 July 2020 entitled "State support for entrepreneurship in the Arctic zone of the Russian Federation".

the population from 15 to 74 years old are active Internet users[3], which corresponds to the level of the most developed countries in the world. Therefore, digital traces can be a fairly representative source of data on Arctic populations.

Official statistics on the size and migration of the population of the Arctic territories were obtained from the Unified Interdepartmental Statistical Information System[4] (UISIS) and the Database of Municipal Indicators[5] (DMI) of Rosstat. Digital traces of the population were analyzed using three data sources: the Yandex service "Keyword Statistics", the project "Virtual Population of Russia" and the Tutu.ru service data set. Let's look at their capabilities and limitations.

Yandex Keyword Statistics[6] is a service for assessing user interest in topics. It allows you to get information about the popularity of a particular query in the search engine in the context of regions. The names of the regions of the Arctic were entered as queries, which made it possible to assess the interest that Yandex users in some regions had in other regions. The presence of search queries does not guarantee that users are planning to move to or visit a region. Nevertheless, interest in a region can characterize the intensity of cultural, social or economic interactions.

The project "Virtual Population of Russia"[7] was implemented with the support of the Russian Geographical Society and the Keldysh Institute of Applied Mathematics of the RAS. It contains geo-referenced data for January-March 2015 from user profiles of the most popular social network in Russia at that time, VK. The data of the project make it possible to analyze migratory movements at the regional and municipal levels by age groups, as well as friendships between people. Of the 88 million accounts that indicate a place of residence or a place of study, 9 million have more than one place of residence, which makes it possible to analyze migration flows. The migration dataset takes into account only the chronologically last change of location for each user. The set's limitations include the availability of data for only one point in time, as well as the fact that people tend to indicate not the municipality where they actually live, but the nearest large city to it (Zamiatina, Yashunsky 2018). In addition, users are less likely to record short-term and return migrations.

The dataset of the Tutu.ru ticketing service[8] about traveling around the country was created to predict the spread of the coronavirus infection COVID-19[9]. It contains information on the number of movements between cities by plane, train and bus[10] in April 2019. The number of passengers is not limited to the number of tickets sold through the Tutu.ru service, but has been restored to 100%. According to the developers of the dataset, buses are the most inaccurate part of the dataset due to the presence of "grey" carriers. For the Arctic settlements, there are only 11 routes (mainly from Arkhangelsk and Petrozavodsk). Therefore, in this study, we restrict

---

[3] Statistical information on socio-economic development of the Arctic zone of the Russian Federation. Rosstat. https://rosstat.gov.ru/storage/mediabank/arc_zona.html

[4] Unified interdepartmental statistical information system. Rosstat and Ministry of Digital Development of Russia. https://www.fedstat.ru/

[5] Database of municipal indicators. Rosstat. https://www.gks.ru/dbscripts/munst/

[6] Keyword Statistics. Yandex. https://wordstat.yandex.ru/

[7] Virtual population of Russia. http://webcensus.ru/

[8] Dataset Tutu.ru and data from Open Data Science model. https://story.tutu.ru/dataset-tutu-ru-i-dannye-modeli-open-data-science/

[9] Scenarios of infection in specific cities based on the dataset of movement of people in Russia. Habr. https://habr.com/ru/company/tuturu/blog/494700/

[10] Covid19-tutu. GitHub. https://github.com/ods-ai-ml4sg/covid19-tutu/blob/master/data/raw_data.csv

ourselves to data on aircraft (558 routes) and trains (712 routes). Air and rail data also have limitations. In the eastern part of the country, market coverage is worse. This applies in particular to helicopter routes between the cities of the Far East and the propeller-driven aviation of Yakutia.

Migration and transport flows were studied using network analysis methods (Danchev, Porter 2021). Network science tools are used to study migration both at the interregional (Maier and Vyborny 2008) and intercountry (Danchev and Porter 2018) levels. A network is a set of nodes and a set of links (edges) between them. Migration and transport networks are most conveniently represented as directed and weighted. The directions of the links correspond to the directions of movement, from the place of departure to the place of arrival, while the weights correspond to the number of people who have moved. Six networks were built. Three of them include only links with at least one node in the Arctic: migration, air and railway passenger flows. Three more similar networks were built according to data for the whole of Russia and used for comparative analysis.

The limitations of the study include the fact that the data sources used are not in all cases synchronous. They refer to the period from 2015 to 2022. It should also be noted that migration and traffic flows characterize different types of migration: official data from Rosstat that of long-term migration, and data from ticketing services that of any movements, including short-term ones for recreation, medical treatment, education, work or family. Estimating rotational migration, which is important for the Arctic, is complicated by the fact that rotational workers are difficult to separate from other passengers, and travel to the work site may be not on regular, but on charter flights, on which data are less available.

Using the NetworkX package in the Python programming language, algorithms for calculating the main characteristics of migration and transport networks were developed. To search for communities (clusters) in networks, an asynchronous label propagation algorithm (Raghavan, Albert, Kumara 2007) was used, based on the idea that related nodes usually belong to the same cluster. Network visualization was carried out using the GraphPlot.jl and Graphs.jl packages in the Julia language. Network visualization algorithms tend to place nodes in such a way that connected nodes are close to each other and the number of link intersections is minimized. For networks with more than one cluster, a modified force-directed graph drawing algorithm of the Fruchterman-Reingold was used (Fruchterman, Reingold 1991). For the network of aviation passenger flows, in which it is difficult to distinguish clusters, the stress majorization graph drawing algorithm was used (Gansner, Koren, North 2004). The VegaLite.jl package was used to create chart maps, and the chorddiag package in the R programming language was used to build a chord diagram.

Connectivity of the Arctic territories: digital data and official statistics

Connectivity is understood as the degree of connection (link) of cities or territories, expressed through the presence and quantity of migration, transport movements, or other interactions between them. First, we will look at the connectivity of the Arctic territories with each other at the regional level (Figure 2). We will analyze four indicators, three of them based on digital traces of the population and the last one on official statistics:

- the number of friendships per person of the virtual population of the region with residents of another region according to the social network VK in January-March 2015. The indicator is not symmetrical. Although the two regions have the same number of friendships in both directions, their virtual populations differ. Therefore, for example, one virtual resident of Karelia has more friendships with residents of the Murmansk region than vice-versa;

- the number of migration movements per 1,000 people of the virtual population, according to the data of the VK social network. The chronologically last change of residence according to user accounts for January-March 2015 is used;
- the popularity among a region's residents of a search query with the name of another region in the Yandex search engine. Regional popularity is a region's share of displays for a given query divided by the share of all search results displayed for that region.
  If the regional popularity exceeds 1, there is a high interest in this request in the region; if it is less than 1, it is low;
- the number of migration movements per 1,000 people of the real population by regions of departure and arrival according to UISIS data. The average annual value for 2015-2021 is used.

Figure 2. Some indicators of the connectivity of the Arctic regions

		a) Number of friendships per person of the virtual population of the region (Virtual Population of Russia, 2015)									
		connected region									
		№	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
analyzed region	Murmansk obl.	(1)		0.55	0.51	0.01	0.12	0.02	0.14	0.04	0.00
	Rep. Karelia	(2)	0.77		0.31	0.00	0.10	0.02	0.13	0.01	0.00
	Arkhangelsk obl.	(3)	0.42	0.18		0.16	0.36	0.03	0.15	0.08	0.00
	Nenetskiy AO	(4)	0.21	0.09	6.50		1.38	0.17	0.13	0.02	0.00
	Rep. Komi	(5)	0.13	0.08	0.51	0.05		0.04	0.13	0.01	0.00
	Yamalo-Neneetskiy AO	(6)	0.05	0.03	0.09	0.01	0.10		0.16	0.02	0.00
	Krasnoyarskiy Kray	(7)	0.05	0.04	0.07	0.00	0.04	0.02		0.07	0.00
	Rep. Sakha (Yakutia)	(8)	0.06	0.01	0.15	0.00	0.02	0.01	0.30		0.00
	Chukotskiy AO	(9)	0.07	0.03	0.08	0.00	0.03	0.02	0.16	0.13	

	b) Number of migration movements per 1000 persons of the virtual population (Virtual Population of Russia, 2015)									
	region of arrival									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	№
region of departure		1.54	0.84	0.01	0.17	0.04	0.30	0.10	0.01	(1)
	2.09		0.34	0.01	0.09	0.03	0.11	0.02	0.00	(2)
	1.30	0.48		0.30	1.15	0.08	0.19	0.04	0.00	(3)
	0.97	0.44	24.6		3.30	1.31	0.24	0.05	0.00	(4)
	0.31	0.16	0.87	0.10		0.19	0.19	0.03	0.00	(5)
	0.09	0.03	0.10	0.02	0.27		0.29	0.01	0.00	(6)
	0.16	0.05	0.08	0.00	0.05	0.06		0.11	0.01	(7)
	0.22	0.07	0.45	0.01	0.09	0.11	1.79		0.02	(8)
	0.77	0.51	0.84	0.00	0.51	0.64	3.02	1.22		(9)

		c) Regional popularity of search query naming other region (Yandex Keyword statistics, 13.04.2022)									
		requested region									
		№	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
analyzed region	Murmansk obl.	(1)		2.64	1.11	0.99	0.68	0.72	0.13	0.50	0.68
	Rep. Karelia	(2)	3.35		0.78	0.75	0.63	0.55	0.08	0.57	0.88
	Arkhangelsk obl.	(3)	1.20	1.09		4.79	1.43	0.87	0.10	0.54	0.84
	Nenetskiy AO	(4)	1.08	0.71	11.1		8.92	4.95	0.13	1.19	1.85
	Rep. Komi	(5)	0.52	0.57	1.35	2.25		1.38	0.10	0.68	0.75
	Yamalo-Neneetskiy AO	(6)	0.27	0.43	0.37	1.08	0.87		0.23	0.94	2.13
	Krasnoyarskiy Kray	(7)	0.18	0.29	0.19	1.27	0.43	0.74		1.03	0.88
	Rep. Sakha (Yakutia)	(8)	0.27	0.35	0.21	0.80	0.89	0.68	0.36		2.46
	Chukotskiy AO	(9)	0.62	0.69	0.45	3.26	0.49	3.38	0.57	6.41	

	d) Number of migration movements per 1000 persons of real population of region of departure (Rosstat, average for 2015-2021)									
	region of arrival									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	№
region of departure		1.50	1.22	0.02	0.17	0.04	0.17	0.03	0.02	(1)
	1.72		0.27	0.01	0.07	0.02	0.05	0.00	0.01	(2)
	1.04	0.19		0.35	0.55	0.05	0.05	0.01	0.01	(3)
	0.40	0.06	9.49		2.13	0.16	0.16	0.04	0.00	(4)
	0.24	0.08	0.82	0.12		0.17	0.07	0.02	0.01	(5)
	0.07	0.03	0.10	0.01	0.22		0.25	0.04	0.01	(6)
	0.05	0.01	0.02	0.00	0.02	0.05		0.16	0.02	(7)
	0.02	0.01	0.01	0.00	0.01	0.03	0.56		0.02	(8)
	0.31	0.18	0.11	0.00	0.16	0.17	0.85	0.40		(9)

Source: Compiled by the author based on data from the Virtual Population of Russia project, Yandex (dated April 13, 2022) and Rosstat.

In general, regions that are geographically close to each other are more connected, as evidenced by the green colors of many cells located near the diagonal of the tables. All connectivity indicators reach the highest value between the Nenets Autonomous Okrug and the Arkhangelsk Oblast, of which it is a part. Low connectivity values are recorded in Yakutia, where the population is predominantly rural, with no large cities (an exception is the connection between Yakutia and the Krasnoyarsk Territory). The linear correlation between migration according to official statistics and social network data is 0.851 (n = 71, links from the Nenets Autonomous Okrug to the Arkhangelsk region are excluded as statistical outliers). The correlation between migration according to statistics and friendships has a high value – 0.789. The correlation of migration with the intensity of search queries is significantly lower – 0.542.

It can be concluded that, although in absolute terms migration rates obtained from official statistical databases and digital traces of the population differ, they show similar patterns. The connectivity of territories in the digital environment usually means that in reality there is indeed a high connectivity (migration, social, cultural) between them. Moving on to the next level of detail, let’s consider the networks of intermunicipal movements and their main characteristics (Table 1).

Table 1. Networks of migration and passenger movements of the population of the Russian Arctic and Russia

Indicator	Migration		Air transport		Rail transport	
	AZRF	Russia	AZRF	Russia	AZRF	Russia
Size of network (number of nodes)	2112	2201	85	173	160	574
Number of links	32199	334529	558	2951	712	12125
Strongly connected	no	no	no	no	no	no
Weakly connected	yes	yes	yes	yes	yes	no
Reciprocity	0.421	0.419	0.828	0.855	0.775	0.760
Average link weight	6.1	11.0	978.1	2071.6	384.2	689.9
Average distance of movement, km	…	…	1707.7	1632.9	956.3	601.3
Average degree	30.5	304.0	13.1	34.1	8.9	42.2
Average strength	186.9	3319.1	12842.1	70674.0	3419.8	29146.4
Network density	0.007	0.069	0.078	0.099	0.028	0.037
Average path length	1.875	1.936	1.925	2.246	2.489	2.236 *
Diameter of network **	4	3	5	5	6	5 *
Heterogeneity parameter	14.675	2.475	3.157	2.581	3.934	3.409
Degree_assortativity_coefficient	-0.519	-0.202	-0.635	-0.364	-0.644	-0.213
Clustering coefficient	0.525	0.468	0.302	0.507	0.398	0.640
Number of clusters when decomposing by label propagation method	2	1	1	1	4	6

Source: Compiled by the author based on data from webcensus.ru and Tutu.ru.

*Note: * - to calculate the value, the network was converted to weakly connected by removing 5 nodes related to an isolated section of the railway on Sakhalin; ** - to calculate the diameter, all networks were converted to non-directional. AZRF - Arctic zone of the Russian Federation.*

In terms of the size and number of links, migration networks lead, since they are not limited by the capabilities of the transport infrastructure and can connect any settlements. Of the transport networks, rail is larger than air, since populated areas have fewer airports than railway stations. However, in the Arctic, due to its remoteness, the proportion of cities with airports is higher than in the country as a whole. The Russian railway network is not connected, as it includes an isolated railway on Sakhalin Island. When calculating some indicators, this railway was not included. In transport networks, the indicator of reciprocity is higher, i.e., movements are more often carried out in both directions.

The average link weight (average number of moves per flow) is higher in air networks due to there being fewer possible routes. In air travel, the average travel distance is also higher than by rail - for the Arctic, almost twice as high (1707.8 versus 956.3 km). The average degree of a node shows the average number of links or neighbors of nodes throughout the network. In the Arctic migration network, the average degree is an order of magnitude lower than in the Russian migration network. In transport networks, the Arctic's lag is also large due to its remoteness from the main centers of settlement. This is also evidenced by the lower values of density of the Arctic networks. Since transport and migration networks are weighted, it is possible to calculate a weighted degree for them - the node strength. The patterns are similar and even more pronounced.

The shortest path is the minimum number of links that must be made along the path connecting 2 nodes. The average path length is calculated by averaging the lengths of the shortest paths over all pairs of nodes. The lowest values of the average path length are observed in migration networks, due to the large number of links. In Arctic networks, the diameter (the length of the longest shortest path) is greater than or equal to the value for the whole of Russia. The largest diameter (6 movements) is in the railway network, and the smallest (4) in the migration network.

The more hubs (nodes with a higher degree) in the network, the higher the heterogeneity parameter. In the Arctic, heterogeneity is higher and is highest in migration networks. Hubs are Moscow, St. Petersburg and the administrative centers of the regions. However, in other networks too the heterogeneity parameter is quite high. There are hubs in all six networks; they will be discussed in more detail below. In all networks, the assortativity coefficient is negative. This suggests that high-degree nodes (hubs) are more often connected to low-degree ones. Moreover, this dependence is manifested more clearly in the Arctic networks.

The clustering coefficient shows the proportion of pairs of neighbors of a node that are connected to each other. In the Arctic, it is the highest in migration networks. As a rule, either hubs or settlements located close to each other are more connected. In air transport, the network clustering coefficient is relatively low. The maximum number of clusters is in railway networks (4 in the Arctic and 6 in Russia). The composition of the clusters will be discussed below.

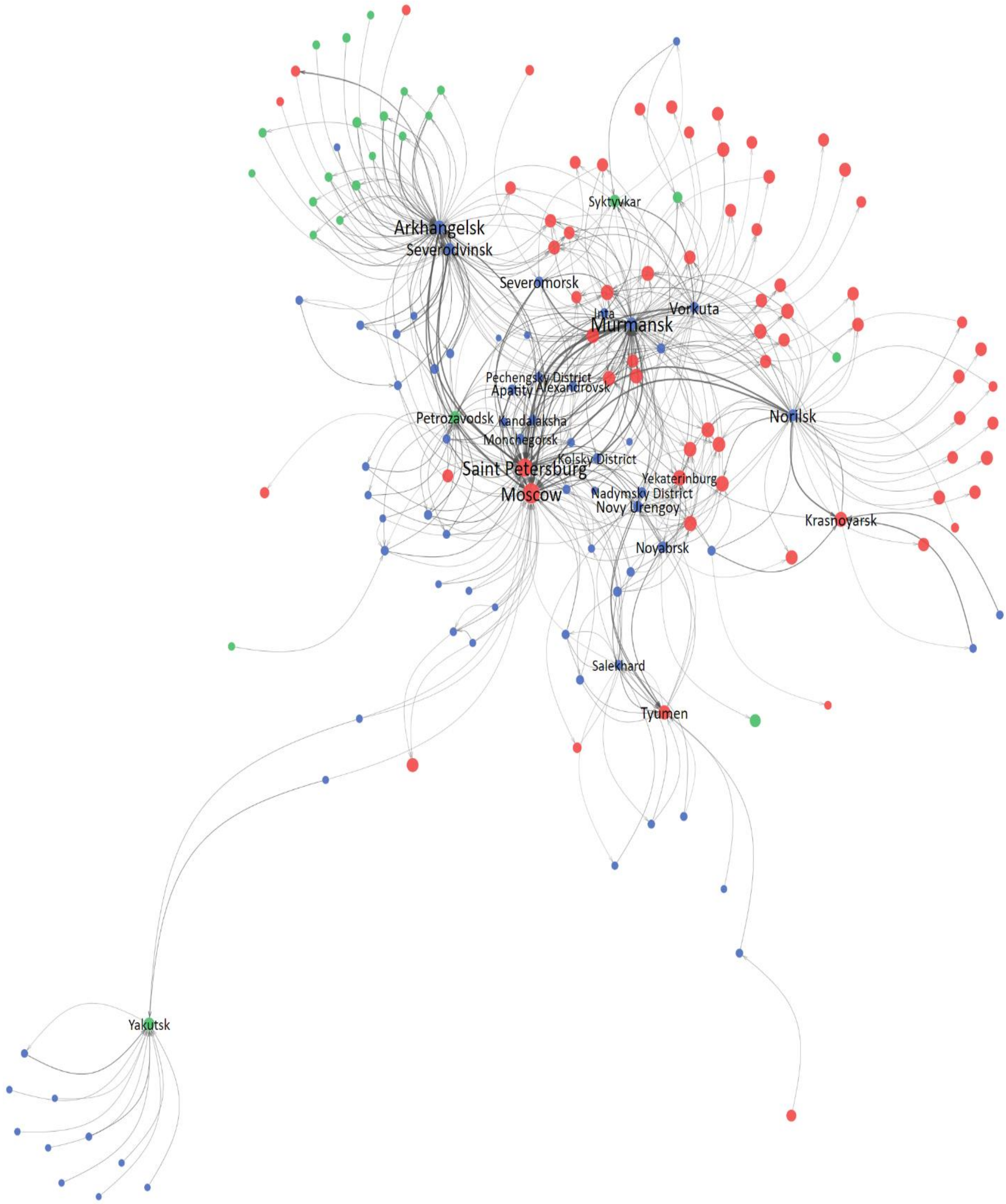
Network analysis showed differences in the patterns of migration and transport movements between Russia and its Arctic part in almost all indicators. The main reason for the differences is the low density of connections, due to the spatial remoteness of the macroregion, as well as the high influence of hubs - the most significant network nodes. Let's focus attention on them.

Migration flows and hubs in the Russian Arctic

Clusters (communities) are sets of nodes with a higher density of connections within than between sets. Two clusters in the network of migration movements have been highlighted (Figure 3). The first includes some rural areas of Yakutia (located in the lower left part of the figure), and the second, the remaining municipalities.

The largest hub in the Arctic migration network is Murmansk (Table 2). It accounts for 17.3% of movements. It is followed by Arkhangelsk (12.9%), St. Petersburg (12.7%), Moscow (8.9%) and Norilsk (8.0%). Moreover, the migration balance is much better in the federal capitals: in St. Petersburg and Moscow, incoming flows greatly exceed outgoing ones. In most Arctic cities, according to digital traces, there persists a negative migration balance. Vorkuta (-7.5 thousand movements) and Norilsk (-7.3 thousand) especially stand out.

Figure 3. Network of intermunicipal migrations in the Russian Arctic



Source: Compiled by the author based on data from the Virtual Population of Russia project.

Note: Only flows larger than 50 people are reflected. Municipal entities of the Arctic zone are marked in blue, the Far North of Russia (except for the Arctic) is marked in green, and other regions of Russia are marked in red. The thickness and brightness of the line is proportional to the size of the flow, the size of the circle is proportional to the population of the municipality.

Table 1. Characteristics of the largest nodes of the network of migration movements of the Russian Arctic

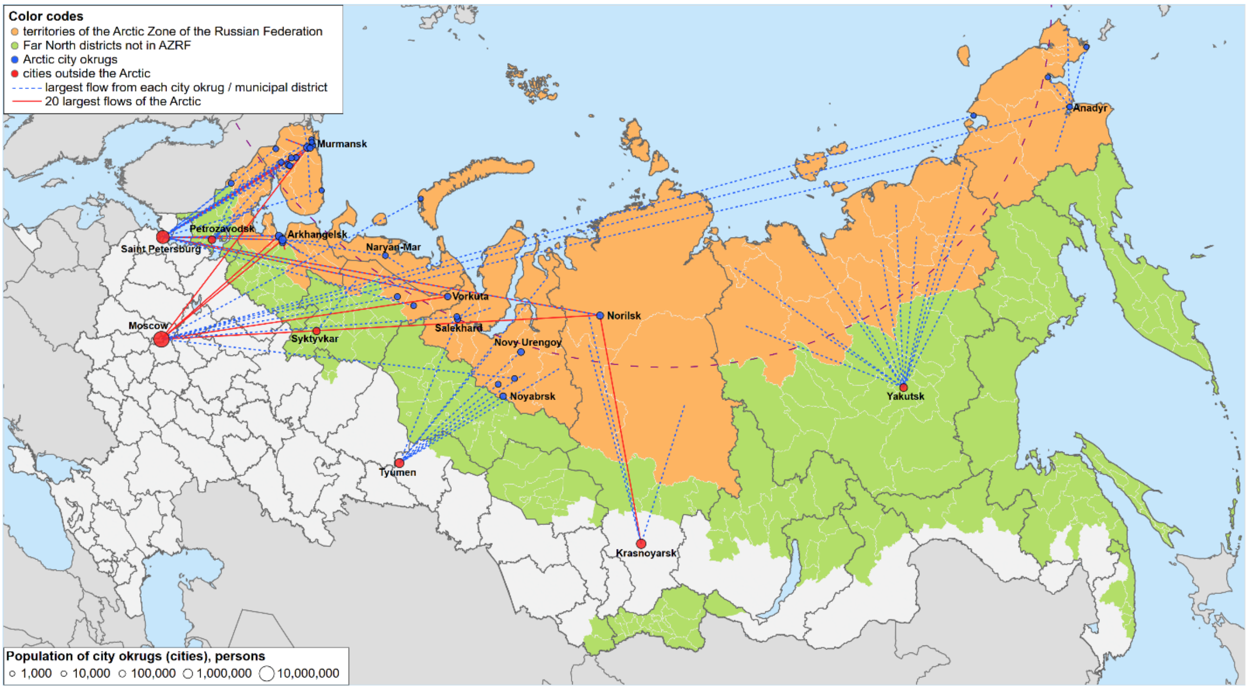
№	City okrug / municipal district	Outflows		Inflows		Balance	
		quantity, units	size, persons	quantity, units	size, persons	quantity, units	size, persons
1	Murmansk	810	16.235	1.421	18.109	611	1.874
2	Arkhangelsk	611	11.167	853	14.473	242	3.306
3	St. Petersburg	64	2.945	74	22.287	10	19.342
4	Moscow	68	3.308	75	14.483	7	11.175
5	Norilsk	753	11.655	985	4.349	232	-7.306
6	Severodvinsk	488	6.292	668	5.122	180	-1.170
7	Vokruta	749	9.283	603	1.734	-146	-7.549
8	Novy Urengoy	396	3.971	911	3.706	515	-265
9	Severomorsk	374	4.671	723	2.636	349	-2.035
10	Noyabrsk	340	3.288	801	2.901	461	-387
11	Apatity	378	3.199	397	1.880	19	-1.319
12	Nadymskiy district	362	3.408	601	1.644	239	-1.764
13	Aleksandrovsk	362	3.234	617	1.728	255	-1.506
14	Krasnoyarsk	35	745	60	3.492	25	2.747
15	Monchegorsk	308	2.650	418	1.264	110	-1.386
16	Inta	416	3.249	237	555	-179	-2.694
17	Petrozavodsk	36	1.130	48	2.607	12	1.477
18	Tyumen	26	485	49	3.239	23	2.754
19	Pechengskiy district	345	2.693	404	937	59	-1.756
20	Usinsk	334	2.113	477	1.515	143	-598
21	Salekhard	231	1.553	490	1.993	259	440
	Total	32.199	199.052	32.199	199.052	0	0

Source: Compiled by the author based on data from the Virtual Population of Russia project.

Note: Ranked in descending order of total movements.

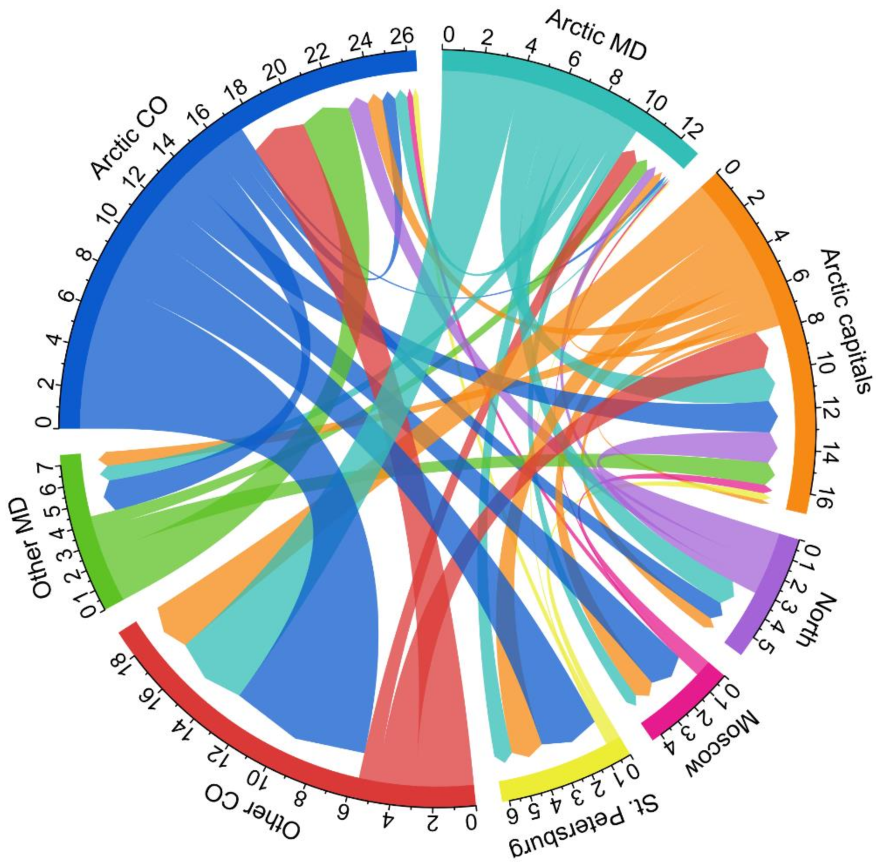
Of the 20 largest flows, only 5 link municipalities both of which are located in the Arctic: from the Kola district and Severomorsk to Murmansk, from Severodvinsk to Arkhangelsk and back, from the Pinezhsky district to Arkhangelsk. Another 1 connects the Arctic with the municipality of the Far North - from the Kholmogorovsky district to Arkhangelsk (Figure 4). The rest connect the Arctic city okrugs with Moscow, St. Petersburg and Krasnoyarsk. Residents of the European part of the Arctic more often move to St. Petersburg, while those in the Asian part move to Moscow (exceptions are Norilsk, Usinsk and Novaya Zemlya). Residents of city okrugs are more likely to move to Moscow and St. Petersburg, while residents of municipal districts are more inclined to move to regional capitals (Figure 5). This may be due to the availability of relocation resources for residents of cities and urban-type settlements, most often specializing in the resource extraction industry or the transportation of natural resources. The administrative centers of the regions located in the Arctic (Arkhangelsk, Murmansk, Salekhard, Naryan-Mar and Anadyr) give about the same number of people as they get. But while people arrive mainly from the northern and Arctic municipalities, they leave for cities outside the Arctic.

Figure 4. The largest migration flows in the Russian Arctic



Source: Compiled by the author based on data from the Virtual Population of Russia project using Natural Earth geodata (https://www.naturalearthdata.com/).

Figure 5. Migration flows of the Russian Arctic by groups of municipalities, %



Source: Compiled by the author based on data from the Virtual Population of Russia project.

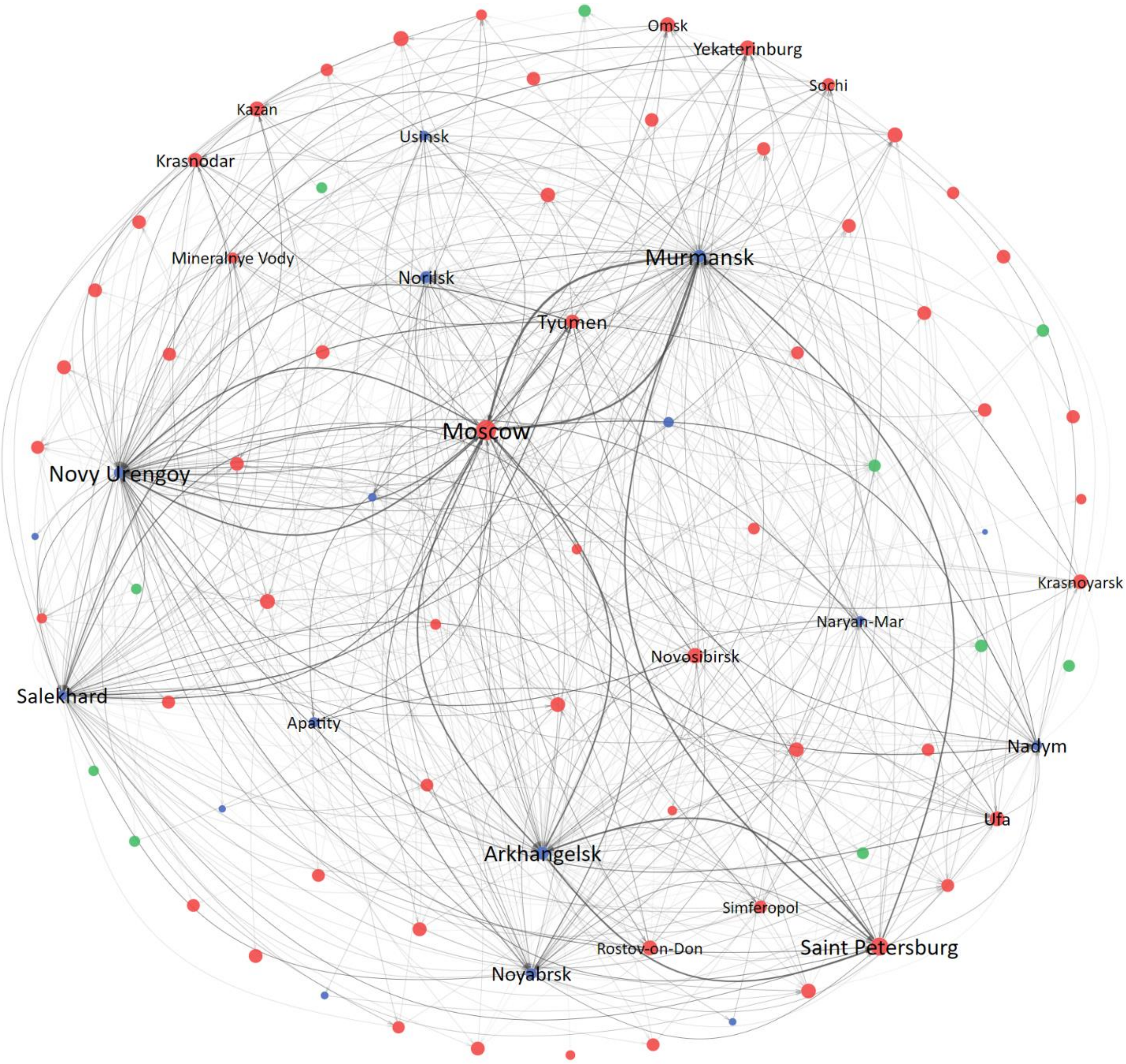
Note: CO – city okrug, MD - municipal district.

Numerical data confirm the main migration trends in the Arctic recorded by official statistics (Fauzer, Smirnov 2020) and reveal them with a high degree of detail. The population is becoming concentrated in a small number of large cities and their environs. Population decline from older resource cities and rural areas remains high.

Transport flows and mobility of the Arctic population

It was not possible to identify explicit clusters in the network of air travel movements. Almost all parts of the Arctic have similar patterns of movement with a high proportion of outflows to Moscow and back (Figure 6).

Figure 6. Air passenger traffic network in the Russian Arctic

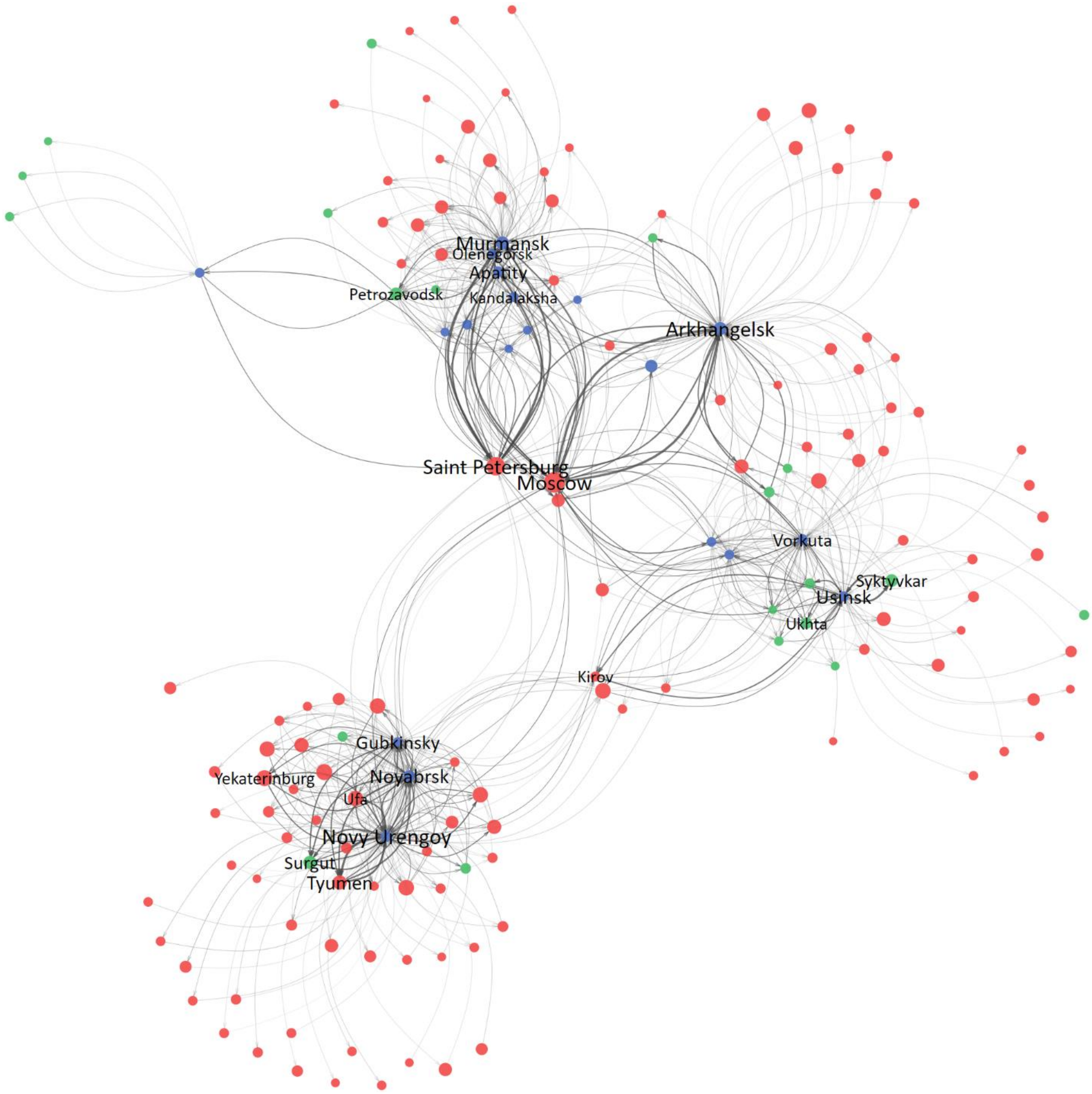


Source: Compiled by the author according to the Tutu.ru service.

Note: Settlements in the Arctic Zone are marked in blue, in the Far North of Russia (except for the Arctic) in green, and in other regions of Russia in red.

In the network of railway movements, clusters are distinguished that roughly correspond to the branches of the Russian Railways (RZD): Oktyabrskaya, Severnaya, Sverdlovskaya (Figure 7). Moreover, the Severnaya Railway is divided into 2 clusters corresponding to branches to Arkhangelsk and Vorkuta. An intermediate position between the cores of the clusters is occupied by Moscow, St. Petersburg and Kirov, whose railways are not included in the listed branches of Russian Railways and are used as transfer hubs.

Figure 7. Rail passenger traffic network in the Russian Arctic



Source: Compiled by the author according to the Tutu.ru service.

Note: Settlements in the Arctic Zone are marked in blue, in the Far North of Russia (except for the Arctic) in green, and in other regions of Russia in red.

In the air travel network, 15.1% of movements are made from Moscow, and another 17.7% - to Moscow (Table 3). Thus, the capital accounts for a third of all movements. In addition to Moscow, there are several other major hubs: Murmansk (25.4% of movements), Novy Urengoy

(22.8%), St. Petersburg (17.4%), Arkhangelsk (16.9%) and Salekhard (12.8%). In the railway network, the share of large hubs is lower. Novy Urengoy accounts for 19.6% of movements, Moscow and Arkhangelsk - 16.5% each, St. Petersburg - 16.0%. The shares of Murmansk (12.9%), Usinsk (11.6%) and Noyabrsk (11.1%) are high. The large passenger turnover of Novy Urengoy is associated, among other things, with rotational migrations.

Table 3. Characteristics of the largest nodes of the network of air and rail passenger traffic in the Russian Arctic

№	City	Outflows		Inflows		Balance	
		quantity, units	size, persons	quantity, units	size, persons	quantity, units	size, persons
				Air transport			
1	Moscow	13	82,372	13	96,451	0	14,079
2	Murmansk	52	76,607	47	62,057	-5	-14,550
3	Novy Urengoy	45	68,159	48	56,435	3	-11,724
4	St. Petersburg	12	47,251	12	47,531	0	280
5	Arkhangelsk	40	43,975	40	48,169	0	4,194
6	Salekhard	42	37,810	33	32,313	-9	-5,497
7	Tyumen	7	21,830	8	24,302	1	2,472
8	Noyabrsk	26	18,952	27	18,001	1	-951
9	Nadym	22	18,444	16	13,590	-6	-4,854
10	Norilsk	34	13,306	24	9,954	-10	-3,352
	Total	558	545,791	558	545,791	0	0
				Rail transport			
1	Novy Urengoy	56	29,643	49	23,982	-7	-5,661
2	Moscow	18	21,960	18	23,280	0	1,320
3	Arkhangelsk	38	22,898	35	22,156	-3	-742
4	St. Petersburg	15	22,047	15	21,630	0	-417
5	Murmansk	33	20,642	28	14,747	-5	-5,895
6	Usinsk	29	17,022	22	14,798	-7	-2,224
7	Noyabrsk	36	15,356	33	15,013	-3	-343
8	Tyumen	3	10,581	3	12,987	0	2,406
9	Gubkinskiy	33	11,983	31	8,835	-2	-3,148
10	Apatity	25	10,604	22	9,922	-3	-682
	Total	712	273,581	712	273,581	0	0

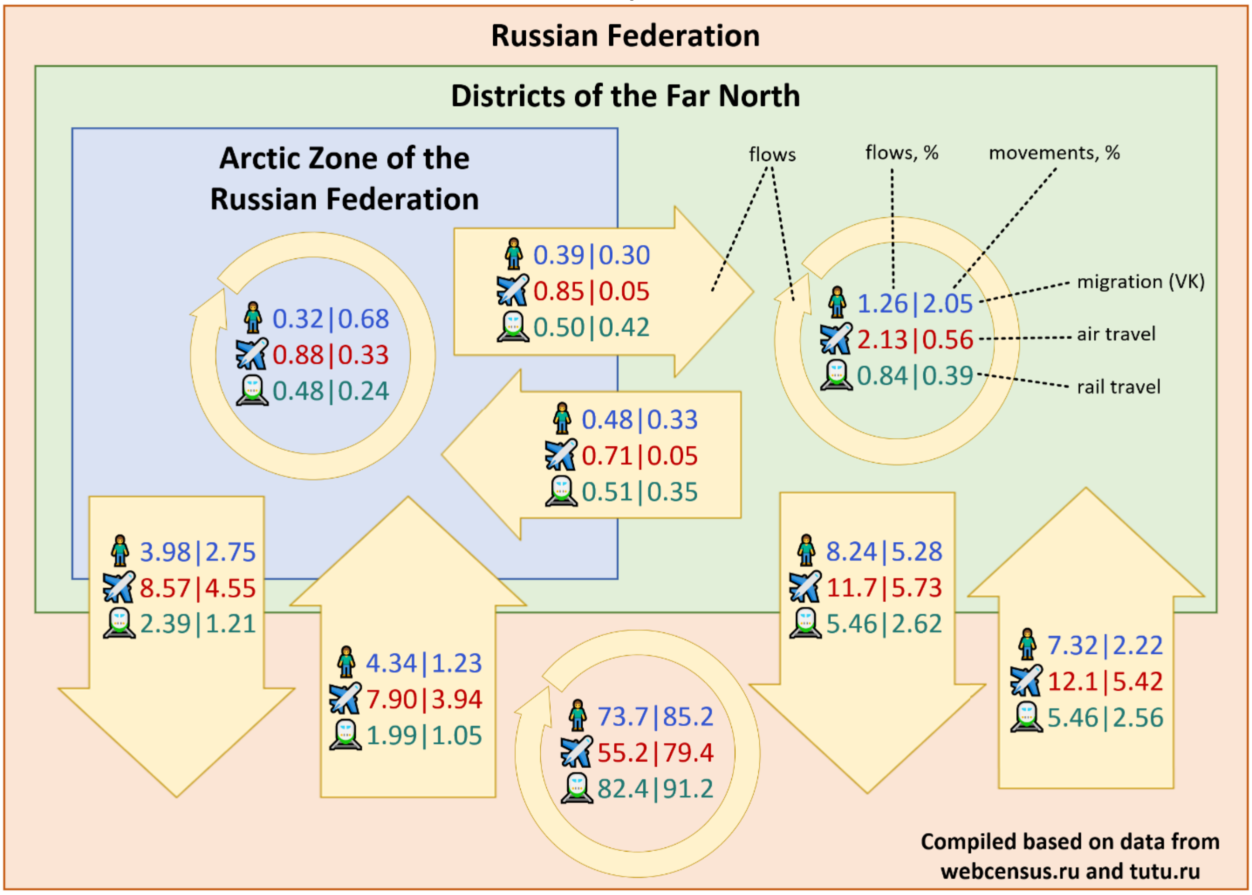
Source: Compiled by the author based on Tutu.ru data as of April 2019.

Note: Ranked in descending order of total moves.

The difference in the number of incoming and outgoing flows from the same nodes is explained by the short analysis time period (1 month), which could lead to underestimation of rare flights; a lack of data on small flows (less than 10 people for rail transport and less than 50 for air); and the presence of compound routes, including more than 1 movement with transfers. If the tickets were bought with different bookings or from different companies, it is very difficult to identify composite routes. This leads to an overestimation of the share of hubs in traffic flows.

The largest flows in air transport are between Murmansk and Moscow, and in rail transport, between Arkhangelsk and Moscow. None of the 20 largest air flows connect 2 Arctic cities. In rail transport, the situation is similar, but there are 5 flows that connect Arctic cities with cities located in the non-Arctic part of the Far North. These are flows from Usinsk to Syktyvkar and back, from Novy Urengoy to Surgut and back, from Noyabrsk to Surgut. Most of the largest flows connect the Arctic municipalities with the capital cities and administrative centers of the regions in the central and southern parts of the country. To consider the connectivity of the Arctic territories in more detail, let's combine all types of the examined migration and transport flows into one scheme (Figure 8).

Figure 8. Distribution of migration and transport flows by directions, % of the total number of flows/movements



Source: Compiled by the author based on data from the Virtual Population of Russia project and the Tutu.ru service.

Although only 1.8% of Russia's population lives in the Arctic, they account for 5.3% of migratory movements, 3.3% of rail movements, and 8.9% of air movements. Consequently, the Arctic population is more mobile, and air transport, which connects remote settlements with federal centers, is of particular importance. There are very few displacements within the Arctic (0.68% of migrations, 0.33% of air and 0.24% of railways), which indicates a low connectivity of the Arctic territories with each other. A much larger share is accounted for by movements between Arctic and non-Arctic settlements (4.6, 3.0 and 8.6% of all movements in the country, respectively). In this regard, the importance should be noted of completing the construction of such large infrastructure projects as the Northern Latitudinal Railway, Belkomur (White Sea – Komi Rep. – Ural railway) and Barentskomur (Barents Sea – Komi – Ural railway). At the same time, to ensure the connectivity of territories in the Arctic, it is necessary to expand the use of regional and local aviation.

Conclusion

The study showed that new data sources that have emerged due to the development of digital technologies make it possible to obtain detailed and timely data on migration processes, while network analysis provides suitable tools for systematizing and comprehending this information. In the territorial context, digital traces can reach the level of settlements or even contain the coordinates of individual places. They reflect different types of migration by duration, direction and reason. Data obtained from social media can contain very detailed socio-demographic characteristics of the population, and search queries and digital texts make it possible to analyze the migration intentions and preferences of residents without large time and material costs. Digital traces make it possible to include millions of Internet users in research, and at the same time to gather information at the micro level, including from remote and hard-to-reach areas. Processing methods, approaches to interpretation and ethical aspects of the use of digital data on the movements of people are still being developed, and further expansion of their explanatory and predictive potential can be expected.

The study of the digital traces of the population of the Russian Arctic made it possible to identify key migration and passenger flows in the macroregion. It has been confirmed that the connectivity of the territories of the Arctic with each other is quite low, and the bulk of the movements concern flows with cities outside the Arctic. It is shown that there are differences in the models of demographic behavior between residents of the Arctic capitals, the administrative centers of the regions, and of other urban districts and municipal districts, as well as between European and Asian territories. The migration and transport hubs of the Russian Arctic have been identified. Moscow and St. Petersburg account for more than a fifth of migration, a third of rail and half of air movements. Moreover, in the federal capitals, incoming migration flows are much larger than outgoing ones. The identification of clusters in migration networks showed a high degree of isolation of territories in the north of Yakutia, and in railway networks, the division of the network into four parts due to the limitations of the existing transport infrastructure. Together with the weak development of the motorway network in most of the Russian Arctic, this hinders the development of horizontal cooperation between residents and organizations in the Arctic Zone.

Since the demographic dynamics in the Arctic are determined primarily by migration flows, taking into account their characteristics at the district and settlement levels will make it possible to build more accurate forecasts of the size and composition of the population. In crisis situations, the speed of generating digital traces is of great importance. Thus, digital sources of data on population migration and morbidity make it possible, without waiting for the publication of official statistics, to develop forecasts for the development of the coronavirus pandemic and make management decisions based on them. The study presents a static picture of patterns of migratory movements, without taking into account the time factor. To conduct such a study in dynamics, the tools of temporal networks can be used. Particular attention in the future study of the problem should be given to seasonal fluctuations in migration, which are significant for remote areas with raw materials. In addition, digital traces can be used to study international and rotational migrations, which are of particular importance for the labor markets of the Arctic in the face of a declining resident population.

References

Dudina V.I. (2021). "Reassembling sociology": the digital turn and the search for new theoretical optics. *Sociological Studies*, 11, 3-11. https://doi.org/10.31857/S013216250016829-4.

Zamyatina N.Yu., Yashunsky A.D. (2018.). Virtual geography of the virtual population. *Monitoring of Public Opinion: Economic and Social Changes*, 1, 117-137. https://doi.org/10.14515/monitoring.2018.1.07.

Kalabikhina I.E., Lukashevpich N.V., Banin E.P., Alibaeva K.V., Reblei S.M. (2021). Automatic extraction of social network users' attitudes on reproductive behavior issues. *Program Systems: Theory and Applications*, 12:4(51), 33-63. https://doi.org/10.25209/2079-3316-2021-12-4-33-63.

Kitchin R. (2021). Data-driven, networked urbanism. In E. Lapina-Kratasyuk, O. Zaporozhets, A. Vozyanov (Eds.), *City Networks: People. Technologies. Authorities* (pp. 58-80). Moscow: New Literary Review.

Smirnov A.V. (2021). Digital society: theoretical model and Russian reality. *Public Opinion Monitoring: Economic and Social Changes,* 1(161), 129-153. https://doi.org/10.14515/monitoring.2021.1.1790.

Smirnov A.V. (2022). Digital Demography Methods for Forecasting Migration Processes. *Economy of the regions*, 18(1), 133-145. https://doi.org/10.17059/ekon.reg.2022-1-10.

Srnicek N. (2020). *Platform capitalism*. Moscow: Publishing House of the Higher School of Economics.

Sudakova A.E. (2020). Migration of scientists: the Digital footprint and scientometry. *Perspectives of Science and Education*, 3(45), 544-557. https://doi.org/10.32744/pse.2020.3.39.

Tarde G. (2016). *Monadology and sociology*. Perm: Giele Press.

Fauzer V.V., Lytkina T.S. (2017). Migration processes in the Russian North. *Social policy and sociology*. 16:1 (120), 141-149. https://doi.org/10.17922/2071-3665-2017-16-1-141-149.

Fauzer V.V., Smirnov A.V. (2020). Migration of the Russian Arctic population : models, routes, results. *The Arctic: Ecology and Economy*, 4(40), 4-18. https://doi.org/10.25283/2223-4594-2020-4-4-18.

Ahmad I., Flanagan R., Staller K. (2020). Increased internet search interest for GI symptoms may predict COVID-19 cases in US hotspots. *Clinical Gastroenterology and Hepatology*, 18(12), 2833-2834. https://doi.org/10.1016/j.cgh.2020.06.058.

Alburez-Gutierrez D., Aref S., Gil-Clavel S., Grow A., Negraia D.V., Zagheni E. (2019). Demography in the Digital Era: New data sources for population research. In: *SIS2019. Smart statistics for smart applications.* Milano: Pearson. https://doi.org/10.31235/osf.io/24jp7.

Billari F.C., D'Amuri F., Marcucci J. (2013). *Forecasting births using Google. Population Association of America Annual Meeting*. https://paa2013.princeton.edu/papers/131393.

Boullier D. (2017). Big data challenges for the social sciences: from society and opinion to replications. In *eSymposium*, 7(2), 1-17. https://www.boullier.bzh/wp-content/uploads/EBul-Boullier-Jul2017.pdf

Cesare N., Lee H., McCormick T., Spiro E., Zagheni E. (2018). Promises and pitfalls of using digital traces for demographic research. *Demography*, 55, 1979-99. https://doi.org/10.1007/s13524-018-0715-2.

Danchev V., Porter M.A. (2018). Neither global nor local: Heterogeneous connectivity in spatial network structures of World migration. *Social Networks*, 53, 4-19. https://doi.org/10.1016/j.socnet.2017.06.003

Danchev V., Porter M.A. (2021). Migration networks: applications of network analysis to macroscale migration patterns. In M. McAuliffe (Ed.), *Research handbook on international migration and digital technology* (pp. 70-90). Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781839100611.

Gansner E., Koren Y., North S. (2004). Graph Drawing by Stress Majorization. *Lecture Notes in Computer Science*, 3383, 239-250. https://doi.org/10.1007/978-3-540-31843-9_25.

Gebru R., Krause J., Wang Y., Chen D., Deng J. Aiden E.L., Fei-Fei L. (2017). Using deep learning and Google street view to estimate the demographic makeup of neighborhoods across the United States. *PNAS*, 114(50), 13108-13113. https://doi.org/10.1073/pnas.1700035114.

Golder S.A., Macy M.W. (2014). Digital footprints: opportunities and challenges for online social research. *Annual Review of Sociology*, 40(1), 129-152. https://doi.org/10.1146/annurev-soc-071913-043145.

Edelmann A., Wolff T., Montagne D., Bail C. (2020). Computational Social Science and Sociology. *Annual Review of Sociology*, 46, 61-81. https://doi.org/10.1146/annurev-soc-121919-054621.

Fruchterman T.M.J., Reingold E.M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129-1164. https://doi.org/10.1002/spe.4380211102.

Heleniak T., Bogoyavlenskiy D. (2014). Arctic Populations and Migration. In Larsen J.N., Fondahl G., Rasmussen H. (Eds.), *Arctic Human Development Report. Regional Processes and Global Linkages* (pp. 53-104). Copenhagen: Nordic Council of Ministers. https://doi.org/10.6027/TN2014-567.

Hughes C., Zagheni E., Abel G., Wi´sniowski A., Sorichetta A., Weber I., Tatem A.J. (2016). *Inferring migrations: Traditional methods and new approaches based on mobile phone, social media, and other big data*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2767/61617.

Igntatow G. (2016). Theoretical foundations for digital text analysis. *Journal for the Theory of Social Behaviour*, 46(1), 104-120. https://doi.org/10.1111/jtsb.12086.

Katzenbach C., Bächle T.C. (2019). Defining concepts of the digital society. *Internet Policy Review*, 8(4). https://doi.org/10.14763/2019.4.1430.

Kitchin R. (2014). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, 1(1), 1-12. https://doi.org/10.1177/2053951714528481.

Lazer D., Radford J. (2017). Data ex Machina: introduction to big data. *Annual Review of Sociology*, 43(1), 19-39. https://doi.org/10.1146/annurev-soc-060116-053457.

Ledford H. (2020). How Facebook, Twitter and other data troves are revolutionizing social science. *Nature*, 582, 328-330. https://doi.org/10.1038/d41586-020-01747-1.

Maier G., Vyborny M. (2008). Internal migration between US States: A social network analysis. In J. Poot, B. Waldorf, L.W. Wissen (Eds.), *Migration and Human Capital*. Cheltenham, UK: Edward Elgar Publishing. URL: https://www.econstor.eu/handle/10419/117573.

McCormick T.H., Lee H., Cesare N., Shojaie A., Spiro E.S. (2017). Using Twitter for demographic and social science research: tools for data collection and processing. *Sociological Methods & Research*, 46(3), 390-421. https://doi.org/10.1177/0049124115605339.

Petrov A.N., Welford M., Golosov N., DeGroote J., Degai T., Savelyev A. (2021). The “second wave” of the COVID-19 pandemic in the Arctic: regional and temporal dynamics. *International Journal of Circumpolar Health*, 80(1). https://doi.org/10.1080/22423982.2021.19254461.

Raghavan U.N., Albert R., Kumara S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3). https://doi.org/10.1103/physreve.76.036106.

Taylor L., Floridi L., van der Sloot L. (Eds.). (2017). *Group privacy: New challenges of data technologies.* Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-46608-8.

Zagheni E., Weber I., Gummadi K. (2017). Leveraging Facebook’s advertising platform to monitor stocks of migrants. *Population and Development Review*, 43(6178), 721-734. https://doi.org/10.1111/padr.12102.

Zamyatina N., Yashunsky A. (2017). Migration cycles, social capital and networks. A new way to look at Arctic mobility. In M. Laruelle (Ed.), *New Mobilities and Social Changes in Russia’s Arctic Regions* (pp. 59-84). London and New York, Routledge.

Демографическое обозрение / Demographic Review
2022;9(2) 81-101 DOI: https://doi.org/10.17323/demreview.v9i2.16207

Excess summer mortality at young ages: do long school holidays matter?

Polina O. Kuznetsova
(polina.kuznetsova29@gmail.com), Institute for Social Analysis and Forecasting at RANEPA, Russia.
Tatyana M. Maleva
(maleva-tm@ranepa.ru), Institute for Social Analysis and Forecasting at RANEPA, Russia.

Abstract: *In this paper we analyze the seasonality of total mortality at younger ages in Russia compared with other countries. To decompose the time series into a trend, seasonality, and a random component and to perform seasonality tests, the TRAMO SEATS procedure was applied to short-term mortality data from the Short-term Mortality Fluctuations Human Mortality Database in 2000-2020. Summer seasonality was estimated as the ratio of average mortality in summer to average mortality per year after the trend component was excluded from the time series.*

In Russia, there exists a significant summer seasonality of child and adolescent mortality: at the age of 5-9 years, mortality in the summer months exceeds the average mortality per year by 21%; at the age of 10-14 years old, by 32%; and at the age of 15-19 years old, by 19%. In the vast majority of countries included in the analysis, the seasonality of child mortality in general and the summer seasonality in particular are noticeably lower than in Russia, especially in the age groups of 5-9 and 10-14 years. It is only after the age of 15 that excess summer mortality at young ages becomes relatively common in other countries.

The pronounced seasonality of mortality in younger age groups may, among other things, be the result of long school holidays, during which many children find themselves without proper supervision and accompaniment. To test this hypothesis, we compared the length of summer holidays in the STMF countries with child and adolescent seasonality. We show that there is a significant positive relationship between these indicators: longer holidays correspond to more pronounced excess summer mortality, and this result remains stable for different age groups.

Keywords: *mortality, seasonality, adolescent mortality, TRAMO SEATS, Human Mortality Database, Russia.*

Funding: *The article was prepared as part of the research work of the state assignment of the RANEPA.*

Acknowledgments: *The author is grateful to two anonymous referees and Marina Kartseva for valuable comments that allowed us to significantly improve this text.*

For citation: *Kuznetsova P., & Maleva T. (2022). Excess summer mortality at young ages: do long school holidays matter? Demographic Review, 9(2), 81-101. https://doi.org/10.17323/demreview.v9i2.16207*

Introduction

The mortality of the Russian population has a pronounced seasonality: deaths occur more often in winter, at the height of seasonal respiratory diseases. However, for certain age groups, this rule is not followed, and mortality in summer is noticeably higher. This can be seen in the weekly mortality data of the Short-term Mortality Fluctuations Human Mortality Database (hereinafter STMF, see (Human Mortality Database 2022)). The seasonality coefficients, calculated as the ratio of the average monthly number of deaths in June-August and the average monthly number of deaths in 2019, reach a maximum at the age of 5-14 years and then decrease to below one after the age of 45, which corresponds to the transition to the winter seasonality of mortality (Table 1).

Table 1. Seasonal mortality coefficients by age, Russia, men and women, 2000-2020

Year	Men						Women					
	2000	2005	2009	2015	2019	2020	2000	2005	2009	2015	2019	2020
0-4	0.99	1.02	0.99	0.99	1.00	1.04	1.02	1.01	0.97	0.97	1.03	1.02
5-9	1.52	1.39	1.32	1.30	**1.21**	**1.28**	1.36	1.36	1.14	1.02	**1.13**	**1.26**
10-14	1.38	1.43	1.34	1.34	**1.30**	**1.37**	1.40	1.33	1.28	1.10	**1.18**	**1.16**
15-19	1.24	1.28	1.20	1.25	**1.19**	**1.25**	1.22	1.24	1.10	1.03	1.06	**1.18**
20-24	1.19	1.20	1.17	1.14	1.09	1.16	1.15	1.09	1.02	1.06	1.03	1.08
25-34	1.10	1.07	1.06	1.04	1.05	1.06	1.10	1.05	0.96	0.99	1.01	1.03
35-44	1.01	1.00	0.99	1.00	1.02	1.01	0.99	0.95	0.94	0.96	1.00	1.00
45-54	0.97	0.96	0.95	0.96	0.97	0.97	0.92	0.94	0.91	0.93	0.96	0.94
55-64	0.94	0.96	0.96	0.95	0.96	0.93	0.91	0.91	0.92	0.94	0.95	0.91
65-74	0.92	0.95	0.93	0.94	0.97	0.91	0.91	0.92	0.92	0.93	0.96	0.91
75-85	0.91	0.94	0.92	0.92	0.96	0.90	0.90	0.92	0.92	0.93	0.95	0.93
85 years and older	0.88	0.93	0.92	0.93	0.93	0.89	0.88	0.91	0.91	0.91	0.94	0.93
Whole population	0.96	0.97	0.96	0.95	**0.97**	**0.93**	0.91	0.93	0.92	0.93	**0.95**	**0.93**

Source: Authors' calculations based on STMF data (Human Mortality Database 2022).

Note: Instead of 2010, which was characterized by extreme weather conditions in the European part of the country, 2009 was used. Data for 2019 were added to the pandemic 2020 data.

In the 2010s, the improvement in mortality in the general population and in adolescents in particular contributed to a slight decrease in the seasonality of mortality in 5-year age groups from 5 to 19 years; however, its absolute values, especially for males, still remain high (see highlighted values in table 1). In our opinion, this may be one of the consequences of children being unsupervised during long school holidays, as a result of which boys, often more active and more prone to risky behavior (Weden, Zabin 2005; Chun, Mobley 2010; Knyazev et al. 2004; Boriskin et al. 2018), can suffer more. It should be especially noted that a significant excess of the number of summer deaths over winter ones occurs against the background of a significant increase in mortality during the New Year holidays, which is concentrated, however, at older ages (Nemtsov, Fattakhov 2021).

When making calculations, it is necessary to use proper estimates of seasonality. In Table 1, we present estimates of seasonality based on a comparison of data from the initial time series of mortality, without taking into account long-term changes in its dynamics. In modern scientific research, econometric methods are used to extract individual components (trend, seasonality, random fluctuations) from the original time series, which makes it possible to more correctly assess seasonality and conduct intertemporal and intergroup comparisons (UNECE 2020).

Since we are interested in the relationship between the summer seasonality of mortality and the duration of summer holidays, this study will mainly study the age groups that are most related to school age: 5-9, 10-14 and 15-19 years old. Due to the relatively small contribution of these age groups to overall mortality, as well as to the lack of open data at the required level of detail, seasonality at younger ages in Russian data, as far as we know, has not yet been considered in detail. In this regard, of great interest is the STMF (Human Mortality Database 2022) database, which contains information on weekly mortality among various sex and age groups in several countries of the world, including Russia.

The purpose of this study is to analyze the seasonality of mortality at younger ages in Russia and, if possible, to test the hypothesis about the effect of the duration of school holidays on the range of intra-annual fluctuations. The article includes a brief review of recent work on seasonality in mortality, a description of the data and methodology of the empirical part of our study, a presentation of the results and a brief discussion. The main conclusions of the work are presented in the final section.

Seasonality of Mortality in Children and Adolescents: A Brief Review of Current Research

The seasonality of mortality is a well-known and well-studied phenomenon. Currently, in countries with a temperate climate, for the population as a whole it is usually expressed in a winter peak and a summer minimum of mortality. For example, in a number of developed countries with different climates, annual mortality minima between the ages of 2 and 57 are most often reached in late summer or early autumn, due to favorable weather conditions, as well as to the positive effect of summer holidays (Falagas et al. 2009). (Healy 2003), which compared the seasonality of mortality in 14 countries, observed the highest winter seasonality in Portugal (21% higher).

In Russia, the situation is similar: during the epidemiologically more or less safe pre-COVID period, the number of deaths in January, March and, to a lesser extent, December, consistently exceeded the average annual level, and in July-October was usually below it (Shcherbakova 2010). Researchers also note an increase in mortality during heat waves, including in Russia in 2010, when a long period of hot weather and atmospheric pollution caused by forest fires in Moscow alone caused 11,000 excess deaths (Shaposhnikov et al. 2014). Another important factor in the seasonality of mortality in Russia is the significant excess mortality during the New Year holidays (Nemtsov, Fattakhov 2021).

Information concerning the seasonality of mortality in children and adolescents (with the exception of infant mortality) in the scientific literature is noticeably more fragmented. In the United States, there is a pronounced summer seasonality in adolescent and young adult mortality, driven primarily by external cause mortality (Parks et al. 2018). In particular, the peak of mortality at the age of 5–14 years from all causes occurs in the summer and coincides in time with the peak of mortality from external causes.

In Japan (Shinsugi 2015) in 2000-2010 the main cause of death in children and adolescents was external causes, some of which have a pronounced seasonality. Thus, the peak of deaths from transport accidents and drownings occurs in the summer, suicides in the spring, and accidents with respiratory risks in the winter.

In the third most populous city in China (Guangzhou), in 2008-2018 the mortality of children aged 0-14 years from external causes of death increased significantly during the summer holidays and especially in August. Of the three most common causes of death from injuries (transport accidents, accidents with a threat to breathing, and drowning), the most pronounced seasonality of mortality is that from drowning, the local maximum of which (more than 35% of all deaths per year) occurs in July-August (Xu et al. 2020).

In Hungary the annual maximum mortality of children aged 0-14 years from transport accidents and drownings is observed in July (Lantos, Nyári, McNally 2019).

There is a small number of works that analyze the effect of vacations and holidays on mortality in younger ages. The analysis carried out on the data for 2012-2015 in Fiji (Murray, Carter 2017) found that deaths from accidental drowning between the ages of 1 and 29 were significantly higher in months with a high number of holidays and weekends. In a study on mortality from unintentional drowning among 5–17-year-olds, carried out on Australian data (Peden, Barnsley, Queiroga 2019), it was shown that the relative risk of drowning on a holiday is 2.4 times higher than on a school day, which probably says something about the protective effect of formal schooling. The authors of a study on the seasonality of mortality among children aged 0-14 years in the Chinese city of Guangzhou also emphasize that there is a higher risk of drowning during school holidays (Xu et al. 2020).

There are relatively few studies on the seasonality of mortality in Russia. They are mainly devoted to excess winter mortality of the population as a whole, the level of which is noticeably lower than in many European countries (McKee et al. 1998; Kontsevaya et al. 2014), as well as to the impact of extreme weather conditions on mortality (Revich, Shaposhnikov 2010; Shaposhnikov et al. 2014). (McKee et al. 1998) also noted significant excess summer mortality in young adults, especially from alcohol-related causes.

In a study of the mortality of adolescents aged 15-17 in the Omsk region in 2000-2012, (Zakorkina 2015) demonstrated a significant seasonal component in the number of deaths from external causes. For the three causes that make the greatest contribution to traumatic mortality, the share of seasonal fluctuations was: for suicide - 7% (local maxima both in May, August and September, and in December), for transport accidents - 35% (seasonal rise in May -September), for events of undetermined intent - 30% (seasonal rise in June-September with a local maximum in August). Among the reasons for the pronounced summer seasonality of mortality from a number of external causes, the author names the absence of classes in educational institutions and adolescents' relative lack of activities.

The seasonality of demographic processes in Russia is studied in (Rodionova, Kopnova 2019). The authors apply seasonal ARIMA models to monthly Rosstat data on the number of births, deaths, and marriages. The application of this method makes it possible to obtain models with good predictive and statistical properties. In particular, it is shown that the time series of the number of deaths contains a deterministic seasonality, while there is no seasonality for infant mortality.

Of interest to our study are also works concerning the dynamics and structure of mortality in children and adolescents from external causes, since it is mortality from a number of external causes that has a pronounced summer seasonality. A report by the World Health Organization

(Sethi et al. 2017) compares the dynamics of traumatic mortality among children aged 0-14 years in the countries of the European region. It is shown that from 2000 to 2015 the region had increased inequality between high-income countries and low- and middle-income countries in the level of deaths from unintentional injuries (accidents, poisoning, falls, accidents caused by fire and heat, drowning). In part, this increase in inequality could be due to the persistence of the unfavorable situation in Russia (a large middle-income country in the region), where mortality from unintentional injuries during the study period decreased more slowly than in developed countries.

The work (Ivanova et al. 2009) presents an analysis of the mortality of the Russian population aged 15-19 years. The positive dynamics observed in the mid-2000s largely affected endogenous causes of death (neoplasms and diseases of the nervous system). At the same time, there was an increase in mortality from cardiovascular diseases, which, according to the authors, may partly mask mortality from drug poisoning. An important change in post-Soviet teenage mortality statistics has been an increase in mortality from events of undetermined intent, which in turn may include a proportion of deaths from murder and drug poisoning. Another possible reservoir of masked mortality from external causes is the group of ill-defined conditions (see also (Semenova et al. 2018; Semenova et al. 2021) about this).

Data and methodology

The calculations used weekly data on mortality in 2000-2020 from the Short-term Mortality Fluctuations (STMF) Human Mortality Database. Weekly data on the number of deaths were converted into monthly data, which were then normalized according to the length of the months, i.e., recalculated for 30 days.

To decompose the time series of the number of deaths by age groups, we used the parametric TRAMO-SEATS procedure developed by the Central Bank of Spain (Caporello, Maravall 2004; Maravall, López-Pavón, Pérez-Cañete 2015). We used its implementation in the JDemetra software (Grudkowska 2015; UNECE 2020). The procedure uses a seasonal autoregressive integrated moving average model (SARIMA) and makes it possible to select 3 components from the initial time series X_t: trend-cyclical T_t, which includes both a long-term trend and smooth cyclical fluctuations around it, seasonal S_t and irregular I_t. We used the multiplicative decomposition $Xt = T_t * S_t * I_t$, which assumes that the time series does not take negative values and that seasonal fluctuations increase with the growth of the indicator.

The TRAMO-SEATS method is one of the seasonal adjustment procedures for data. JDemetra also implements the X13 semiparametric method developed by the US Census Bureau. Each method has its advantages and disadvantages; both are actively used in scientific expertise and researchers have no clear preferences in favor of one or the other. In practice, they often give similar results (UNECE 2020).

In the implementation of the TRAMO-SEATS procedure in the JDemetra statistical package, the presence of seasonality is determined based on 6 tests. These include seasonal lag autocorrelation, the Friedman test, the Kruskal-Wallis test, spectral peaks, a periodogram and seasonal dummies (UNECE 2020: 56). A decision on the need for seasonal adjustment of the original time series is made if the majority of tests confirm the existence of seasonality (UNECE 2020: 57, table 4.2). A detailed description of the tests is given in (Grudkowska 2015: 106-110); for examples of calculations see (O'Keeffe 2017; Asif et al. 2019).

We defined the coefficient of summer seasonality of mortality as the ratio of the average monthly number of deaths in June-August 2015-2019 to the average monthly number of deaths over the entire period.

When conducting international comparisons, data on mortality in countries with small populations were also used. Of the 26 countries included in the analysis of the seasonality of mortality by 5-year age groups, in 5 (Croatia, Slovenia, Latvia, Lithuania, Estonia) the population does not exceed 5 million people, and in 3 more countries (Denmark, Finland, Slovakia) does not exceed 6 million. To increase the absolute values of the mortality rate at younger ages, we tried to enlarge the groups studied by combining men and women, comparing data for a three-month period (summer), and also calculating the seasonality indicator for the total values of the indicator for 5 years from 2015 to 2019.

When conducting a comparative analysis, we used data on the length of school summer holidays in different countries, collected from open sources. The main source was a site[1] which provides data on the length of summer holidays for younger students in European countries. We collected the missing information for each country separately, using data from the world-schools website[2], as well as Wikipedia. In some countries, vacation length varies by region. In such cases, we tried to take the average of the observed values.

Table 2. Duration of summer holidays in the countries included in the comparative analysis (for countries located in the Southern Hemisphere, the duration of winter holidays)

Name	Duration of summer holidays, weeks	Name	Duration of summer holidays, weeks	Name	Duration of summer holidays, weeks
Australia	5	France	8	Norway	8
Austria	9	Germany	6	Poland	13
Belgium	9	Greece	12	Portugal	11
Bulgaria	13.5	Hungary	11	Russia	14
Canada	9	Israel	10	Scotland	6.5
Chile	11	Italy	14	Slovakia	9
Croatia	10	Korea	8	Slovenia	10
Czech Republic	9	Latvia	13	Spain	13
Denmark	6	Lithuania	13	Sweden	6
England	6.5	Netherlands	6	Switzerland	6
Estonia	12	New Zealand	6	Taiwan	9
Finland	10.5	Northern Ireland	6.5	USA	10.5

Source: Compiled by the authors based on information from websites: https://jakubmarian.com/school-holidays-by-country-in-europe-map/, https://world-schools.com; https://www.aucklandforkids.co.nz/new-zealand-school-and-public-holiday-dates; https://holidayswithkids.com.au/school-holidays-and-australian-public-holidays

As can be seen from the data in Table 2, the countries included in the analysis differ significantly in terms of vacation duration. The shortest summer school holidays (no more than 7 weeks) are found, as a rule, in countries where the predominant religion is Protestantism: in Great Britain (England, Northern Ireland, Scotland), former English colonies (Australia and New

[1] https://jakubmarian.com/school-holidays-by-country-in-europe-map/
[2] https://world-schools.com

Zealand), countries of Northern Europe (Denmark, Sweden) and several countries of Western Europe (Germany, the Netherlands, Switzerland).

On the contrary, the longest vacations (12 weeks or more) are found in the countries of the former USSR (Russia, the Baltic countries), Eastern Europe (Bulgaria, Poland) and countries of Southern Europe (Greece, Spain, Italy). The longest holidays (14 weeks) in the sample of countries considered are in Russia and Italy.

Results and discussion

Seasonality of mortality in Russia

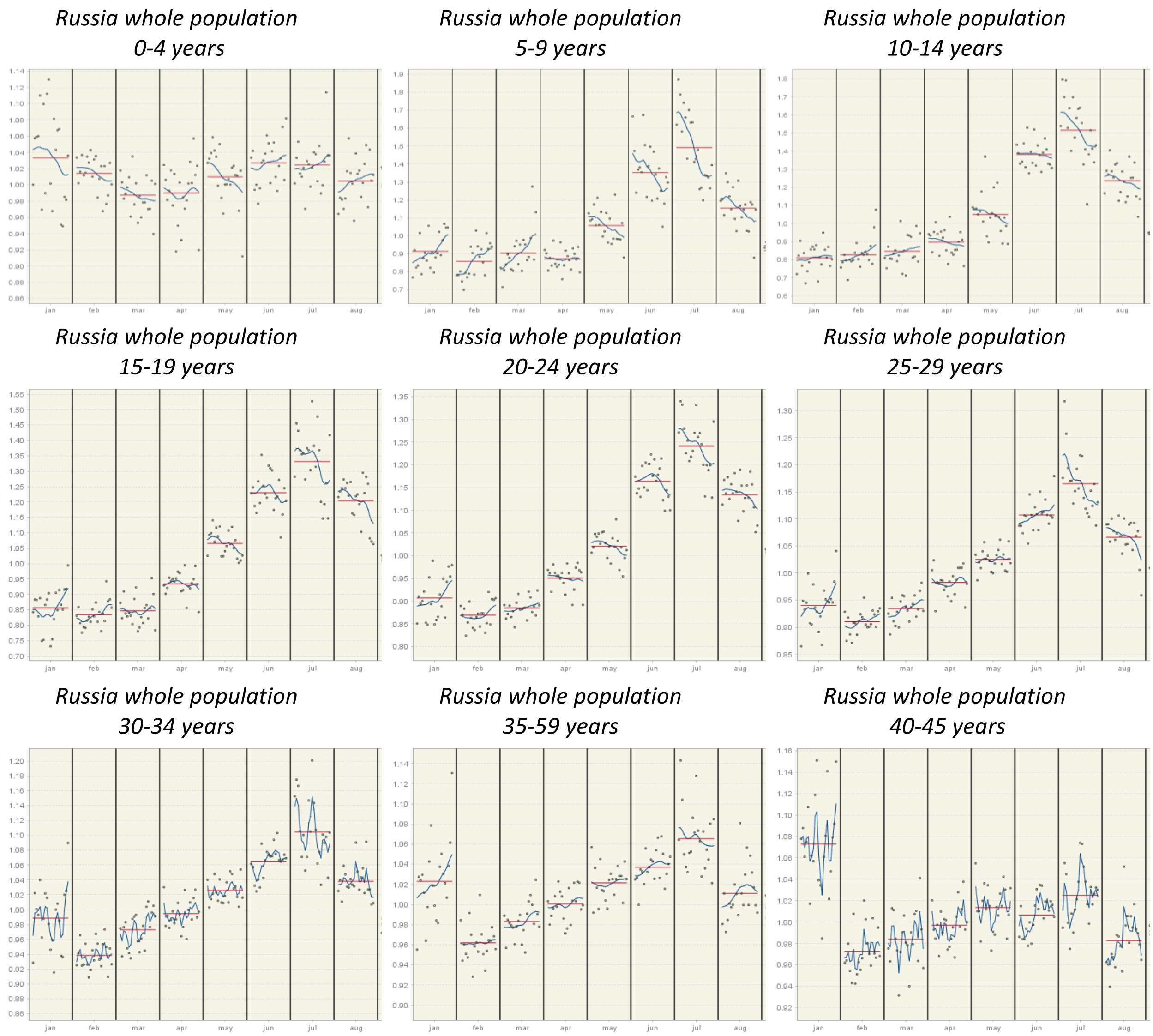
To assess the seasonality to the baseline mortality data in 5-year age and sex groups, the iterative procedure TRAMO SEATS (RSA3) was applied. With its help, the initial series was decomposed into a trend, seasonal and random (irregular) components. Seasonally adjusted was defined as the ratio of the observed mortality value to the trend value; in the literature, it is called the seasonal-irregular (S-I) component. Figure 1 shows the seasonally adjusted data for the population aged 0-44 years, calculated on the 2015-2019 data obtained for the original series from 2000 to 2020.

The data are presented by months, with individual points corresponding to the values of the seasonally irregular component in 2000–2020, the red line to its average value, and the blue line to the trend during the observation period.

Figure 1 clearly shows how the summer seasonality increases at younger ages: first appearing in the group up to 5 years (along with the summer local peak, there is also a winter peak observed in January-February), it then increases sharply at the age of 5-9 years and reaches a maximum for 10-14-year-olds. The highest mortality values are in July. Further, the summer peaks of mortality decrease relative to the average values and the factor of increased January mortality and overall mortality of the winter and cold season as a whole begins to manifest itself more and more.

The data in Figure 1 also shows that in the 2010s the summer seasonality of mortality in the age groups of 5–9 and 10–14 years decreased, as evidenced by a declining trend (blue lines) for the summer months and, conversely, an increasing trend for the winter months. We assume that this decrease was due to the favorable dynamics of mortality from external causes, including those for which annual maxima occur during the warm season.

Figure 1. Seasonally irregular component of mortality in age groups from 0 to 44 years, males and females, 2000-2020



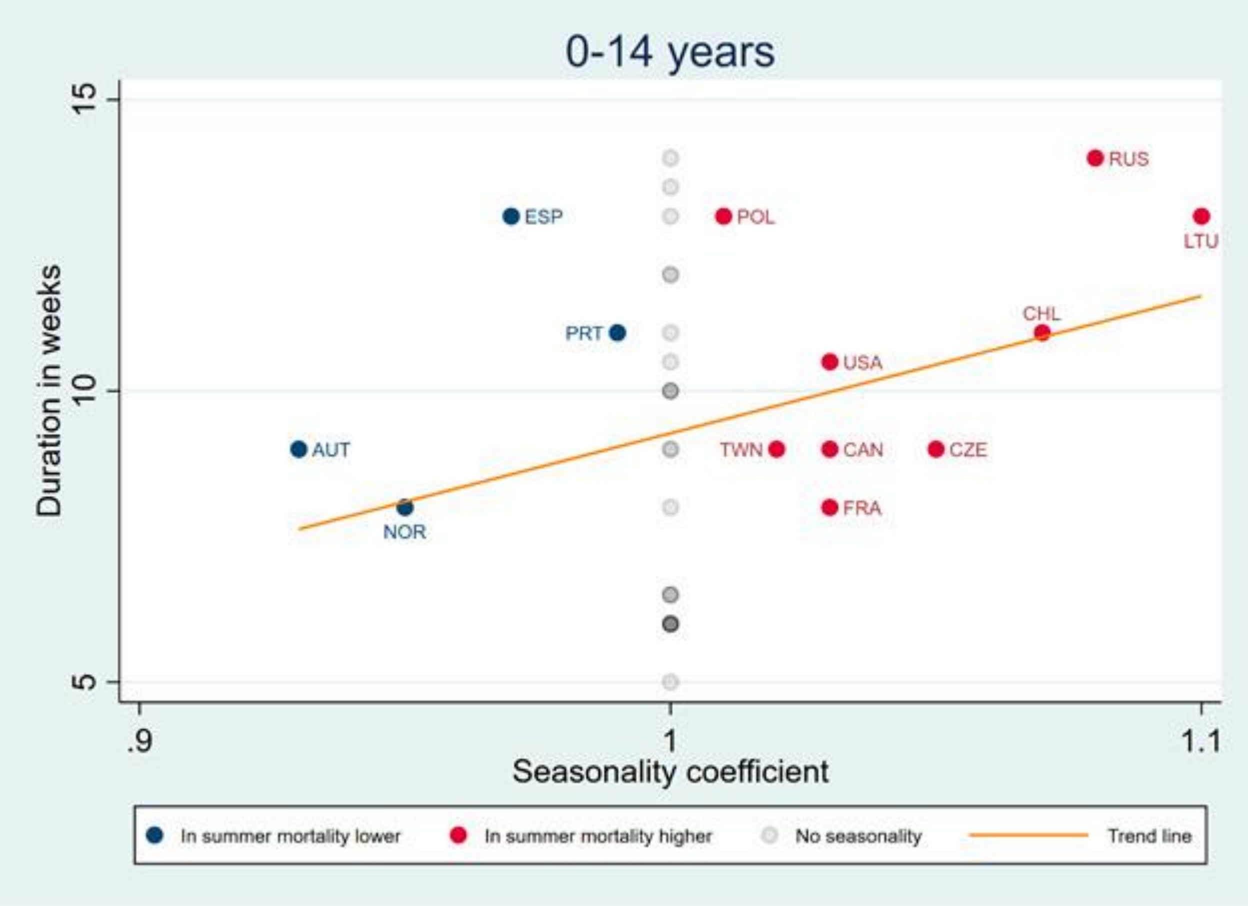
Source: Authors' calculations based on STMF (Human Mortality Database 2022) data using JDemetra software.

Seasonality of mortality between the ages of 0-14 years in 38 STMF countries

In order to compare the seasonality of mortality in Russia with other countries, we first reviewed STMF data for 36 countries on mortality by months at age 0-14 years (Iceland and Luxembourg were excluded from the analysis due to their small populations). We supplemented this data with information on the duration of summer vacations collected from open sources.

Figure 2 shows the relationship between the length of summer holidays and the summer seasonality factor for the population aged 0-14 years. Values for countries with summer seasonality are in red, with winter seasonality in blue and no seasonality in gray (darker shading corresponds to values from multiple observations).

Figure 2. Coefficient of mortality seasonality at the age of 0-14 years (mortality in the summer months in relation to the average annual mortality), males and females, 2015-2019



Source: Authors' calculations based on STMF data (Human Mortality Database 2022).

Note: AUT - Austria, CAN - Canada, CHL - Chile, CZE - Czech Republic, ESP - Spain, FRA - France, LTU - Lithuania, NOR - Norway, POL - Poland, PRT - Portugal, RUS - Russia, TWN - Taiwan, USA - USA.

From the data presented, it can be seen that summer maxima in mortality are more common in countries with longer summer holidays. Apart from Russia, summer seasonality in mortality for those aged 0-14 years in 2015-2019 is also observed in Lithuania, as well as, though less pronounced, in Chile, the Czech Republic, the USA, Canada, France, Poland and Taiwan. A lack of seasonality is more common in countries with relatively shorter holidays (darker gray dots at the bottom of Figure 2, corresponding to larger groups of countries with undetectable mortality seasonality).

In order to take into account possible gender differences, we also considered the ratio of the seasonality of mortality and the duration of holidays separately for men and women (Figure A1 of the Appendix). The main difference was that for males the seasonality confirmed by tests is always summer, while for female mortality this is not necessarily the case. For example, in Austria, Norway, Spain and Portugal, mortality in summer is less than the average during the year. We explain this by the higher contribution of external causes of death to male mortality. Diseases rather than injuries play a greater role in female mortality, and therefore the effect of adverse weather and epidemiological conditions characteristic of the cold season is stronger.

However, it should be noted that the seasonality of mortality in the extended age group from 0 to 14 years is a composite of the seasonality of infant, child and adolescent mortality, which differ significantly in structure. In mortality at the age of 0-4 years, over 90% are endogenous (not external) causes, and in groups of 5-9 and 10-14 years, their share falls below 60% (Figure 3). In this regard, the seasonality of schoolchildren's mortality may differ significantly from the seasonality of mortality in the aggregated group from 0 to 14 years. Thus, the result obtained - the presence of a positive relationship between the length of school holidays and the pronounced excess summer mortality of children - needs to be examined.

Seasonality of mortality in 5-year age groups for 28 STMF countries

In a more detailed analysis of the seasonality of mortality at younger ages, a reduced sample of 28 countries was used for which data on mortality in 5-year age groups at younger ages are available. Large countries such as the United States (STMF contains only raw data on mortality of those aged 0-24 years) and Germany (only data on mortality for ages 0-29 years) did not meet this criterion. The list of countries selected for comparison includes France, Italy, Spain, Poland, Taiwan, Chile, the Netherlands, Belgium, Greece, Czech Republic, Portugal, Sweden, Hungary, Austria, Switzerland, Bulgaria, Croatia, Denmark, Estonia, Finland, Latvia, Lithuania, Norway, Slovenia and Slovakia, as well as parts of the UK (England and Wales; Scotland and Northern Ireland). Due to their small populations, Iceland and Luxembourg were not included in the analysis.

Table 3 provides information on the seasonality of mortality in four 5-year age groups in the countries of this list (in cases where seasonality was confirmed by statistical tests). Confirmed seasonality of mortality at the age of 10-14 years is observed in 5 countries - Lithuania, Poland, Russia, France and Croatia. In Russia, it is the highest - mortality in summer exceeds the annual average by 32% (in Lithuania by 24%, in Croatia by 15%, in Poland and France by 5% or less). It should be noted that Croatia and Lithuania are countries with small populations and the results obtained should be treated with caution.

At the age of 15-19 years, seasonality is more common: in 11 out of the 26 cases examined. In Russia, summer seasonality at this age is also higher than in other countries - 1.19 (1.18 in Bulgaria, 1.15 in Croatia, 1.12 in Hungary, 1.09 in Sweden, 1.08 in Poland and France, 1.05 in Lithuania).

Why is the mortality of children and adolescents in Russia higher in better weather? Unlike older people, they die more often from a number of external causes, for which the peak values of mortality occur in the warm season: from suicide in May (less often June or July), from transport injuries in August-September (Shcherbakova 2010), and from events of undetermined intent in June-September (Zakorkina 2015: results for the Omsk region).

In 2019, the proportion of suicides, transport accidents and deaths with undetermined intent in the mortality of males for those aged 10-14 years was 5%, 11% and 15%, respectively, and 14%, 17% and 21% for those aged 15-19 years. Thus, the total share of these three causes with presumably summer seasonality in male mortality was 31% for those aged 10-14 and 52% for those aged 15-19 years. For female mortality, the share of these three causes is also high: 26% at the age of 10-14 and 48% at the age of 15-19. In addition, common causes of accidental death (W00-X49) at these ages are "accidental drowning and submersion while in a natural body of water" and "falling from one level to another" (correspondingly 5% and 2% of all male deaths for those aged 10-14 years, as well as 3% and 2% of all male deaths for those aged 15-19), both of which may also have a summer seasonality.

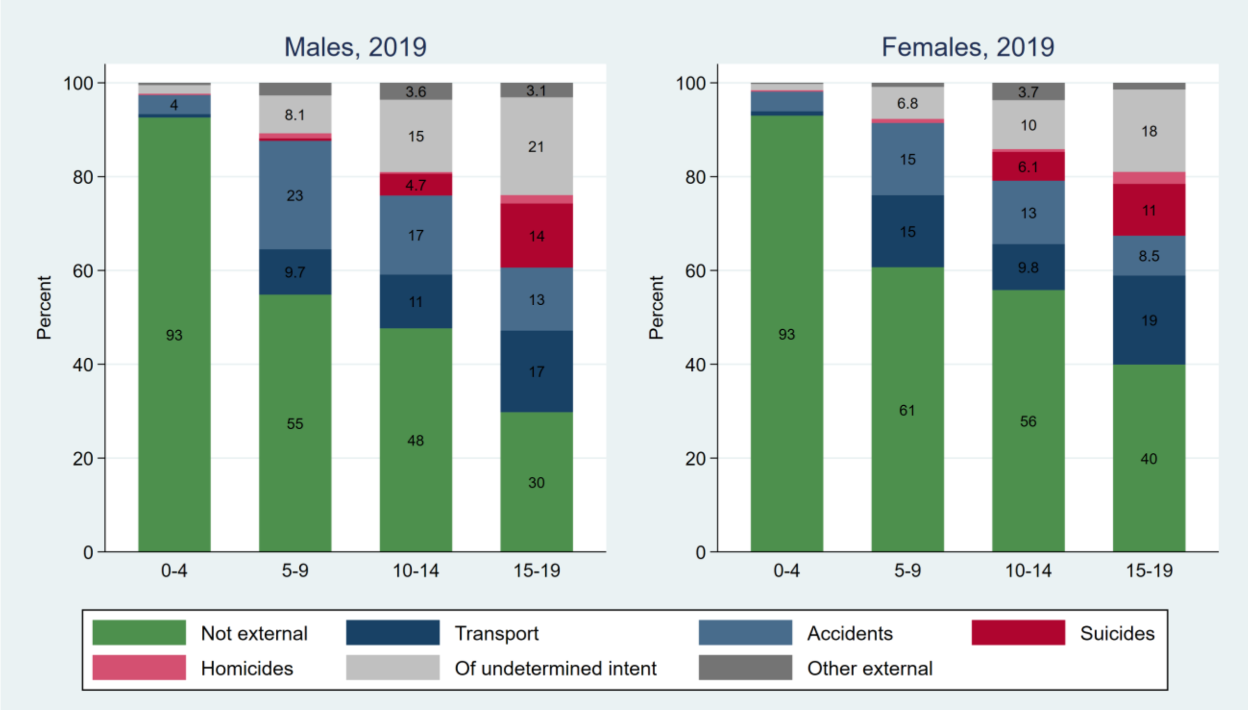
Table 3. Information on the seasonality of mortality at younger ages, countries from the STMF database, 2015-2019

Country	Age	Seasonality coefficient (average mortality in June-August 2015-2019/average mortality in 2015-2019)	Month with minimum mortality	Month with maximum mortality
Bulgaria	15-19	1.18	11	8
Croatia	5-9	0.85	10	1
Croatia	10-14	1.15	9	1
Croatia	15-19	1.15	10	8
Denmark	15-19	0.81	9	1
Estonia	0-4	1.06	9	8
France	5-9	0.99	6	4
France	10-14	1.01	11	2
France	15-19	1.08	4	8
Hungary	5-9	0.92	8	3
Hungary	15-19	1.12	2	6
Lithuania	10-14	1.24	3	8
Lithuania	15-19	1.05	9	2
Poland	5-9	1.05	9	10
Poland	10-14	1.05	1	8
Poland	15-19	1.08	4	8
Portugal	0-4	0.99	10	1
Portugal	15-19	1.16	4	7
Russia	0-4	1.03	11	7
Russia	5-9	1.21	11	7
Russia	10-14	1.32	12	6
Russia	15-19	1.19	2	6
Spain	0-4	0.97	6	12
Spain	5-9	0.98	11	2
Spain	15-19	1.02	10	1
Sweden	15-19	1.09	12	11
Taiwan	0-4	1.02	11	2
Taiwan	5-9	1.02	3	2

Source: Authors' calculations based on STMF data (Human Mortality Database 2022).

As can be seen in Figure 3, the structure of mortality among children and adolescents has noticeable gender differences. For female mortality, the contribution of external causes in general and unintentional injuries in particular is lower. At the same time, regardless of gender, there is a very high proportion of events of undetermined intent: 21% of all deaths for those aged 15-19 years for men and 18% for women.

Figure 3. The contribution of various causes to the mortality of children and adolescents in Russia, males and females, 2019



Source: Authors' calculations based on RosBRiS data (CDI NES 2022)[3].

Figure A2 of the Appendix also displays information on changes in the structure of child mortality over the past 20 years. During this period, there was a decrease in the contribution of external causes to the mortality of children and adolescents, especially noticeable for younger groups. For example, the proportion of deaths from external causes for boys aged 5-9 decreased from 61% in 2000 to 45% in 2019. In addition to this positive trend, one can also note the deterioration in the quality of mortality statistics due to a significant increase in the share of events of undetermined intent. In particular, for boys aged 15-19 years, the proportion of deaths from events of undetermined intent increased from 9% in 2000 to 21% in 2019, which, according to experts, may mask some of the deaths from socially significant causes (homicides and suicides).

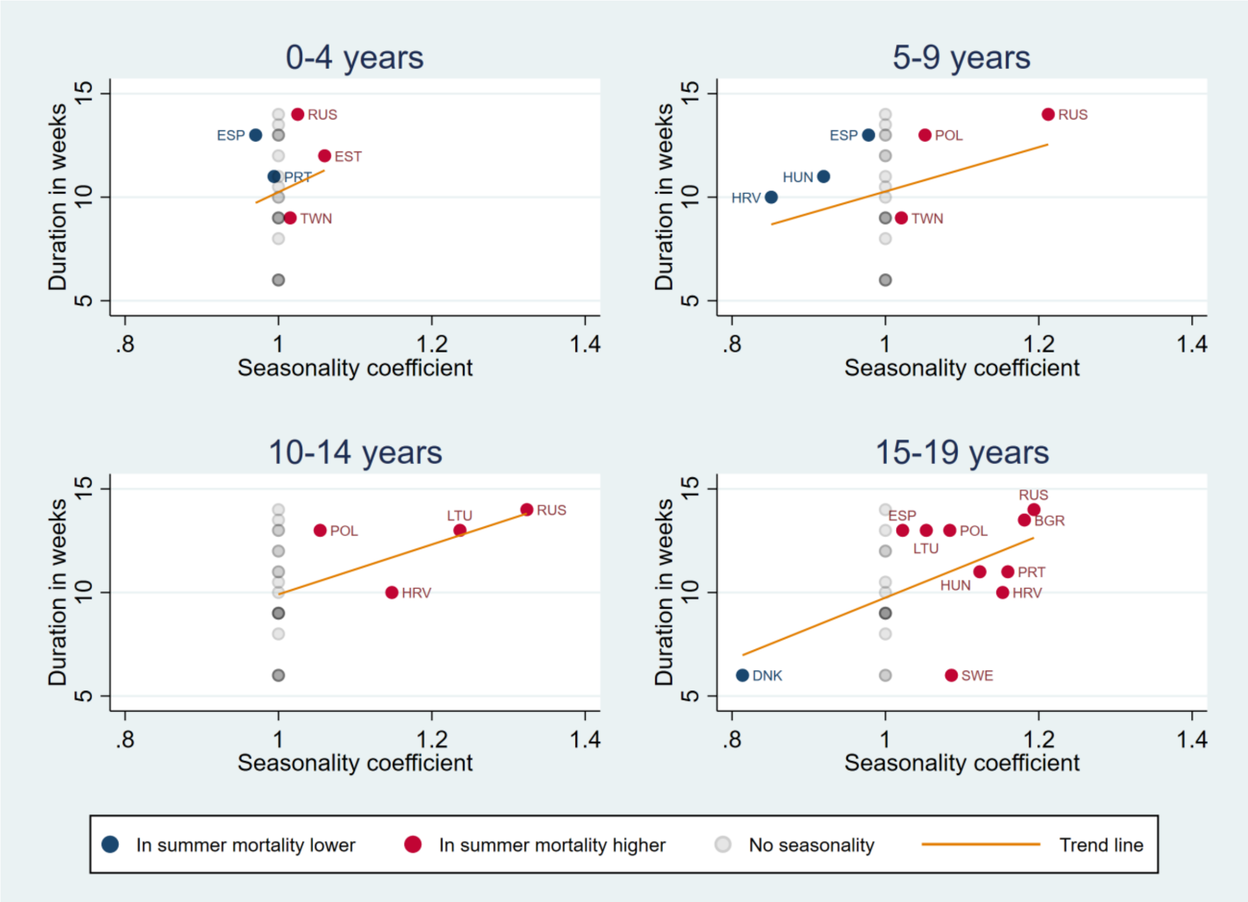
When analyzing the current state and dynamics of the structure of child mortality, it is useful to view the position of Russia in the context of other countries. Figure A3 of the Appendix displays information on the structure of child mortality in 6 countries of the world: 2 states of the Eurasian Economic Union (Belarus and Kazakhstan), two large countries of Eastern Europe (Poland and Romania) and the largest economies in Europe (Germany) and the world (USA). International comparisons, firstly, indicate the similarity of the structure of child mortality in Russia and its immediate neighbors. In addition, a comparison with the Eastern European countries that are members of the EU reveals a smaller contribution of external causes to the mortality of younger schoolchildren (5-9 years old) there than in Russia. By the age of 15-19, these differences are mostly smoothed out. However, the high contribution of events of undetermined intent, which was repeatedly mentioned above, remains a unique Russian phenomenon, not exceeding 10% in other countries (maximum values in Kazakhstan and Poland).

[3] http://demogr.nes.ru/index.php/ru/demogr_indicat/data_description

One of the reasons for the high excess summer mortality of children and adolescents in Russia may be the long summer holidays, a period when children are more often left without adult supervision. As noted in the brief literature review, a number of studies based on data from other countries have demonstrated excess mortality of school-aged children on weekends and holidays, as well as during school holidays (Murray, Carter 2017; Xu et al. 2020; Parks et al. 2018).

Our estimates generally confirm this hypothesis. Figure 4 graphically shows the relationship between the coefficient of summer seasonality of mortality (the ratio of summer deaths to annual average deaths) and the length of summer vacation in the 28 countries of the STMF database. From the data presented, it is clearly seen how higher summer mortality, from being a relatively rare phenomenon under the age of 15 years, becomes very common at the age of 15-19 years. It can also be noted that the above average summer mortality at the age of up to 15 years in Russia is greater than in all 28 countries included in the comparative analysis.

Figure 4. Summer seasonality of early childhood mortality and duration of summer school holidays, males and females, 2015-2019, 28 countries from the STMF database



Source: Authors' calculations based on STMF data (Human Mortality Database 2022).

Note: AUT – Austria, BGR – Bulgaria, CAN – Canada, CHL – Chile, CZE – Czech Republic, DEU – Germany, DNK – Denmark, FRA – France, ESP – Spain, EST – Estonia, HRV – Croatia, LTU – Lithuania, NOR – Norway, POL – Poland, PRT – Portugal, RUS – Russia, SWE – Sweden, TWN – Taiwan, USA – USA.

The positive relationship between increased summer mortality and the length of school holidays, which we previously demonstrated for children aged 0 to 15 (Figure 2), also persists in the 5-year age groups. This is largely due to the lack of summer seasonality in mortality in countries where vacations are shorter. This is indicated, in particular, by the darker gray dots at the bottom of the figure, corresponding to several countries with undetectable seasonality in mortality.

Thus, we were able to find some empirical evidence that the increased summer seasonality of mortality at younger ages in Russia may be associated with long summer holidays. In Russia, summer school holidays last 14 weeks, which is considerably longer than in most other countries included in the comparative analysis.

Our study has a number of limitations. It should be noted that there is a lack of data with the necessary detail. In particular, it is not possible to analyze the seasonality of mortality in more detailed gender and age groups by groups of causes of death. When conducting international comparisons, it was necessary, among other things, to use data for countries with small populations, which could also affect the quality of the estimates. In addition, it should be noted that the STMF database includes mainly high-income countries, while comparisons would be more appropriate for middle- and upper-middle-income countries. In poorer countries, there may be fewer opportunities to institutionally support children's summer activities, but at the same time, other factors may become more important, for example, family resources for childcare, the use of child labor in the household, etc.

Conclusion

Russia is characterized by pronounced summer excess mortality at younger ages. According to our calculations in 2015-2019, the average mortality in summer exceeded the average mortality per year for the population aged 5-9 years by 21%, for those 10-14 years old by 32%, and for those 15-19 years old by 19%.

In most of the countries included in the comparative analysis, the seasonality of child mortality in general and the summer seasonality in particular are noticeably less pronounced than in Russia, especially in the age groups of 5-9 and 10-14 years. It is only after 15 years that excess summer mortality at younger ages becomes relatively common in other countries (Figure 4).

Gender differences in the seasonality of child mortality mainly consist in the fact that for males, the seasonality confirmed by tests is always summer, while for female mortality this is not necessarily the case (Figure A1 of the Appendix for 36 countries from the STMF database). We explain this feature by the higher contribution of external causes of death to male mortality. Diseases rather than injuries play a greater role in female mortality, and therefore the effect of unfavorable winter conditions on them is stronger.

The high excess summer mortality of children and adolescents in Russia is observed against the background of a significant contribution to mortality by external causes, including those for which the annual maximum falls on the warm season (Figure 3). About 50% of deaths at the age of 15–19 years in 2019 were due to transport accidents, suicides and events of undetermined intent, i.e., causes of death with presumably summer seasonality (Shcherbakova 2010; Zakorkina 2015).

The pronounced seasonality of mortality in younger age groups may, among other things, be the result of long school holidays, during which many children find themselves without family and school supervision. To test this hypothesis, we compared the length of holidays in the countries included in the STMF database with the seasonality of mortality in children and adolescents. It has been shown that there is a significant positive relationship between these indicators: longer holidays correspond to more pronounced summer seasonality in the number of deaths, and this result is maintained for different age groups.

In Russia, summer school holidays last 14 weeks, which is noticeably longer than in most other countries. It might be useful to have discussions among educators about the academic advisability of taking such a long break from classes. The problem of summer growth in child mortality requires reflection by both experts and the public. It is necessary to expand support for family and institutional leisure and recreation for children during the holidays and, if approved by experts in the field of education, gradually reduce the duration of summer holidays, focusing on the experience of other countries.

In the future, we plan to continue studying the seasonality of mortality in Russia. Among other things, the seasonality of mortality in middle and older ages is of interest. Thus, for example, also worthy of more detailed study is the fact that the significant summer seasonality characteristic of the mortality of children and adolescents persists for young adults, with the gradual transition towards excess winter mortality occurring only after age 45.

References

Boriskin M.L., Ulesikova I.V., Shatyr Yu.A., Mulik I.G., Bulatetsky S.V., Mulik A.B. (2018). Age and gender characteristics of a person's predisposition to risky behavior. Personality in a Changing World: Health, Adaptation, Development, 6(4(23)), 741-756. http://humjournal.rzgmu.ru/art&id=352

Zakorkina N.A. (2015). Influence of seasonal factors on the causes of death of adolescents (15-17 years old) living in the territory of the Omsk region. Omsk Scientific Bulletin, 1(138), 46-49.

Ivanova A.E., Semenova V.G., Kondrakova E.V., Mikhailov A.Yu. (2009). Main trends and regional features of mortality in Russian adolescents. Social aspects of public health, 10(2), 7.

Kontsevaya A.V., Lukyanov M.M., Khudyakov M.B., Klyashtorny V.G., Balanova Yu.A., Kalinina A.M., Boytsov S.A. (2014). Seasonal and monthly changes in mortality in regions of the Russian Federation with different climatic and geographical characteristics. Russian Journal of Cardiology, 11(115), 25-30. https://doi.org/10.15829/1560-4071-2014-11-25-30

Rodionova L.A., Kopnova E.D. (2019). Statistical approaches to the analysis and modeling of seasonality in demographic data. Demographic Review, 6(2), 104-141. https://doi.org/10.17323/demreview.v6i2.9874

Semyonova V.G., Ivanova A.E., Evdokushkina G.N., Ryazantsev S.V. (2018). Mortality of children in Moscow from causes associated with violence. Social aspects of public health, 64(6), 8. https://dx.doi.org/10.21045/2071-5021-2018-64-6-8

Semyonova V.G., Ivanova A.E., Sabgaida T.P., Zubko A.V., Zaporozhchenko V.G., Gavrilova N.S., Evdokushkina G.N., Chizhikova I.O. (2021). Losses of the Russian population from external causes and the specifics of their accounting. Social aspects of public health, 67(2), 7.

CDR of the NES. (2022). Russian Database on Fertility and Mortality (ROSBRiS). URL:http://www.demogr.nes.ru/index.php/ru/demogr_indicat/data (data loaded 03.07.2022).

Shcherbakova E. (2010). Demographic Results 2009 (Part II). Demoscope Weekly, 411-412. http://www.demoscope.ru/weekly/2010/0411/barom03.php

Asif M., Nawaz K., Zaheer Z., Thygesen H., Abu-Shaheen A., Riaz M. (2019). Seasonality of deaths with respect to age and cause in Chitral District Pakistan. *Plos one*, 14(12): e0225994. https://doi.org/10.1371/journal.pone.0225994

Caporello G., Maravall H.A. (2004). Program TSW. *Revised manual: Version May 2004. Documentos ocasionales/Banco de España, 0408*. https://dialnet.unirioja.es/servlet/articulo?codigo=1033479

Chun H., Mobley M. (2010). Gender and grade-level comparisons in the structure of problem behaviors among adolescents. Journal of Adolescence, 33(1), 197-207. https://doi.org/10.1016/j.adolescence.2009.03.010

Falagas M.E., Karageorgopoulos D.E., Moraitis L.I., Vouloumanou E.K., Roussos N., Peppas G., Rafailidis P.I. (2009). Seasonality of mortality: the September phenomenon in Mediterranean countries. *CMAJ*, 181(8), 484-486. https://doi.org/10.1503/cmaj.090694

Grudkowska S. (2015). Jdemetra+ user guide. URL: https://ec.europa.eu/eurostat/cros/system/files/jdemetra_user_guide.pdf

Healy J.D. (2003). Excess winter mortality in Europe: a cross country analysis identifying key risk factors. *Journal of Epidemiology & Community Health*, 57(10), 784-789. https://doi.org/http://dx.doi.org/10.1136/jech.57.10.784

Human Mortality Database (2022). Short-term mortality fluctuations (STMF) data series. Retrieved from https://www.mortality.org/ (Accessed 03.06.2022).

Knyazev G.G., Slobodskaya H.R., Kharchenko I.I., Wilson G.D. (2004). Personality and substance use in Russian youths: The predictive and moderating role of behavioural activation and gender. *Personality and Individual Differences*, 37(4), 827-843. https://doi.org/10.1016/j.paid.2003.10.010

Lantos T., Nyári T.A., McNally R.J. (2019). Seasonal variation of mortality from external causes in Hungary between 1995 and 2014. *PLoS one*, 14(6): e0217979. https://doi.org/10.1371/journal.pone.0217979

O'Keeffe S.T. (2017). Summertime blues? A re-examination of the seasonality of web searches for restless legs and leg cramps. *Sleep medicine*,1(37),119-23.

Maravall A., López-Pavón R., Pérez-Cañete D. (2015). Reliability of the automatic identification of ARIMA models in program TRAMO. In *Empirical Economic and Financial Research* (pp. 105-122). Springer, Cham.

McKee M., Sanderson C., Chenet L., Vassin S., Shkolnikov V. (1998). Seasonal variation in mortality in Moscow. *Journal of Public Health*, 20(3), 268-274. https://doi.org/10.1093/oxfordjournals.pubmed.a024768

Murray K., Carter P. (2017). Fatal drownings in Fiji: an effective parsimonious model that can explain the number of cases from January 2012 to April 2015. *Asia Pacific Journal of Public Health*, 29(1), 28-34. https://doi.org/10.1177/1010539516685610

Nemtsov A.V., Fattakhov T.A. (2021). Diagnostic and age composition of excess mortality associated with the New Year holidays in Russia. *Population and Economics,* 5(4): 1-20. https://doi.org/10.3897/popecon.5.e72550

Parks R.M., Bennett J.E., Foreman K.J., Toumi R., Ezzati M. (2018). National and regional seasonal dynamics of all-cause and cause-specific mortality in the USA from 1980 to 2016. *eLife*, 7: e35500. https://doi.org/10.7554/eLife.35500

Peden A.E., Barnsley P.D., Queiroga A.C. (2019). The association between school holidays and unintentional fatal drowning among children and adolescents aged 5–17 years*. Journal of paediatrics and child health*, 55(5), 533-538. https://doi.org/10.1111/jpc.14235

Revich B., Shaposhnikov D. (2010). The effects of particulate and ozone pollution on mortality in Moscow, Russia. *Air Quality, Atmosphere & Health,* 3(2), 117-23. https://doi.org/10.1007/s11869-009-0058-7

Sethi D., Aldridge E., Rakovac I., Makhija, A. (2017). Worsening inequalities in child injury deaths in the WHO European region. *International journal of environmental research and public health*, 14(10), 1128. https://doi.org/10.3390/ijerph14101128

Shinsugi, C., Stickley, A., Konishi, S., Ng, C. F. S., & Watanabe, C. (2015). Seasonality of child and adolescent injury mortality in Japan, 2000–2010. *Environmental health and preventive medicine*, 20(1), 36-43. https://doi.org/10.1007/s12199-014-0421-7

Shaposhnikov D., Revich B., Bellander T., Bedada G.B., Bottai M., Kharkova T., Kvasha E., Lezina E., Lind T., Semutnikova E., Pershagen G. (2014). Mortality related to air pollution with the Moscow heat wave and wildfire of 2010. *Epidemiology (Cambridge, Mass.)*, 25(3), 359-364. https://doi.org/10.1097/EDE.0000000000000090

UNECE (2020). *Practical Guide to Seasonal Adjustment with JDemetra+*. United Nations. Retrieved from https://www.un-ilibrary.org/content/books/9789210050043 (Accessed 03.06.2022)

Weden M.M., Zabin L.S. (2005). Gender and ethnic differences in the co-occurrence of adolescent risk behaviors. *Ethnicity and Health* 10(3), 213-234.

WHO (2022). World Health Organization Mortality Database. Retrieved from http://www.who.int/healthinfo/mortality_data/en/ (Accessed 25.08.2022).

Xu X.H., Dong H., Li L., Liu W.H., Lin G.Z., Ou C.Q. (2020). Trends and seasonality in cause-specific mortality among children under 15 years in Guangzhou, China, 2008–2018. *BMC public health*, 20(1), 1-9. https://doi.org/10.1186/s12889-020-09189-0

Демографическое обозрение / Demographic Review
2022;9(3) 24-43 DOI: https://doi.org/10.17323/demreview.v9i3.16468

Who needs (or doesn't) three years of parental leave?

Irina E. Kalabikhina
(kalabikhina@econ.msu.ru), Lomonosov Moscow State University, Russia.
Polina O. Kuznetsova
(polina.kuznetsova29@gmail.com), Institute for Social Analysis and Forecasting at RANEPA, Lomonosov Moscow State University, Russia.

Abstract: *In this work we study the actual length of parental leave in Russia and the factors influencing it. In our estimates we use a survival analysis model applied to 20-year panel data from the nationally representative survey of the HSE RLMS.*
According to the results, shorter actual parental leave is more often observed among women who have better connections with the labor market (a higher level of education, work experience) and wider external and intra-family resources for childcare. In contrast, older age, having another child and living in areas with lower mean wages discourage young mothers from returning to work. We also show that the probability of an earlier return to work after the birth of a first child in the 2010s decreased significantly compared to the 2000s, which we explain by the introduction of the maternity capital program and increased state support of families with children, which together with the stagnation of labor incomes could reduce the interest of families in an earlier return of young mothers to work.
International comparisons show that Russia belongs to the group of countries with the longest paid maternity and parental leave; at the same time, it is only in the 4th quintile in terms of the labor force participation rate among women aged 15-64. Taking these results into account, we suggest some improvements to the parental leave policy. Among other possibilities, we suggest considering flexible leave (reducing its duration while keeping the same total payment) and discontinuous leave (giving parents the right to go back on unpaid leave if necessary), as well as the introduction of non-transferable paternity leave.

Keywords: *mothers, parental leave, maternity leave, employment of mothers, family policy, survival analysis, RLMS.*

Funding: *The research was conducted with the financial support of the State task of Moscow State University No. 122041800047-9 "Reproduction of the population in socio-economic development" and the Russian Foundation for Basic Research (RFBR) under grant No. 19-29-07546 "The Impact of Human Capital on Present and Future Economic Growth in Russia".*

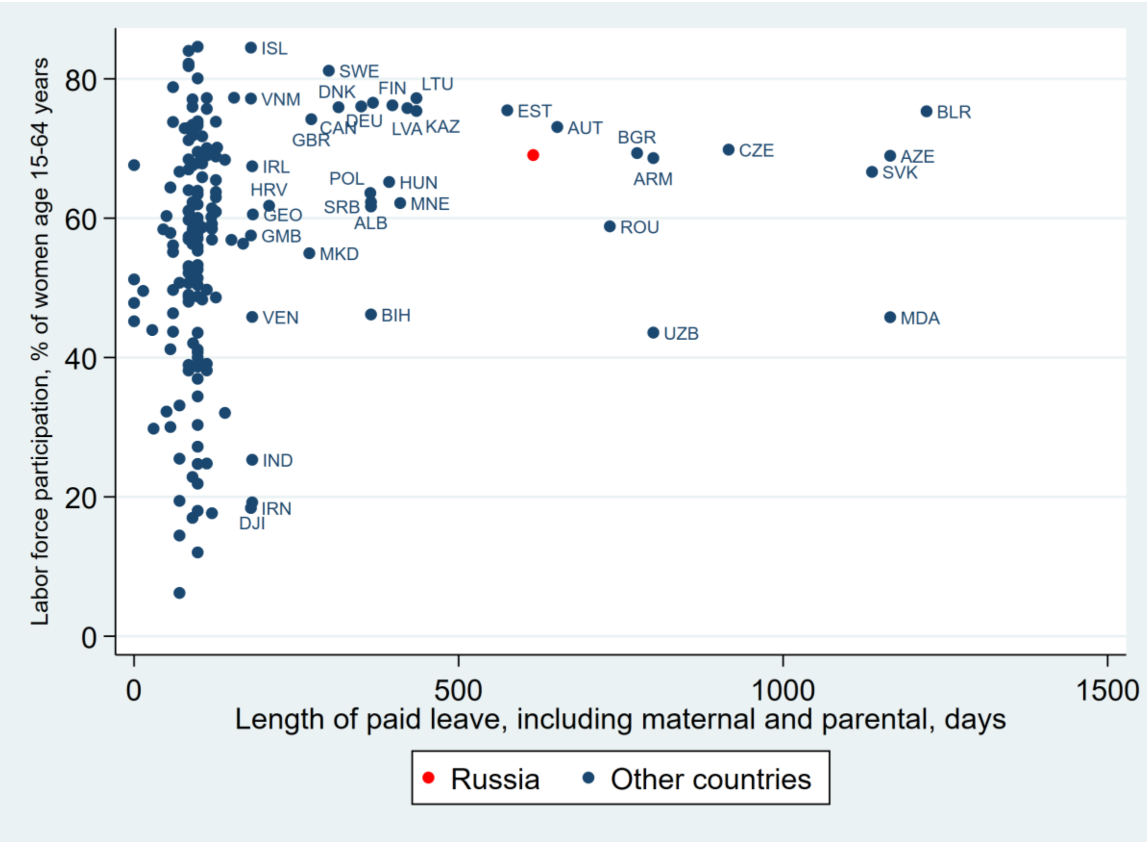
For citation: *Kalabikhina I. & Kuznetsova P. (2022). Who needs (or doesn't) three years of parental leave? Demographic Review, 9(3), 24-43. https://doi.org/10.17323/demreview.v9i3.16468*

1. Introduction

Despite relatively high overall female employment, Russia lags behind most developed countries in terms of the share of working women with children under the age of three: 24% versus 59% on average for OECD countries (Kolesnik, Pestova, Donina 2021). Among the reasons for this lag are the lack of childcare services, the lack of places in the preschool education system, the lack of partner support in caring for a child and the longer duration of parental leave compared with other countries.

Figure 1 summarizes information on the economic activity of women depending on the length of paid parental leave (including leave for pregnancy and birth) in 189 countries in 2019. A long leave is observed only in the countries of the former USSR (Azerbaijan, Armenia, Belarus, Moldova, Uzbekistan) and in the former countries of the socialist bloc (Bulgaria, Romania, Slovakia, Czech Republic). At the same time, the level of economic activity of women in Russia is not high compared to other countries. Russia belongs to the 75st percentile of the countries of the world in terms of the level of economic activity of women aged 15-64 years.

Figure 1. Length of paid leave and economic activity of women in 189 countries, 2019



Source: Authors' calculations based on World Bank data https://genderdata.worldbank.org/indicators/#gender

Over the past decades, many countries have reformed their parental leave systems, thus providing researchers with data from a kind of natural experiment on the effect of the length and generosity of paid parental leave on fertility and subsequent female employment. For example, a reform carried out in 2007 in Germany included the replacement of a targeted allowance for up to two years with an earnings-related allowance paid over a period of one year. It has been shown (Bergemann, Riphahn 2022; Kluve, Tamm 2013) that, despite the lack of impact on employment in the long term, the short-term effect of the reform was noticeable: mothers began to return to paid work earlier, immediately after the expiration of benefits. The reform had the greatest impact on high-income mothers, among whom employment of those with children older than a year increased significantly (Chirkova 2019).

In France, on the contrary, an experiment was undertaken to lengthen paid leave for a particular group of recipients. Starting in 2004, mothers of first children were offered the opportunity to extend paid parental leave for up to six months in the event of part-time employment. (Joseph et al. 2013) showed that for those who took longer leaves, employment increased at the end of the vacation, but relative wages decreased, with the effect most pronounced for the most educated women.

Numerous empirical studies suggest that the impact of job security for young mothers on the likelihood and quality of their future employment is mixed. Pettit and Hook (2005) have shown that the effect of parental leave duration on subsequent employment is not linear. The dependence of the probability of a mother's employment on the length of leave has an inverted U-shape: for not too large values, this effect is positive; the presence of a leave helps to maintain the connection of mothers with the labor market, but its extension significantly reduces the likelihood of employment.

A comparative study based on data for OECD countries (Thévenon, Solaz 2013) also confirms an inverted U-shape of the relationship between employment and vacation length: paid vacation, if less than two years, has a small positive effect on women's employment and contributes to a small narrowing of the gender gap in employment, while longer leaves reduce the likelihood of employment. In the case of wages, the dependence is unequivocal: longer paid leave contributes to the growth of the gender gap in wages. After a long maternity leave, a significant part of professional skills and human capital is lost, which subsequently becomes an important factor in the motherhood penalty, and also leads to an increase in the gender wage gap and a deterioration in the situation of the most vulnerable groups of workers, including mothers raising a child without a father (Morosow, Jalovaara 2019).

There is evidence that longer leave has a negative impact on the quality of a mother's subsequent employment. Finnish data have shown that a childcare allowance up to age three contributes to a longer absence from the workforce, which subsequently negatively affects women's working career (Morosow, Jalovaara 2019). In a study using data for Germany and the United States, it was shown that mothers who previously had a high employment status return to their jobs faster and are less likely to leave their careers (Grunow, Aisenbrey 2016). In another study, the authors showed that long periods of absence of paid work for women increase the risk of downward mobility and reduce the chances of career growth not only in Germany and the United States, but also in Sweden, where the level of support for maternal employment is traditionally high (Aisenbrey, Evertsson, Grunow 2009).

In our study, we wanted to look at how Russian women manage their leave. We were interested in which demographic and social groups of women tend to "sit out" their maternity leave or, conversely, go to work earlier than others.

The literature distinguishes 4 groups of factors influencing the decision concerning the use and duration of parental leave (Lapuerta, Baizán, González 2011). First, there are the individual characteristics of the parent on leave and the household in which he/she lives. Among the factors influencing the length of the leave, researchers note the level of education, income, type of ownership of the enterprise the parent works for, the birth order of the child in the family (Evertsson, Duvander 2011). A higher level of education tends to correspond to a mother's greater connection to the workforce and therefore contributes to her earlier return to work. Women from low-income households stay on leave longer. The duration of a mother's leave in connection with the birth of a first child is slightly longer compared to subsequent ones (Hobson, Duvander, Halldén 2006). Studies also show non-random selectivity for the birth of a third and subsequent children (Andersson 2008).

The group of individual factors also includes the characteristics of the parent's employment, including the type of enterprise ownership, the number of working hours (the possibility of part-time work), working conditions, distance from home (Crompton 2006), and the level of wages. For example, work in a private enterprise has been shown to reduce leave time (Bygren and Duvander 2006), and workers with higher wages tend to return to work faster (Sundström and Duvander 2002).

The second group of factors are the characteristics of the partner, primarily the indicators of his education and employment (Lundberg, Pollack 1993). There is evidence both for the significance of these factors (Lappegard 2008) and for their significantly smaller effect on the length of leave compared to the individual characteristics of the parent on leave (Kuhlenkasper, Kauermann 2010). The third type of factors are related to gender values and societal norms, including the social acceptability of different family care arrangements and the division of care time and effort between partners (Paye and Sinyavskaya 2010; Zhou and Kan 2019). The fourth group of factors includes various institutional characteristics of the national and/or regional system of assistance to families in the care of children.

Russian studies of the employment of women with children mainly deal with the topic of the motherhood penalty (see, for example, (Nivorozhkina, Nivorozhkin, Arzhenovsky 2008; Oshchepkov 2020)). Based on Russian data, the size of the motherhood penalty was obtained, which is estimated as the percentage of losses in wages of women after the birth of a child compared with women without children with similar individual characteristics. For example, in the early 2000s, the penalty was estimated at 8% (Arzhenovsky, Artamonova 2007), and in a later work (Biryukova, Makarentseva 2017) it was about 4% for 2014. It was also shown that for more educated mothers, the relative size of the penalty is greater (Yermolina et al. 2016). In addition, there is empirical evidence that despite the presence of a motherhood penalty, the difference in wages between women with children and those without is generally smoothed out by the time a child reaches the age of 3-4 years (Yermolina et al. 2016).

At the same time, as far as we know, there are no Russian studies that examine the factors of the length of maternity leave. In our study, we will try to fill this gap and compare the results with estimates for other countries.

The article has the following structure. Section 2 presents the data and methodology of a quantitative study and summarizes the hypotheses formed based on the results of a literature review and a preliminary analysis of data for subsequent testing using regression analysis. This is followed by a discussion of the estimates obtained during the quantitative study (Results ection). The work concludes with a brief discussion about improving parental leave policy and the main conclusions of the study.

2. Data, methodology and research hypotheses

The object of our study is the duration of maternity leave, defined as the period of absence from the workforce from the moment a child is born. Using such a broad definition of leave allows us to include previously unemployed women in the analysis.

Sources of data on the length of maternity leave in Russia

We managed to find two sources of data on the duration of maternity leave in Russia: Russian Monitoring of the Economic Situation and Health of the Population of the National Research University Higher School of Economics (hereinafter RLMS) and Comprehensive Monitoring of Living Conditions of the Population of Rosstat (hereinafter CMLC). Each of these surveys has its own advantages and disadvantages.

The RLMS data have a panel structure that makes it possible to track events in the lives of respondents over several years. At the same time, we were able to estimate the duration of maternity leave in the RLMS only in years, not in months, as is done in most international studies. This is due to the lack of information in the RLMS about how many months the respondent worked in a given year (in a normal situation, it is not important). Thus, it is not possible to specify exactly when the woman returned from maternity leave.

The Rosstat CMLC data are not panel data, i.e., each time they represent a situation at a specific point in time without the possibility of comparing individual observations for different years. However, the survey for those on parental leave asks respondents to “Specify the month and year you went on leave, and in which month of this year you started work”, which makes it possible to calculate the duration of the leave in months.

However, an attempt to include these data in the analysis was unsuccessful. The start date of the leave given by the respondents does not correlate well with the date of birth of the child, and the results of the regression analysis of the duration of the leave turned out to be counterintuitive (higher education contributes to a longer leave) and contradicted both the calculations based on the RLMS data and the results of studies for other countries. Most likely, the problem lies in the fact that respondents often find it difficult to name the exact date of a very distant event.

Empirical basis and construction of variables

The empirical basis of the study are the RLMS data for 2000-2020. The analysis used the longitudinal structure of the survey data, which makes it possible to follow the changes that occur with a woman and the household in which she lives from the moment of birth of a child to the start of work (completed observation episode) or until the last time she was present in the survey database (censored episode). We interpreted the duration of the leave as broadly as possible, defining it as the time spent by a woman not in the workforce from the moment the child was born.

We excluded the self-employed from the number of employed, since this type of employment is often unstable (Bobkov et al. 2017). However, the calculations confirmed the robustness of the results to a broader definition of employment (including the self-employed).

We only looked at first births during women’s participation in the survey. This method of selection made it possible to avoid repeated inclusions: each respondent was taken into account in the model only once. The birth of subsequent children was accounted for by a special variable.

The proportion of first births in our sample was 61%, second births 30%, third and subsequent births 9%.

Thus, the object of our study is young mothers - women who had children during their presence in the RLMS sample, and it is important to note that 'young' in this case refers not to the age of the mother, but to the age of the child.

We used a number of independent variables to identify factors influencing duration of leave (see Table 1 for a complete list). The individual characteristics of the mother were considered, including age, marital status (registered or unregistered marriage), level of education, and the presence of paid work just before the child's birth. In addition, the decision to go to work can be influenced by the number and ages of other children in the family, in connection with which we included in the model information about the child's birth order, as well as about the birth of one or more other children during the period of observation.

To take into account a family's having external assistance in the care and upbringing of children, the independent variables "relative assistance" and "non-relative assistance" were included in the model (constructed, respectively, based on the questions "During the past 7 days, have relatives who live separately helped take care of the child?" and "During the past 7 days, have other people who are not your relatives helped take care of the child?"). Information on external care assistance in the survey is provided for each minor child. When forming the variables, we used data for the child for whose care the woman was on leave at the time of observation.

In addition to external assistance, assistance to the young family can be provided by cohabiting relatives, in connection with which we added the variables "living together with the parent(s)" and "living together with the parent(s) of the husband". When constructing variables, we used data on family ties between family members from the RLMS household file.

The model also included information about the woman's subjective perception of the family's financial situation and the likelihood of losing her job. A low level of satisfaction with the financial situation was determined using the question "How worried are you that you will not be able to provide for yourself in the next 12 months?" (response "Very worried"). Fear of losing her job was assessed using the question "How worried are you that you might lose your job?" (response "Very worried").

To take into account regional specificities, we included the variable "quintile group of regions by the level of average wages" in the model. The values of the average salary were considered in comparable prices, i.e., taking into account the size of the cost of a fixed set of consumer goods and services.

As a dependent variable, we look at the number of years spent away from work since the birth of a child. The average length of leave thus determined was on average 2.95 years for employed and unemployed women, 2.86 years for first births and 3.06 years for second and subsequent births (Table 1). For women with work experience, the average length of leave was lower - 2.56 years, for those with higher education - 2.66 years. The mean values of the independent variables included in the analysis are also presented in Table 1.

Table 1. Mean values of the variables included in the regression analysis

	Model 1 (all cases)	Model 2 (cases of first birth)	Model 3 (cases of second and subsequent births)
Duration of an "out of workforce" episode due to the birth of a child	2.946	2.875	3.06
Married, including unregistered	0.882	0.848	0.937
Locality type			
Moscow and St. Petersburg	0.089	0.090	0.088
Regional centers	0.327	0.366	0.264
Other cities	0.277	0.274	0.282
Urban-type settlement and village	0.307	0.270	0.366
Education			
Not above general secondary	0.177	0.156	0.212
Primary vocational	0.212	0.187	0.250
Secondary professional	0.232	0.247	0.206
Higher professional	0.380	0.410	0.332
Birth order of child			
First	0.615	1.000	
Second	0.298		0.775
Third and subsequent	0.087		0.225
Another child was born	0.216	0.275	0.123
Mother's age			
Under 25	0.179	0.267	0.039
25-29 years	0.330	0.422	0.184
30-34 years	0.261	0.205	0.350
35 years and older	0.230	0.106	0.427
Has work experience	0.646	0.614	0.698
Period of child's birth			
2000-2005	0.213	0.228	0.189
2006-2010	0.280	0.298	0.250
2011-2015	0.330	0.326	0.337
2016-2020	0.177	0.147	0.224
Availability of care assistance			
Non-relatives	0.058	0.066	0.047
Relatives	0.335	0.351	0.309
Low assessment of financial situation	0.390	0.374	0.417
Fear of losing job	0.446	0.420	0.487
Lives with child's grandparents	0.308	0.358	0.229
Number of observations	1952	1201	751

Source: Authors' calculations based on RLMS data.

Quantitative research methodology

To study the factors of the duration of women's absence from work after childbirth, we use a regression analysis of survival using a nonparametric Cox model in discrete time (Cox 1972; Klein, Moeschberger 2003).

As the dependent variable, we consider the number of years spent out of the workforce since the birth of the child. For a discrete random variable T, which reflects the length of a young mother's absence from the workforce, the risk function λ is defined as follows:

$$\lambda(t)=P(T=t|T\geq t). \quad (1)$$

For a state that lasted *t* years, this function reflects the probability of its termination within the next year. The discrete Cox model defines the functional form of the relationship between risk and explanatory variables as follows (Klein, Moeschberger 2003: 259):

$$\frac{\lambda(t,X,\beta)}{1-\lambda(t,X,\beta)} = \frac{\lambda_0(t)}{1-\lambda_0(t)}\exp(\beta' X), \quad (2)$$

where *X* is the vector of explanatory variables, β the estimated coefficients, and $\lambda_0(t)$ the basic risk function equal to its value in the absence of the impact of explanatory variables, i.e., when β^' *X*=0. Vector *X* contains information about the individual characteristics of the young mother, her partner and her family (see Table 1).

RLMS data on the status of a young mother in the labor market is incomplete, due to interval censoring and censoring on the right. Censoring on the right is observed for incomplete episodes of going back to work, when the episode of the woman's absence from the workforce during the observation period has not ended. The use of duration analysis methods makes it possible to solve the problem of estimation bias due to right-hand censoring.

Interval censoring occurs because in some cases there are gaps in the observations, and we cannot understand whether the woman went to work at that time or not. The removal of observations with omissions reduced the sample by about 10 per cent.

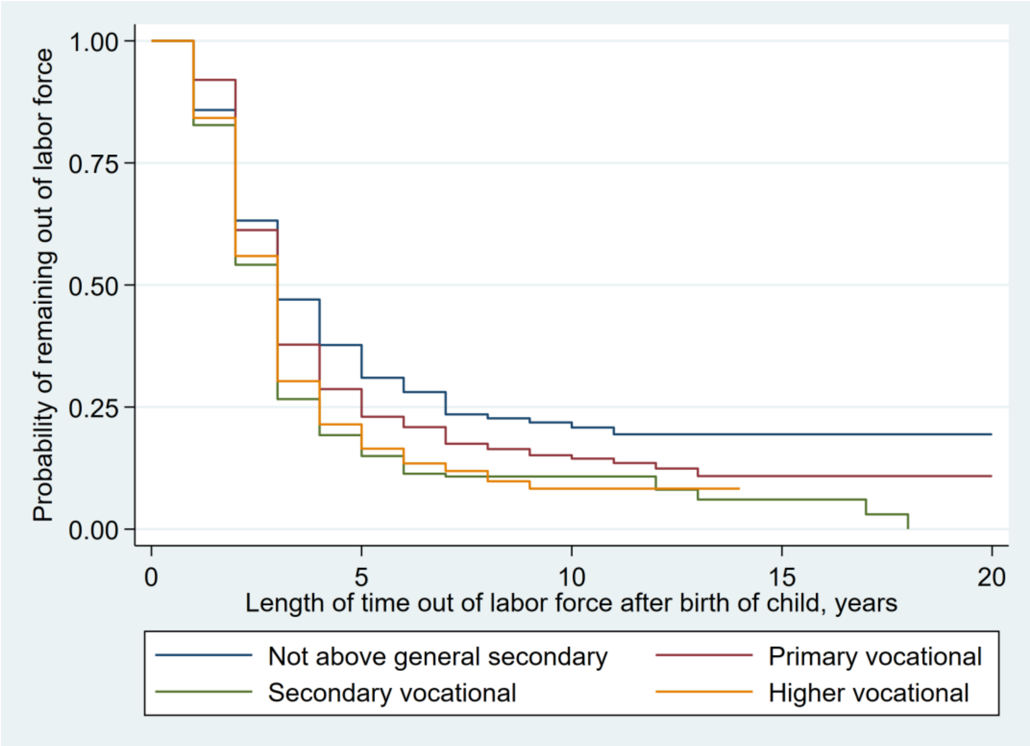
Hypotheses

Kaplan-Meier diagrams make it possible to study pairwise dependences of the length of mothers' leave from work on the individual and household factors of interest to us. We have included information on the probability distribution of staying out of the workforce for groups of women depending on the level of education, the child's birth order and the time period of his birth (Figures 2-4).

A comparison of the distribution of the probability of remaining on parental leave as a function of education level indicates an earlier return to work for women with higher and secondary vocational education (Figure 2). The graph also clearly shows that in the long term, 10-15 years after the birth of a child, the economic activity of women with a low level of education is noticeably lower than for everyone else.

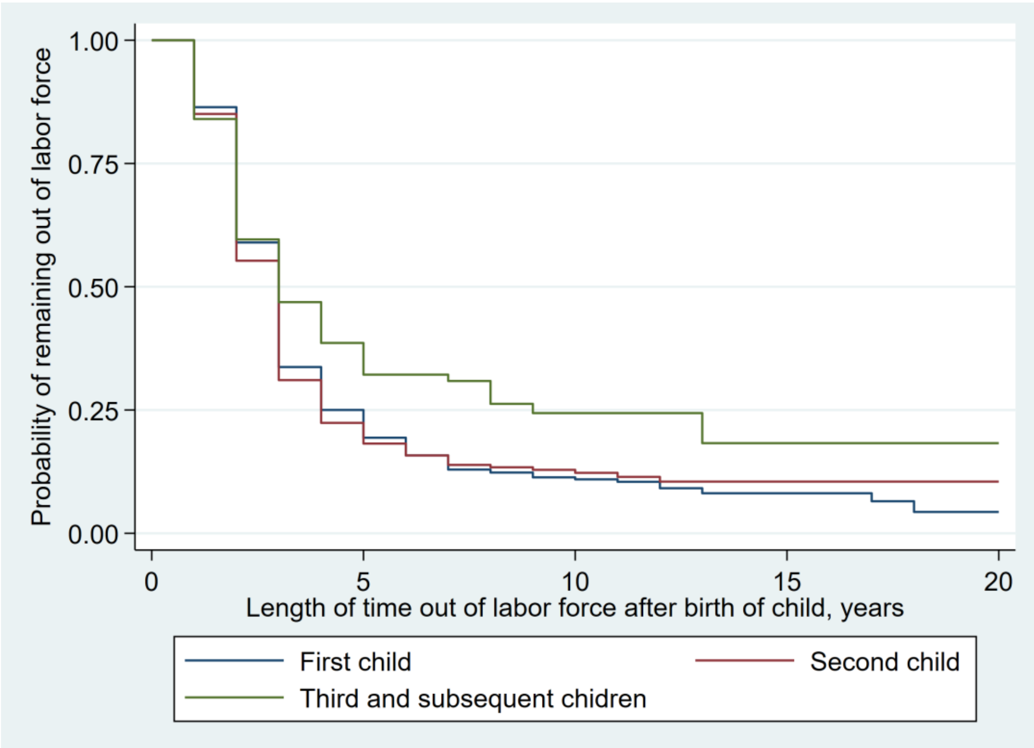
Also of interest is the effect of a child's birth order on a mother's return to work (Figure 3). There is evidence in the literature that in the case of a first child, the actual length of leave is slightly longer than in other cases (Hobson, Duvander, Halldén 2006). The data in Figure 3 do not allow us to reach a definite conclusion; we can only note that the probability of an early return to work for second births is consistently higher than for first ones. With births of a higher order, there is no complete clarity, but it is clear that in the long term, the economic activity of mothers with many children is expected to be significantly lower than for the rest.

Figure 2. Kaplan-Meier diagram for the probability of not working as a function of a mother's education level



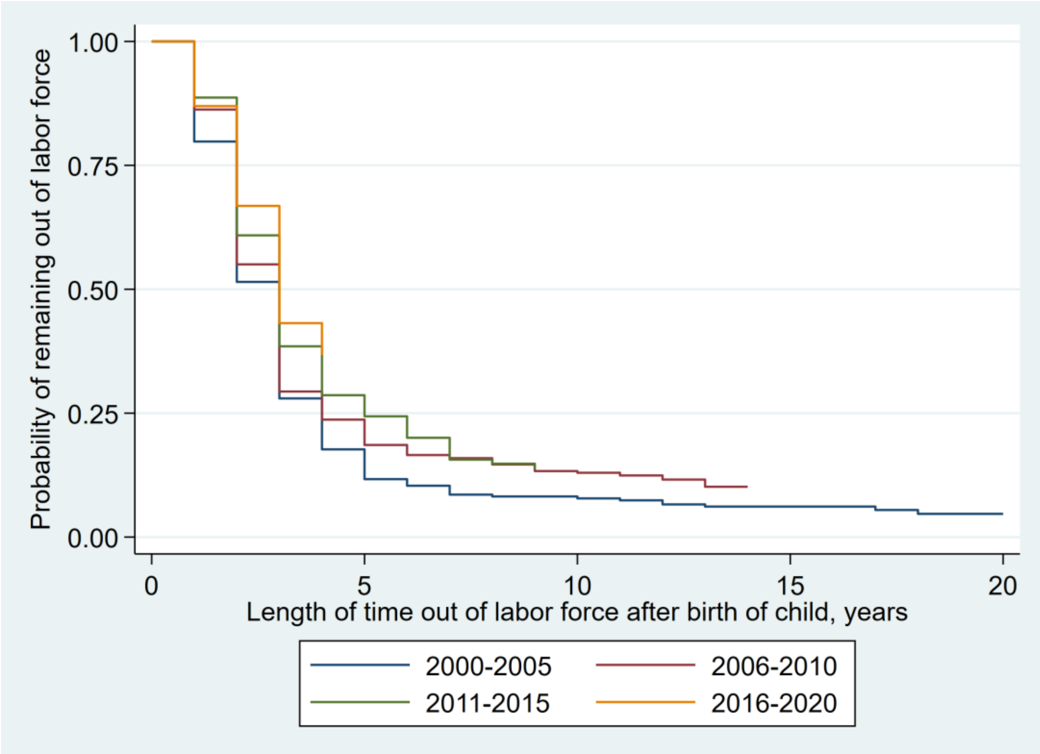
Source: Authors' calculations based on RLMS data.

Figure 3. Kaplan-Meier diagram for the probability of not working as a function of a child's birth order



Source: Authors' calculations based on RLMS data.

Figure 4. Kaplan-Meier diagram for the probability of not working as a function of a child's period of birth



Source: Authors' calculations based on RLMS data.

Another factor whose impact on the likelihood of an early return to work is of interest is the period of a child's birth. The 20 years under study were eventful in terms of the country's demographic and economic development. During this time, a number of demographic policy measures were implemented and periods of economic growth and relative stability were replaced by noticeable economic shocks. How did these affect the labor activity of mothers? The answer is clear (Figure 4): in the 2010s, young mothers returned to work later than in the 2000s. The details of this process, whether it depends on the child's birth order, will be looked at more closely using regression analysis.

A review of earlier studies on the factors of young mothers returning to work, as well as a preliminary analysis of the characteristics of mothers and their households using Kaplan-Meier diagrams, allow us to formulate the following empirical research hypotheses.

1. Contributing factors to a mother's early return to work include work experience (Makay 2017) and a high level of education (Grunow, Aisenbrey 2016), a higher birth order of the child (Hobson, Duvander, Halldén 2006), the availability of help with childcare from relatives and non-relatives both in and outside the household, and living in a region with a developed labor market.
2. The chances of an earlier return to work are reduced by a higher age of the mother and a later period of birth (the effect of expanding state support for families and children, as well as the stagnation of real wages, which reduces the gain from returning to a paid job).
   The effect of a woman's marital status on the probability of ending maternal leave can be multidirectional: on the one hand, the presence of a partner can make the financial situation of the family more stable and contribute to a longer leave. On the other hand,

the partner can take on some of the childcare responsibilities, thereby making it easier for the mother to go to work.

3. The nature of the dependence of the probability of returning to work on the characteristics of the mother and the household in which she lives may depend on the child's birth order. To take into account this possibility, 3 econometric models were considered: in addition to the full sample, subsamples for mothers of a first, as well as a second and subsequent children, were considered.

Regression analysis, in contrast to pairwise comparisons, makes it possible to assess the influence of the listed factors, other things being equal, i.e., taking into account the influence of other characteristics of the mother and her family, which makes it possible to significantly refine the results of the preliminary analysis.

3. Results

We assessed the chances of new mothers returning to work using a non-parametric Cox model of event duration analysis. We consider the duration of a woman's leave from work from the moment (year) of the birth of a child, with the indicator of the exit from this state being the woman's return to work. Three sampling options were considered: all observations, observations for the first child, observations for the second and subsequent children. It was not possible to separately consider the sample for mothers with many children (having three or more children) due to the small number of such observations. The results of the regression analysis are presented in Table 2.

Some of the factors that have a significant impact on the chances of young mothers returning to work are in line with our expectations. So, for example, the birth of another child significantly reduces the likelihood of a woman returning to work, while work experience, as well as the fear of losing her job, on the contrary, significantly increase it.

The mother's high level of education also has an expected effect on the duration of the leave: the presence of secondary and, in particular, higher vocational education contributes to an earlier return to work. For mothers of second and subsequent children, the effect of education is especially pronounced. Thus, the hypothesis of shorter leaves for the more educated and highly skilled, repeatedly confirmed in international studies (Grunow, Aisenbrey 2016; Hobson, Duvander, Halldén 2006), also turned out to be correct for Russian data.

The presence of a partner (both in a registered and unregistered marriage) does not affect the duration of the leave. In the calculations, we also considered other versions of the model, including various characteristics of the woman's husband/partner, such as the partner's employment, higher education, a registered marriage, etc. According to the results obtained on Russian data, the significance of the impact of the woman's family status and the partner's characteristics on the duration of the leave is not confirmed. In the scientific literature, as a rule, it is noted that the presence of a partner allows a woman to be on maternity leave longer, although it should also be noted that in foreign works the leave is shorter (Morosow, Jalovaara 2019; Kuhlenkasper, Kauermann 2010).

External assistance from non-relatives is not common. Thus, on average, about 6% of women in the sample used it (see Table 1). As a rule, this assistance is paid (more than 65% of cases of assistance for the care of children under 5 years of age), which suggests that this is mainly babysitting services. According to the results, the presence of non-relative assistance significantly contributes to an earlier return to work in the case of the first child and is an insignificant factor for the subsample of mothers of second and subsequent children.

Regardless of the child's birth order, the factor of help from relatives who are not part of the household turned out to be important. Cohabitation with the child's grandparents (only in the case of the first child) also encourages women to go to work.

Let us note the influence of the period of a child's birth, especially of a first child. The probability of returning to work over the past 10 years has been steadily declining compared to the period from 2000 to 2005. This is partly due to the strengthening of the state's demographic policy. In 2007, the state introduced a maternity capital program, which may have had an impact on the family's need to earn money for housing and children's education, as well as other family support measures. The expansion of the child allowance system, which has taken place regularly since 2007, may have reduced the need for families to earn additional income. Another explanation for women's longer absence from the workforce in the 2010s compared to the 2000s could be that labor income grew much more slowly during this period, which in turn reduced the economic attractiveness of women returning to work earlier.

In order to take into account the peculiarities of local labor markets, dummy variables for quintile groups by average wages were added to the model. Note that the RLMS sample includes only a part of the regions of Russia. According to the results, living in regions with relatively low wages reduces the likelihood of an earlier return to work. This is understandable - compensation for returning to work in the form of a difference in labor income and childcare benefits is lower in such regions, which makes returning to employment less attractive.

The characteristics of the locality turned out to be insignificant - there were no significant differences in the probability of returning to work for women living in regional centers and other cities compared with residents of rural areas.

The age of the mother reduces the chance of returning to work: compared to women under 25 years of age, those in the age groups of 25-29 years and especially those of 30 and older are less likely to return to work from maternity leave. Of course, the effect of age is partly offset by other characteristics that, on the contrary, contribute to an earlier return from maternity leave and are more common among older women - the presence of a paid job before the birth of a child and the experience of returning to work after the birth of a previous child. Nevertheless, the problem of age exists: other things being equal, it is much more difficult for women after thirty to go to work after the birth of a child.

In a work based on American data (Macran, Joshi, Dex 1996), it is noted that the duration of maternity leave decreases with the age of the mother. We have the opposite results, but most likely the differences are due to the specifics of the American case (very short paid leave), as well as the fact that we look at all women (both working and not working before the birth of a child).

Another important result of the analysis is the significance of a child's birth order. Taking into account other characteristics of the mother, the probability of terminating a leave taken to care for a third and subsequent children is higher than for a leave connected with the birth of a second and especially first child. This effect is typical for an earlier (up to three years) return from maternity leave, although in general, mothers with many children are expected to be out of the labor market more often in the long term. We assume that when a mother returns to work again, the family uses already developed schemes for replacing part of her household work, for example, getting help from relatives both inside and outside the family and using babysitting services and preschool institutions, thus making possible a quicker return to work. It is also possible that the more children there are in the family, the more urgent is the need for additional income, and as a result, the more difficult it is for the mother not to work or to be on unpaid leave.

Table 2. Results of regression analysis of parental leave duration factors, logarithm of the odds ratio of going to work (Cox model)

	Model 1 (all cases)	Model 2 (cases of first birth)	Model 3 (cases of second and subsequent births)
Married, including unregistered	-0.082 [0.087]	-0.059 [0.100]	-0.106 [0.184]
Type of locality; omitted category - village, urban-type settlement			
Regional centers	-0.07 [0.073]	0.01 [0.093]	-0.201* [0.118]
Other cities	-0.02 [0.076]	0.028 [0.100]	-0.083 [0.119]
Education; omitted category - not above general secondary			
Primary vocational	0.009 [0.094]	-0.102 [0.126]	0.118 [0.142]
Secondary vocational	0.239*** [0.092]	0.164 [0.118]	0.309** [0.150]
Higher professional	0.327*** [0.091]	0.233** [0.119]	0.432*** [0.147]
Birth order of child; omitted category - first child, for model 3 – second child			
Second	0.334*** [0.071]		
Third and subsequent	0.501*** [0.122]		0.207* [0.122]
Another child was born	-0.308*** [0.068]	-0.297*** [0.079]	-0.409*** [0.144]
Mother's age; omitted category - under 25 years			
25-29 years	-0.456*** [0.084]	-0.440*** [0.093]	-0.636** [0.266]
30-34 years	-0.850*** [0.098]	-0.912*** [0.120]	-0.958*** [0.265]
35 years and older	-1.275*** [0.112]	-1.241*** [0.149]	-1.435*** [0.269]
Has work experience	0.673*** [0.066]	0.651*** [0.081]	0.781*** [0.120]
The period of the child's birth; omitted category - 2000-2005			
2006-2010	-0.121 [0.075]	-0.152 [0.094]	-0.054 [0.129]
2011-2015	-0.232*** [0.076]	-0.264*** [0.097]	-0.159 [0.126]
2016-2020	-0.327*** [0.106]	-0.512*** [0.154]	-0.118 [0.154]
Availability of care assistance			
Non-relatives	0.181 [0.111]	0.294** [0.133]	-0.041 [0.205]
Relatives	0.352*** [0.061]	0.316*** [0.078]	0.430*** [0.100]
Low assessment of financial situation	-0.09 [0.058]	-0.153** [0.075]	0.026 [0.092]
Fear of losing job	0.452***	0.445***	0.427***

	Model 1 (all cases)	Model 2 (cases of first birth)	Model 3 (cases of second and subsequent births)
	[0.058]	[0.073]	[0.096]
Lives with child's grandparents	0.140**	0.179**	0.093
	[0.066]	[0.084]	[0.115]
Quintile group of regions in terms of average wages (in comparable prices): omitted category – Group 3			
Group 1 (with the lowest wages)	-0.251**	-0.289**	-0.209
	[0.100]	[0.129]	[0.160]
Group 2	-0.213**	-0.276**	-0.11
	[0.093]	[0.117]	[0.158]
Group 4	-0.025	0.029	-0.1
	[0.092]	[0.117]	[0.153]
Group 5	0.023	-0.039	0.124
	[0.088]	[0.112]	[0.146]
Number of observations	1952	1201	751
Number of completed episodes	1356	836	520
Log-likelihood	-9137.065	-5207.907	-3015.683

Source: Authors' calculations based on RLMS data.

*Note: *significance at the level of 10%; **significance at the level of 5%; ***significance at the level of 1%. Standard errors are in parentheses.*

The results of our study are resistant to small changes in the underlying model, such as expanding the definition of employment (including self-employment), changing the definition of marital status and spouse/partner characteristics, using other regional characteristics, etc.

Our estimates have an important limitation. The RLMS data do not allow the estimation of the length of maternity leave in months. Our calculations were carried out in years, which certainly significantly reduces the quality of the estimate, especially considering that paid leave in Russia is one and a half years. It should also be noted that, in our efforts to avoid repeated inclusions, we consider strictly one birth for each woman. The values of most independent variables correspond to the end of the observation episode (a return to work or the last period of being in the sample). Because of this, we cannot distinguish, for example, a single mother from a woman whose marriage broke up after the birth of a child.

4. Discussion and conclusions

In this paper, we have tried to answer the question of what factors contribute to or hinder young mothers from returning to the workforce earlier. To a 20-year panel of a nationally representative study we applied duration analysis to assess a woman's chances of being out of the workforce due to childbirth.

The overall conclusion of our study generally confirms the results obtained earlier for other countries: women who have more diverse resources, both individual and family, return to professional work after the birth of a child faster. Higher levels of education, work experience, family and, to a lesser extent, non-relative assistance, living in a region with higher wages and experience of going to work after the birth of an older child will, all other things being equal, increase the likelihood of women returning to work.

Currently, Russia takes a non-standard approach to parental leave by international standards. In the aggregate, it lasts up to the age of three, which is noticeably longer than in most countries, but half of the leave is unpaid, though the mother's job is secure. The paid part of the leave, consisting of maternity leave and parental leave (140 + 475 days), is also noticeably larger than in most other countries. Due to objective reasons, maternity leave is less variable, and the main cross-country differences are concentrated in the duration of parental leave.

A separate problem is the unpaid part of parental leave. How important is it for Russian women to keep a job, if Russia has one of the lowest unemployment rates even at the peak of economic crises? As has been repeatedly shown, the Russian labor market adapts to crises by reducing real wages in exchange for keeping workers in their jobs (Kapelyushnikov 2022). In practice, the realization of women's right to keep their jobs often leads to discrimination by employers who are trying to get rid of employees who go on maternity leave or put pressure on them to shorten their leave (Center for Social and Labor Rights 2014; Kalabikhina 2017).

This measure is both an obvious burden for employers and at the same time not in demand by many women. As can be seen from our study, young mothers with work experience and a good education often do not use all of their leave, even the paid part, foregoing state support in order to continue working and maintain their professional skills and careers. Their rejection of the guarantees offered can be seen as a request for a more diversified leave policy that caters to the interests of different groups of women.

What can be proposed to improve the policy of granting parental leave? One possibility is a system of flexible leave, in which the woman and the household in which she lives independently determine the duration of the maternity leave, without losing their entitlements for a year and a half, which would be redistributed to the months they have selected. Establishing a minimum length of leave under this scheme would make it possible to bring Russia's experience closer to that of developed countries, where the leave is significantly shorter and the average salary is significantly higher. Secondly, a divided (intermittent) leave could be considered: repeated periods of leave over the allowable three years, if necessary, at the parents' choice. Under such a scheme, only the first part of the leave is paid, but the mother or father can go on leave again with the preservation of the workplace if necessary. Thirdly, thought could be given to shortening the leave to a year with a simultaneous increase in the amount of payment. Another modernization option is the introduction of non-transferable paternity leave, which would allow a combination of maternity (parental) and paternity leave, further contributing to a more equitable distribution of care for children within the household.

We understand that the proposed changes to the parental leave system may be ill received by the public. Ideally, the reduction of paid and unpaid maternity leave should be accompanied by an improvement in the whole system of family-work balance for different socio-demographic groups of parents. This includes the institution of certified nannies, guaranteed places in nursery groups, a variety of formats for these groups, three-party agreements on a friendly climate for parents with the participation of employers, and policies regarding care for the elderly and needy family members, since the time spent taking care of them often competes with childcare time.

In the recent past, the social dilemma associated with motherhood, as a rule, sounded like this: what is more important - fertility or female employment? Over time, these choices have ceased to be mutually exclusive. Flexible employment options and the expansion of childcare services allow women to reduce their absence from the workforce due to the birth of a child.

In the future, we plan to continue to study the impact of parental leave on women's labor activity. Of particular interest is the influence of the distribution of gender roles in the household on the length of maternity leave. Based on British data (Zhou, Kan 2019), it is shown that couples with a large number of children are less likely to adhere to the male breadwinner model, and families with a male breadwinner, in turn, are less likely to have more children. Thus, greater gender equality within the family has ceased to be an obstacle to fertility. This hypothesis could be tested on Russian data.

References

Arzhenovsky S.V., Artamonova D.V. (2007). Estimation of losses in the wages of women with children. *Applied Econometrics*, 3(7), 66-79.

Bobkov V.N., Kvachev V.G., Loktyukhina N.V., Richcheri M. (2017). Criteria, probability and degree of instability of employment, taking into account the peculiarities of the Russian labor market. *Economics of the Region*, 3, 672–683. https://doi.org/10.17059/2017-3-3

Ermolina A.A., Rokhmina E.B., Vasilyeva Yu.M., Trach T.M. (2016). The birth of a child as a factor in the income security of women. How big is the "maternity penalty" in Russia?

Demoscope Weekly, 701-702, 35-40. http://www.demoscope.ru/weekly/2016/0701/analit02.php (data uploaded 07/22/2022).

Kapelyushnikov R.I. (2022). Anatomy of the coronacrisis through the prism of the labor market. *Questions of Economics*, 2, 33-68. https://doi.org/10.32609/0042-8736-2022-2-33-68

Kolesnik D.P., Pestova A.A., Donina A.G. (2021). What to do with the employment of women with children in Russia? The role of preschool educational institutions. *Questions of Economics*, 12, 94-117. https://doi.org/10.32609/0042-8736-2021-12-94-117

Nivorozhkina L.I., Nivorozhkin A.M., Arzhenovsky S.V. (2008). The "price" of motherhood: an econometric assessment. *Bulletin of the Taganrog Institute of Management and Economics*, 1, 3-20.

Oshchepkov A.Yu. (2020). Fathers and children: "premium" for fatherhood in the Russian labor market. *Economic Journal of the Higher School of Economics*, 24(2), 157-190. https://doi.org/10.17323/1813-8691-2020-24-2-157-190

Payet A., Sinyavskaya O. (2010). Employment of women in France and Russia: the role of children and gender attitudes. In S.V. Zakharov, L.M. Prokofiev, O.V. Sinyavskaya (Ed.), *Family Evolution in Europe: East-West* (pp. 304–352). M.: NISP.

Center for Social and Labor Rights (2014). Discrimination, benefits and waiting lists for kindergartens are the main problems of working women. https://trudprava.ru/expert/analytics/genderanalyt/1049 (date of access: 08/03/2122)

Aisenbrey S., Evertsson M., Grunow D. (2009). Is there a career penalty for mothers' time out? A comparison of Germany, Sweden and the United States. *Social Forces*, 88(2), 573-605. https://doi.org/10.1353/sof.0.0252

Andersson G. (2008). Selectivity in higher-order childbearing in Sweden. *Finnish Yearbook of Population Research*, 33-40. https://doi.org/10.23979/fypr.45033

Bergemann A., Riphahn R.T. (2010). Female labour supply and parental leave benefits–the causal effect of paying higher transfers for a shorter period of time. *Applied Economics Letters,* 18(1), 17-20. https://doi.org/10.1080/13504850903425173

Biryukova S., Makarentseva A. (2017). Estimates of the Motherhood Penalty in Russia. *Population and Economics*, 1(1), 50-70. https://doi.org/10.3897/popecon.1.e36032

Bygren M., Duvander A.Z. (2006). Parents' workplace situation and fathers' parental leave use. *Journal of marriage and family*, 68(2), 363-372. https://doi.org/10.1111/j.1741-3737.2006.00258.x

Chirkova S. (2019). The impact of parental leave policy on child-rearing and employment behavior: The case of Germany. *IZA Journal of Labor Policy*, 9(1). https://doi.org/10.2478/izajolp-2019-0007

Cox D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Crompton R. (2006). Employment and the family: The reconfiguration of work and family life in contemporary societies. Cambridge University Press.

Evertsson M., Duvander A.Z. (2011). Parental Leave—Possibility or Trap? Does Family Leave Length Effect Swedish Women's Labour Market Opportunities? *European Sociological Review*, 27(4), 435-450. https://doi.org/10.1093/esr/jcq018

Grunow D., Aisenbrey S. (2016). Economic instability and mothers' employment: A comparison of Germany and the US. *Advances in Life Course Research*, 29, 5-15. https://doi.org/10.1016/j.alcr.2015.09.005

Hobson B., Duvander A.Z., Halldén K. (2006). Men and women's agency and capabilities to create a work-life balance in diverse and changing institutional contexts. *Children, changing families and welfare states*, 13, 267-296. https://doi.org/10.4337/9781847204363.00024

Joseph O., Pailhé A., Recotillet I., Solaz A. (2013). The economic impact of taking short parental leave: Evaluation of a French reform. *Labour Economics*, 25, 63-75. https://doi.org/10.1016/j.labeco.2013.04.012

Kalabikhina I. (2017). Parental responsibilities and discrimination in employment. *Population and Economics*, 1(1), 89-116. https://doi.org/10.3897/popecon.1.e36034

Klein J.P., Moeschberger M.L. (2003). *Survival analysis: techniques for censored and truncated data* (Vol. 1230). New York: Springer.

Kluve J., Tamm M. (2013). Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: evidence from a natural experiment. *Journal of Population Economics*, 26(3), 983-1005. https://doi.org/10.1007/s00148-012-0404-1

Kuhlenkasper T., Kauermann G. (2010). Duration of maternity leave in Germany: A case study of nonparametric hazard models and penalized splines. *Labour Economics*, 17(3), 466-473. https://doi.org/10.1016/j.labeco.2009.12.001

Lappegard T. (2008). Changing the gender balance in caring: Fatherhood and the division of parental leave in Norway. *Population research and policy review*, 27(2), 139-159. https://doi.org/10.1007/s11113-007-9057-2

Lapuerta I., Baizán P., González M.J. (2011). Individual and institutional constraints: An analysis of parental leave use and duration in Spain. *Population Research and Policy Review*, 30(2), 185-210. https://doi.org/10.1007/s11113-010-9185-y

Lundberg S., Pollak R.A. (1993). Separate spheres bargaining and the marriage market. *Journal of political Economy*, 101(6), 988-1010. https://doi.org/10.1086/261912

Macran S., Joshi H., Dex S. (1996). Employment after childbearing: a survival analysis. *Work, Employment and Society*, 10(2), 273-296. https://doi.org/10.1177%2F0950017096102004

Makay Z. (2017). Parental Leave and Career Interruption of Mothers in France and Hungary. In: Régnier-Loilier A. (Ed.), *A Longitudinal Approach to Family Trajectories in France*. INED Population Studies, vol 7. Springer, Cham. https://doi.org/10.1007/978-3-319-56001-4_9

Morosow K., Jalovaara M. (2019). Disadvantaging single parents? Effects of long family leaves on single and partnered mothers' labour market outcomes in Finland. Retrieved from https://www.diva-portal.org/smash/record.jsf?pid=diva2:1341229

Pettit B., Hook J. (2005). The structure of women's employment in comparative perspective. *Social Forces*, 84(2), 779-801. https://doi.org/10.1353/sof.2006.0029

Sundström M., Duvander A.Z. (2002). Gender division of childcare and the sharing of parental leave among new parents in Sweden. *European Sociological Review*, 18(4), 433-447. https://doi.org/10.1093/esr/18.4.433

Thévenon O., Solaz A. (2013). Labour Market Effects of Parental Leave Policies in OECD Countries. *OECD Social, Employment and Migration Working Papers*, 141, 1-68. https://doi.org/10.1787/5k8xb6hw1wjf-en

Zhou M., Kan M.Y. (2019). A new family equilibrium? Changing dynamics between the gender division of labor and fertility in Great Britain, 1991–2017. *Demographic Research*, 40, 1455-1500. https://doi.org/10.4054/DemRes.2019.40.50

Демографическое обозрение / Demographic Review
2022;9(4) 61-103 DOI: https://doi.org/10.17323/demreview.v9i4.16744

Assessment of reliability of results of the 2019 Belarus population census based on an analysis of changes in the ethnolinguistic composition of the population

Aleksandr S. Sokolov
(alsokol@tut.by), F. Skorina Gomel State University, Belarus.

Abstract: *The article examines the dynamics of the ethnolinguistic composition of the population of Belarus in the period between the 2009 and 2019 censuses. It is shown that for a number of indicators (the population of certain nationalities by region, the share of Belarusian, Russian and ethnic languages as mother and home languages) there is an anomalous dynamics, which is expressed, first of all, in multidirectional and non-uniformly scaled changes in these indicators in different regions, which was not observed in previous censuses, an abnormally sharp increase in the share of the Belarusian language and some other characteristics in certain regions, the manifestation of a number of features almost exclusively within the administrative boundaries of certain regions. As a result, analysis of the complex of anomalous results of the 2019 population census suggests that these results are due to a distortion carried out in order to obtain politically desired results: despite the very sharp and opposite dynamics of various indicators across the regions, changes on a national scale turned out to be very insignificant compared to 2009 and correspond to ideological attitudes - the shares of both the Russian and Belarusian languages, both native and home languages, increased by a small amount among the general population and major nationalities. Thus, the official results of the census made it possible to preserve the symbolic status of the Belarusian language as an element of ethno-consolidation and national identification, although in fact the indicators of the use of the Belarusian language continue the downward trend expressed since 1999, and the majority of the population considers Russian as their native and home language.*

Keywords: *Belarus, Russian language, Belarusian language, mother tongue, home language, Russians, Belarusians, Poles, Ukrainians, Tatars, Jews, Lithuanians, 2019 population census.*

For citation: *Sokolov A. (2022). Assessment of reliability of results of the 2019 Belarus population census based on an analysis of changes in the ethnolinguistic composition of population. Demographic Review, 9(4), 61-103. https://doi.org/10.17323/demreview.v9i4.16744*

Introduction

The significance of population censuses and their results has always, especially in the Soviet and post-Soviet space, gone far beyond a simple display of demographic reality and a tool for a purely scientific analysis of demographic processes and their forecasting. According to V.N. Rakachev and A.A. Khalafyan (2012: 150), *"the problem of the completeness and reliability of statistical data has always existed, regardless of the historical period and political regime ... The uniqueness of our country is manifested in the fact that for several decades demographic statistics have been considered an important ideological tool that can influence the consciousness of the masses... Various kinds of fake statistics have been used at various levels: from the ordinary accountant and bookkeeper to such large-scale actions as the All-Union Population Census."*

Data on the ethnic and linguistic structure of the population in the new post-Soviet independent states are attracting increased socio-political interest, as they are designed, among other things, to reflect success in strengthening "national identity" and minimizing any "Russian influence", the main marker of which is a reduction in the spread of the Russian language.

The publication of the final data of the population census of Belarus in 2019 gave reason to suspect a deliberate distortion of a number of its results (Sokolov 2021; Manakov, Sokolov, Suvorkov 2022), in particular, an overstatement of the share of the Belarusian language as both native and home language and, accordingly, an understatement of the share of the Russian language.

Table 1. The share of individual nationalities in the total population of Belarus and regions (according to the 2019 census), %

Nationality	Belarus	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast
Belarusians	84.89	86.92	82.32	87.23	68.29	86.86	88.55	89.35
Russians	7.51	7.26	12.16	7.83	6.38	7.34	5.87	6.07
Poles	3.06	1.10	0.86	0.19	21.73	0.96	1.07	0.21
Ukrainians	1.70	2.79	1.58	1.81	1.05	1.72	1.45	1.19
Jews	0.15	0.07	0.14	0.14	0.09	0.28	0.10	0.12
Tatars	0.09	0.06	0.09	0.07	0.13	0.11	0.08	0.08
Lithuanians	0.06	0.03	0.08	0.03	0.21	0.03	0.03	0.02

Source: (Belstat 2020).

This article is intended to continue research in this direction, to identify and systematize indicators that could be the result of deliberate distortions. The article deals with data on the ethnolinguistic structure of the population. The object of the study are the 5 largest nationalities (Belarusians, Russians, Poles, Ukrainians, Jews), as well as 2 nationalities with a long history in the territory of Belarus (Tatars, Lithuanians). The shares of these nationalities in the total population of Belarus and its regions are shown in Table 1.

In accordance with the Constitution and the Law "On Languages in the Republic of Belarus", the official state languages in the country are Belarusian and Russian, thus ensuring their full development and functioning in all spheres of public life. In addition, citizens are guaranteed the right to use their national languages.

Formulation of the problem

In total, 3 censuses have been conducted in independent Belarus - in 1999, 2009 and 2019. The results of the last census in 2019 turned out to be quite paradoxical:

- for the first time in the history of the census, for a large number of indicators characterizing the country's ethnic and linguistic structure, opposite dynamics were observed in different regions;
- a number of phenomena appeared almost exclusively within the administrative boundaries of individual regions;
- very sharp changes in one direction of individual indicators in one or two regions ensured a change in the same direction of these indicators in the country as a whole, despite the fact that in all other regions these changes were in the opposite direction;
- significant differences in the direction and magnitude of changes in language indicators for individual regions and nationalities are accompanied by a slight change in these indicators for the republic as a whole;
- some indicators showed huge growth (by hundreds and thousands of percent) in the absence of any prerequisites and factors capable of explaining such significant changes, etc.

As a result, all the above-listed peculiarities of the 2019 census, which were absent in previous censuses, provided Belarus as a whole with results on issues of heightened public interest that can be called ideal from a political point of view:

- for both state languages (Russian and Belarusian) there was an increase in the share of the population who indicated them as their native and home languages;
- an increase in the share of the Russian and Belarusian languages, both native and home, was noted for all major nationalities - Belarusians, Russians, Ukrainians;
- while the total population decreased, the absolute number of Belarusians increased;
- the proportion of Belarusian as a native language was not allowed to decrease to less than 50 per cent, nor the proportion of Russian as a native language to exceed that of Belarusian, despite a marked trend in 1999-2009 of a declining share of Belarusian as native and home language (table 2) in all regions of Belarus.

Table 2. Change in the share of Russian and Belarusian as native and home language according to 1999-2019 census data, %

Language	1999	2009	Share change in 1999-2009	2019	Share change in 2009-2019
Whole population					
Russian as native	24.1	41.5	+17.4	42.3	+0.8
Belarussian as native	73.7	53.2	–20.5	54.1	+0.9
Russian as home	36.7	70.2	+33.5	71.4	+1.2
Belarussian as home	62.8	23.4	–39.4	26.0	+2.6
Belarussians					
Russian as native	14.3	37.0	+22.7	38.1	+1.1
Belarussian as native	85.6	60.8	–24.8	61.2	+0.4
Russian as home	58.6	69.8	+11.2	71.0	+1.2
Belarussian as home	41.3	26.1	–15.2	28.5	+2.4

Sources: (Belstat 2020; Minstat 2001: 150).

Such politically ideal results make it possible, on their basis, to issue statements about the absence of both Russification and de-Russification, the preservation of the status of the Belarusian language as an important value for the Belarusian people, and the equality of state languages with a small symbolic priority of Belarusian as “titular” and “national”, and exclude the possibility of using the results of the census for propaganda by both nationalist and conventionally “anti-nationalist” political forces.

However, in our opinion, based on the analysis of the census results for individual nationalities and for different levels of administrative division, such a result became possible only thanks to targeted administrative influence on the census data in order to obtain a politically “convenient” result.

Many researchers emphasize that the Russian language in Belarus dominates as a means of communication, while Belarusian performs mainly symbolic and ethno-consolidating functions (Mechkovskaya 2003: 10; Goritskaya 2021: 61, 233). A decrease in the share of Belarusian as a native language below that of Russian in the entire population would call into question even its symbolic function in society, and among ethnic Belarusians, its role as an element of national identification.

The 2019 census took place during the period of the so-called “soft Belarusianization” (Goritskaya 2021: 15; Bakhlova, Bakhlov 2020: 745; Linchenko, Belyaeva 2020: 105; Shimov 2020: 69; Anarbek 2019: 109), a set of measures to increase the symbolic status of the Belarusian language and expand its presence in public life, but without imposing any open restrictions in relation to Russian. The main goal of this policy was to emphasize the identity of Belarusian society and its fundamental differences from Russian society, including by promoting appropriate narratives and symbols. And a census, as a number of authors point out (Shimov 2020: 69; Goritskaya 2021: 17), is not only a means of reflecting the existing situation, but also a tool for constructing social, in particular sociolinguistic, reality, and can reflect the ideological attitudes and demands of the state, of its individual institutions and even census takers. In this regard, the official results of the census showed full compliance with the ideological demands of this period.

Change in the population size and linguistic structure of Belarusians and Russians

The indigenous nationalities of Belarus - Belarusians and Russians - differ from the rest in that their native and home languages are almost exclusively Belarusian and Russian. For other nationalities, a third component appears in the linguistic structure - the ethnic language.

Changes in the shares of Belarusian and Russian as native and home languages from 2009 to 2019 are clearly related to administrative divisions. Tables 3-8 present data showing changes in the shares of the Russian, Belarusian and ethnic languages for the population of Belarus and individual nationalities by region. Among Belarusians, an increase in the share of Belarusian as a native language, and a very significant one, was recorded only in the Brest oblast and the city of Minsk, i.e., in the regions where this share was the lowest in 2009 (Table 3). For Belarusian as a home language (Table 6), the situation is similar - an increase in the share of the population was recorded only in two regions, the Mogilev oblast and the city of Minsk, where also in 2009 this indicator was the lowest among all regions. Thus, the Brest oblast experienced a sharp increase, the highest among all regions, in the share of Belarusian as a native language, with a simultaneous decrease in its share as a home language, also the highest among all regions. In the Mogilev oblast, on the contrary, the share of Belarusians calling Belarusian their native language decreased and in 2019 became the lowest among all regions, while the share of those calling it their home language increased. In Minsk, both indicators increased, and the share of Belarusians calling Belarusian their home language increased by 5.5 times, becoming comparable to the share of this indicator in the Grodno and Minsk oblasts, which traditionally have a higher share of Belarusian as both their native and home language.

Table 3. Change in the share of Belarusian as a native language for the population of Belarus and individual nationalities, 2009-2019, %

Nationality and census year		Belarus	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast	σ*	Me**
	2009	53.2	53.7	52.5	54.6	59.2	35.2	69.4	55.1	9.4	54.6
Whole population	2019	54.1	77.0	45.3	47.0	54.5	48.7	59.5	46.1	10.5	48.7
	+/–	+0.9	+23.3	–7.2	–7.6	–4.7	+13.5	–9.9	–9.0		
	2009	60.8	59.6	60.6	61.3	68.8	43.5	76.7	61.6	9.4	61.3
Belarussians	2019	61.2	86.2	53.9	53.5	61.4	55.2	66.2	51.3	11.3	55.2
	+/–	+0.4	+26.6	–6.7	–7.8	–7.4	+11.7	–10.5	–10.3		
	2009	2.8	2.6	2.8	2.5	4.9	2.0	3.1	2.6	0.9	2.6
Russians	2019	2.9	11.3	1.0	1.0	3.0	1.3	2.2	1.5	3.4	1.5
	+/–	+0.1	+8.7	–1.8	–1.5	–1.9	–0.7	–0.9	–1.1		
	2009	58.2	55.2	55.6	43.9	58.0	40.9	81.3	33.6	14.4	55.2
Poles	2019	54.5	43.8	43.7	45.7	56.1	43.2	65.1	43.0	8.0	43.8
	+/–	–3.7	–11.4	–11.9	+1.8	–1.9	+2.3	–16.2	+9.4		
	2009	7.9	7.9	5.8	9.1	11.2	5.4	8.8	7.5	1.8	7.9
Ukrainians	2019	8.2	23.3	13.6	2.3	3.9	0.8	2.0	1.7	7.9	2.3
	+/–	+0.3	+15.4	+7.8	–6.8	–7.3	–4.6	–6.8	–5.8		
	2009	9.1	11.6	8.7	11.2	8.6	6.8	16.9	9.7	3.0	9.7
Jews	2019	11.5	27.6	36.0	18.4	3.3	4.6	3.7	4.0	12.5	4.6
	+/–	+2.4	+16.0	+27.3	+7.2	–5.3	–2.2	–13.2	–5.7		
	2009	19.3	7.7	13.6	2.4	43.0	12.9	22.0	3.0	13.1	12.9
Tatars	2019	21.2	36.7	46.6	30.7	33.8	6.2	8.4	0.8	16.5	30.7
	+/–	+1.9	+29.0	+33.0	+28.3	–9.2	–6.7	–13.6	–2.2		
	2009	25.8	15.8	11.9	13.3	33.1	32.8	17.6	3.0	10.3	15.8
Lithuanians	2019	46.1	51.3	62.0	61.3	42.1	36.3	34.8	0.8	19.4	42.1
	+/–	+20.3	+35.5	+50.1	+48.0	+9.0	+3.5	+17.2	–2.2		

*Note: * – standard deviation; ** – median.*

Source: Author's calculations based on Belstat data.

As can be seen in Table 3, for the share of Belarusian as a native language for the entire population and for all major nationalities in 2019, there is a significant decrease in the median values of this indicator by oblasts, while for Belarus as a whole, this share increases for almost all nationalities. As a result, while in 2009 the median was roughly in line with the whole country (as well as the arithmetic average across regions), in 2019 it was much lower. Also, for almost all nationalities, the standard deviation indicator increases, which characterizes the degree of dispersion of values relative to the mean.

Table 4. Change in the share of Russian as a native language for the population of Belarus and individual nationalities, 2009-2019, %

Nationality and census year		Belarus	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast	σ	Me
Whole population	2009	41.5	42.6	44.2	41.8	36.1	52.6	27.4	41.9	7.1	41.9
	2019	42.3	20.1	51.6	50.0	41.7	48.6	37.3	46.6	10.2	46.6
	+/–	+0.8	–22.5	+7.4	+8.2	+5.6	–4.0	+9.9	+4.7		
Belarussians	2009	37.0	38.8	38.4	36.8	29.5	51.4	21.7	37.2	8.4	37.2
	2019	38.1	13.5	46.0	46.3	38.5	44.6	33.7	44.4	10.9	44.4
	+/–	+1.1	–25.3	+7.6	+9.5	+9.0	–6.8	+12.0	+7.2		
Russians	2009	96.3	96.7	96.7	96.7	94.3	96.4	96.2	96.8	0.8	96.7
	2019	96.8	88.1	98.4	98.5	96.6	98.5	97.7	98.3	3.5	98.3
	+/–	+0.5	–8.6	+1.7	+1.8	+2.3	+2.1	+1.5	+1.5		
Poles	2009	33.9	37.7	32.5	46.8	34.1	47.4	15.0	56.1	12.3	37.7
	2019	38.5	49.5	45.4	43.2	37.0	50.2	29.0	44.9	6.9	44.9
	+/–	+4.6	+11.8	+12.9	–3.6	+2.9	+2.8	+14.0	–11.2		
Ukrainians	2009	61.2	51.5	66.7	62.5	64.0	67.9	60.8	65.0	5.1	64
	2019	62.4	41.4	63.9	71.2	69.0	83.1	68.7	31.2	16.9	68.7
	+/–	+1.2	–10.1	–2.8	+8.7	+5.0	+15.2	+7.9	–33.8		
Jews	2009	86.1	81.4	87.9	83.9	86.6	88.0	77.7	85.5	3.5	85.5
	2019	83.6	38.3	62.2	79.9	95.2	93.6	91.7	87.2	19.4	87.2
	+/–	–2.5	–43.1	–25.7	–4.0	+8.6	+5.6	+14.0	+1.7		
Tatars	2009	63.8	73.5	67.5	75.8	48.5	67.8	61.7	69.9	8.4	67.8
	2019	62.0	46.8	34.2	54.9	56.0	80.9	72.9	63.1	14.5	56.0
	+/–	–1.8	–26.7	–33.3	–20.9	+7.5	+13.1	+11.2	–6.8		
Lithuanians	2009	39.3	51.5	48.6	55.0	28.6	22.4	48.9	69.9	14.9	48.9
	2019	34.7	31.8	26.5	26.0	36.0	21.1	41.8	63.1	13.1	31.8
	+/–	–4.6	–19.7	–22.1	–29.0	+7.4	–1.3	–7.1	–6.8		

Source: Author's calculations based on Belstat data.

Table 5. Change in the share of ethnic languages as native for the population of Belarus and individual nationalities, 2009-2019, %

Nationality and census year		Belarus	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast	σ	Me
Poles	2009	5.4	5.6	10.7	5.7	5.2	6.3	2.8	6.9	2.2	5.7
	2019	6.7	6.4	10.5	9.0	6.6	6.0	5.0	6.9	1.8	6.6
	+/–	+1.3	+0.8	–0.2	+3.3	+1.4	–0.3	+2.2	0.0		
Ukrainians	2009	29.2	39.3	26.5	26.7	23.4	23.8	28.7	26.2	5.0	26.5
	2019	29.1	35.2	21.8	26.3	26.9	15.9	28.8	66.3	15.2	26.9
	+/–	–0.1	–4.1	–4.7	–0.4	+3.5	–7.9	+0.1	+40.1		
Jews	2009	1.9	2.5	1.6	1.7	2.8	1.8	2.3	2.5	0.4	2.3
	2019	3.2	32.6	0.6	0.7	0.2	1.2	0.8	2.3	11.1	0.8
	+/–	+1.3	+30.1	–1.0	–1.0	–2.6	–0.6	–1.5	–0.2		
Tatars	2009	14.5	16.8	16.9	18.8	6.5	15.3	14.7	25.5	5.2	16.8
	2019	13.8	14.9	17.2	13.6	9.0	11.9	13.0	23.5	4.3	13.6
	+/–	–0.7	–1.9	+0.3	–5.2	+2.5	–3.4	–1.7	–2.0		
Lithuanians	2009	31.4	28.2	36.1	28.8	35.6	39.9	30.0	25.5	4.8	30.0
	2019	17.2	14.9	10.1	11.3	19.9	41.6	19.5	23.5	9.8	19.5
	+/–	–14.2	–13.3	–26.0	–17.5	–15.7	+1.7	–10.5	–2.0		

Source: Author's calculations based on Belstat data.

Figure 1 shows the dynamics of the share of Belarusians calling Belarusian their native language in 1979-2019 and their home language in 1999-2019. (The question about home language first appeared only in the 1999 census program). It can be seen that, up to 2009, the changes in these indicators across regions were generally synchronous. At the same time, in Minsk, indicators for the Belarusian language have always been significantly lower than in other regions, even in 1999, when, on the wave of accelerated Belarusianization in the first years of independence, the share of the Belarusian language in all regions slightly increased. In 2019, for the first time, changes in different regions became multidirectional, and the share of the Belarusian language increased very sharply precisely in those regions that were never distinguished by its high values.

Table 6. Change in the share of Belarusian as home language for the population of Belarus and individual nationalities, 2009-2019, %

Nationality and census year		Belarus	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast	σ	Me
Whole population	2009	23.4	26.7	22.4	22.7	35.1	5.8	38.9	19.6	10.0	22.7
	2019	26.0	16.2	12.6	14.5	37.9	34.1	37.4	25.1	10.3	25.1
	+/–	+2.6	–10.5	–9.8	-8.2	+2.8	+28.3	–1.5	+5.5		
Belarusians	2009	26.1	29.4	25.4	25.3	38.4	7.0	42.5	21.8	10.7	25.4
	2019	28.5	16.6	14.7	16.4	39.4	38.5	41.3	27.9	11.1	27.9
	+/–	+2.4	–12.8	–10.7	–8.9	+1.0	+31.5	–1.2	+6.1		
Russians	2009	2.1	1.8	2.5	1.8	4.4	0.6	3.3	1.9	1.1	1.9
	2019	2.5	11.1	0.9	0.7	2.4	0.7	2.1	1.1	3.5	1.1
	+/–	+0.4	+9.3	–1.6	–1.1	–2.0	+0.1	–1.2	–0.8		
Poles	2009	40.9	36.9	48.2	23.7	40.7	12.5	67.9	12.5	18.6	36.9
	2019	46.0	28.7	37.4	40.2	48.9	26.2	53.4	31.8	9.5	37.4
	+/–	+5.1	–8.2	–10.8	+16.5	+8.2	+13.7	–14.5	+19.3		
Ukrainians	2009	6.1	7.8	4.8	6.5	9.2	1.5	8.3	4.5	2.5	6.5
	2019	6.4	20.2	1.8	2.8	5.6	0.8	3.1	1.6	6.3	2.8
	+/–	+0.3	+12.4	–3.0	–3.7	–3.6	–0.7	–5.2	–2.9		
Jews	2009	2.0	2.3	1.7	2.5	3.7	1.3	4.6	2.1	1.1	2.3
	2019	2.1	6.6	1.2	1.0	1.3	2.2	2.1	1.9	1.8	1.9
	+/–	+0.1	+4.3	–0.5	–1.5	–2.4	+0.9	–2.5	–0.2		
Tatars	2009	13.9	17.8	3.6	2.6	8.1	2.8	7.2	1.5	5.2	3.6
	2019	13.0	29.8	0.6	0.6	3.5	0.4	6.8	0.4	10.0	0.6
	+/–	–0.9	+12.0	–3.0	–2.0	–4.6	–2.4	–0.4	-1.1		
Lithuanians	2009	24.4	11.0	13.5	6.3	38.1	18.8	16.2	10.3	9.7	13.5
	2019	26.5	28.2	23.2	21.6	27.3	29.4	26.8	29.0	2.8	27.3
	+/–	+2.1	+17.2	+9.7	+15.3	–10.8	+10.6	+10.6	+18.7		

Source: Author's calculations based on Belstat data.

Table 7. Change in the share of Russian as home language for the population of Belarus and individual nationalities, 2009-2019, %

Nationality and census year		Belarus	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk	Minsk oblast	Mogilev oblast	σ	Me
Whole population	2009	70.2	70.1	73.2	72.0	56.5	82.1	56.0	76.5	9.1	72.0
	2019	71.4	81.9	85.0	83.3	60.1	64.0	60.1	68.2	10.4	68.2
	+/–	+1.2	+11.8	+11.8	+11.3	+3.6	–18.1	+4.1	–8.3		
Belarusians	2009	69.8	68.4	71.7	70.1	54.3	87.3	53.2	75.4	11.0	70.1
	2019	71.0	83.2	85.3	83.6	60.5	61.5	58.6	67.7	11.2	67.7
	+/–	+1.2	+14.8	+13.6	+13.5	+6.2	–25.8	+5.4	–7.7		
Russians	2009	96.5	97.4	96.4	96.7	93.3	97.9	95.2	97.1	1.5	96.7
	2019	97.2	88.8	99.1	99.3	97.5	99.3	97.9	96.1	3.5	97.9
	+/–	+0.7	–8.6	+2.7	+2.6	+4.2	+1.4	+2.7	–1.0		
Poles	2009	50.9	59.6	46.3	71.4	50.0	79.8	29.2	84.0	18.3	59.6
	2019	52.4	68.9	60.6	58.2	49.7	71.3	44.5	60.9	8.9	60.6
	+/–	+1.5	+9.3	+14.3	–13.2	–0.3	–8.5	+15.3	–23.1		
Ukrainians	2009	88.4	82.8	91.6	88.6	85.9	94.6	87.2	92.4	3.8	88.6
	2019	89.1	67.9	97.3	95.2	92.4	98.0	94.9	91.3	9.7	94.9
	+/–	+0.7	–14.9	+5.7	+6.6	+6.5	+3.4	+7.7	–1.1		
Jews	2009	95.9	93.9	96.7	95.3	94.2	96.4	94.3	96.6	1.1	95.3
	2019	96.6	92.9	98.5	98.4	98.5	97.2	97.9	87.8	3.8	97.9
	+/–	+0.7	–1.0	+1.8	+3.1	+4.3	+0.8	+3.6	–8.8		
Tatars	2009	83.5	92.4	86	95.7	63.3	92.6	80.3	96.6	10.9	92.4
	2019	85.5	64.3	72.7	99.2	71.6	97.1	92.9	85.7	12.8	85.7
	+/–	+2.0	–28.1	–13.3	+3.5	+8.3	+4.5	+12.6	–10.9		
Lithuanians	2009	65.5	87.3	82.5	89.3	45.1	72.9	80.7	82.8	14.0	82.5
	2019	68.1	48.2	76.6	78.1	66.6	68.5	71.0	62.0	9.3	68.5
	+/–	+2.6	–39.1	–5.9	–11.2	+21.5	–4.4	–9.7	–20.8		

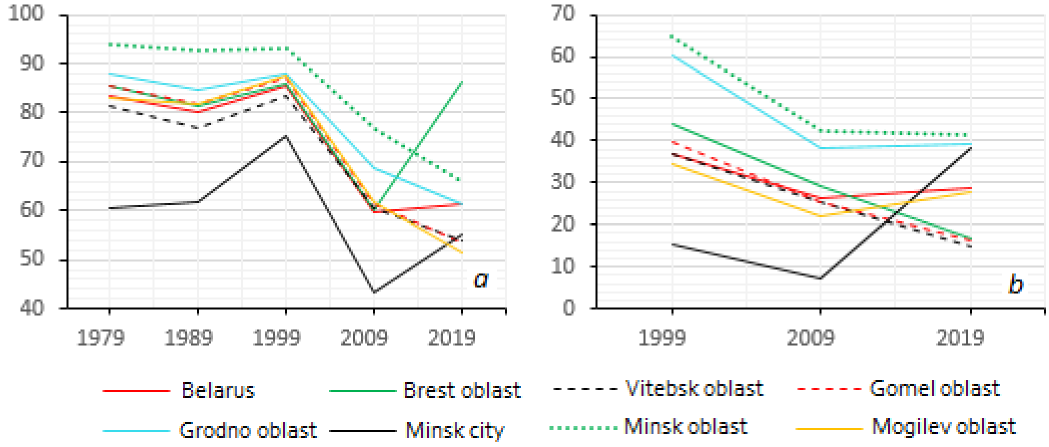
Source: Author's calculations based on Belstat data.

Table 8. Change in the share of ethnic languages as home language for the population of Belarus and individual nationalities, 2009-2019, %

Nationality and census year		Belarus	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast	σ	Me
Poles	2009	1.3	3.2	6.3	2.6	3.3	4.7	0.5	3.2	1.7	3.2
	2019	1.2	7.8	5.3	2.5	2.4	4.8	1.4	1.8	2.2	2.5
	+/–	–0.1	+4.6	–1.0	–0.1	–0.9	+0.1	+0.9	–1.4		
Ukrainians	2009	3.5	7.7	2.1	2.6	1.8	1.6	2.4	2.0	2.0	2.1
	2019	4.0	11.8	0.9	2.0	2.1	1.1	2.0	1.3	3.6	2.0
	+/–	+0.5	+4.1	–1.2	–0.6	+0.3	–0.5	–0.4	–0.7		
Jews	2009	0.8	0.7	0.1	0.3	0.9	0.2	0.3	0.5	0.3	0.3
	2019	0.5	0.4	0.1	0.2	0.1	0.3	0.0	0.1	0.1	0.1
	+/–	–0.3	–0.3	0.0	–0.1	–0.8	+0.1	–0.3	–0.4		
Tatars	2009	0.5	0.1	0.1	1.0	0.6	0.1	0.8	1.4	0.5	0.6
	2019	0.7	0.1	0.0	0.1	0.1	0.1	0.1	6.5	2.2	0.1
	+/–	+0.2	0.0	–0.1	–0.9	–0.5	0.0	–0.7	+5.1		
Lithuanians	2009	5.4	0.6	1.6	0.4	10.9	1.9	0.8	2.6	3.4	1.6
	2019	4.6	22.4	0.2	0.3	5.4	2.1	0.6	2.0	7.4	2.0
	+/–	–0.8	+21.8	-1.4	-0.1	-5.5	+0.2	-0.2	-0.6		

Source: Author's calculations based on Belstat data.

Figure 1. Dynamics of the share of Belarusians calling Belarusian their native (a) and home (b) language, %



Source: Author's calculations based on Belstat data.

At the same time, many sociological studies paint a completely different picture of the spread of the Belarusian language in these regions. Thus, a study conducted simultaneously with the population census in November 2019 showed that in the Mogilev oblast the share of the population using the Belarusian language in everyday life does not exceed 3% (Khamutovskaya 2021: 208). Moreover, in this study, as in the census, respondents were told to choose only one language and there were no such options as “two languages”, “trasyanka”, etc., the presence of which in sociological surveys often explains significant discrepancies between the census results (where such options are absent) and opinion polls.

Also, the results of numerous studies (Hentschel, Kittel 2011: 69; Khentschel 2017: 232) state that in large cities the role of the Belarusian language is smaller than in the whole country, and in Minsk the share of people using it is minimal. In any case, there was not a single study that recorded at least some increase in the use of the Belarusian language by the population of Minsk until the results of the 2019 census appeared, according to which more than a third of the population called Belarusian their home language (34.1% vs. 5.8% in 2009), i.e., far more people than in any other large city of Belarus.

It is interesting to analyze the shares of those calling Belarusian their native and home language among the population that did not indicate nationality in the census form (Table 9). These shares in all regions are noticeably declining, and only in the city of Minsk are they increasing, and many times over, especially the share of Belarusian as a home language, which has increased almost 200 times (from 0.1 to 19.3%), while the value in other oblasts is no more than 0.5%.

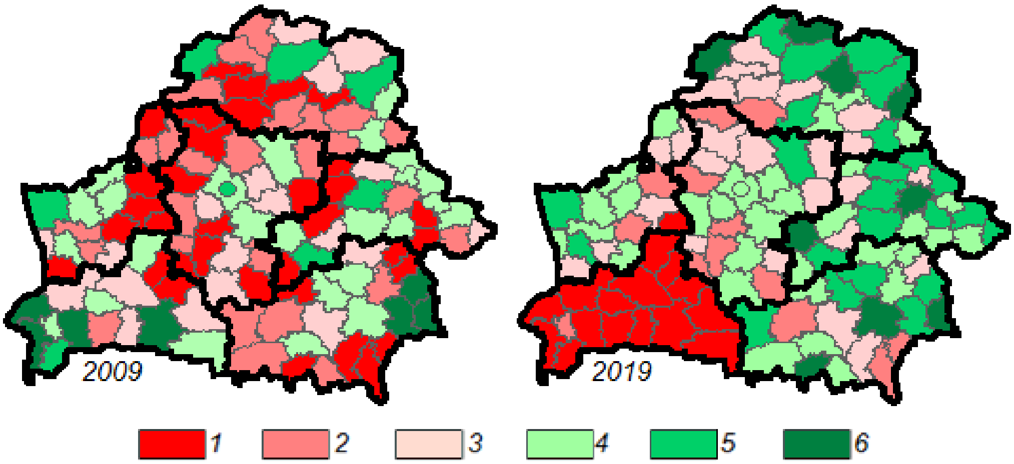
If we look at the change in the share of Belarusian as a native language for Belarusians in terms of administrative regions and types of locality (Figures 2-4), then there is a clear localization of high values of this indicator in 2019 and its increase in 2009-2019 within the administrative boundaries of the Brest oblast for both the urban and rural population. Of the 16 districts of the Brest oblast, only two for the urban population and three for the rural population did not experience an increase in the share of Belarusian as a native language by more than 5%. Of all the other regions of the country, on the contrary, such an increase occurred only in 4 out of 102 regions.

Table 9. Belarusian as native and home language for persons not indicating nationality in the census form

	Number of persons not indicating nationality, persons		Share of persons calling Belarussian native language, %		Share of persons calling Belarussian home language, %	
	2009	2019	2009	2019	2009	2019
Brest oblast	12 883	18 233	4.5	1.2	2.1	0.3
Vitebsk oblast	21 040	24 315	3.1	0.6	1.4	0.2
Gomel oblast	13 292	29 494	3.9	1.4	1.7	0.4
Grodno oblast	14 078	16 248	5.0	1.5	2.6	0.4
Minsk city	134 047	38 810	0.6	4.4	0.1	19.3
Minsk oblast	16 058	34 591	6.3	2.2	4.7	0.5
Mogilev oblast	15 131	26 063	2.7	2.1	1.0	0.5
Belarus	226 529	187 754	2.1	2.2	1.0	4.3

Source: Author's calculations based on Belstat data.

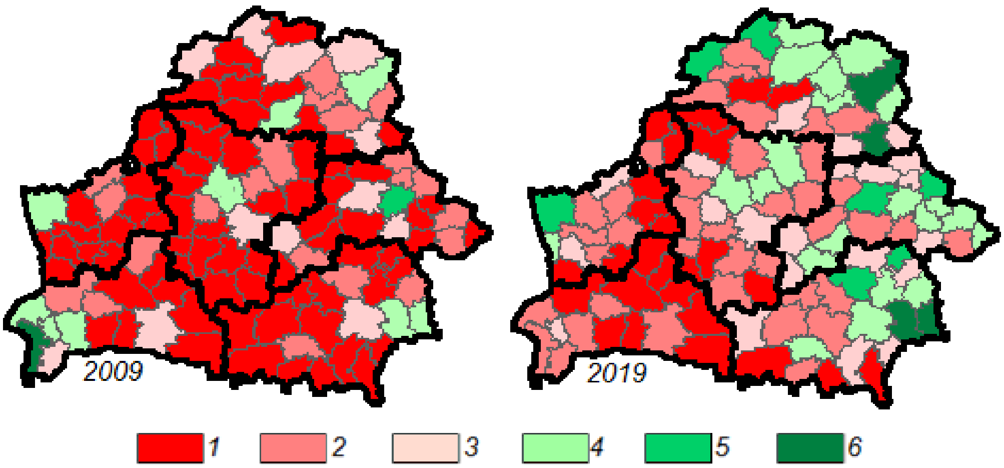
Figure 2. Percentage of Belarusians calling Belarusian their native language (urban population), 2009 and 2019, %



1 - more than 80; 2 - 70-80; 3 - 60-70; 4 -50-60; 5 -40-50; 6 - less than 40

Source: Author's calculations based on Belstat data.

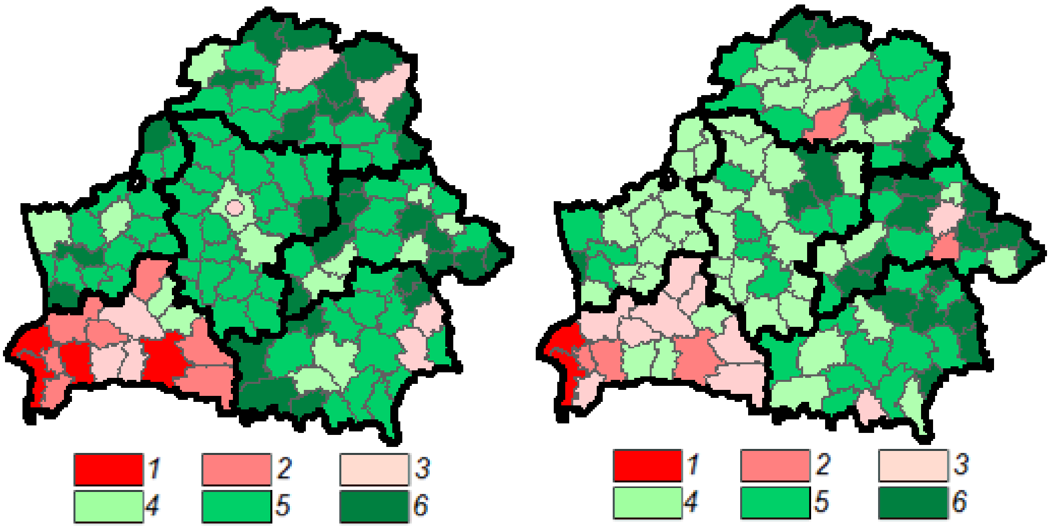
Figure 3. Percentage of Belarusians calling Belarusian their native language (rural population), 2009 and 2019, %



1 - more than 80; 2 - 70-80; 3 - 60-70; 4 -50-60; 5 - 40-50; 6 - less than 50

Source: Author's calculations based on Belstat data.

Figure 4. Change in the share of Belarusians calling Belarusian their native language among the urban (a) and rural (b) population, 2009-2019, %



a

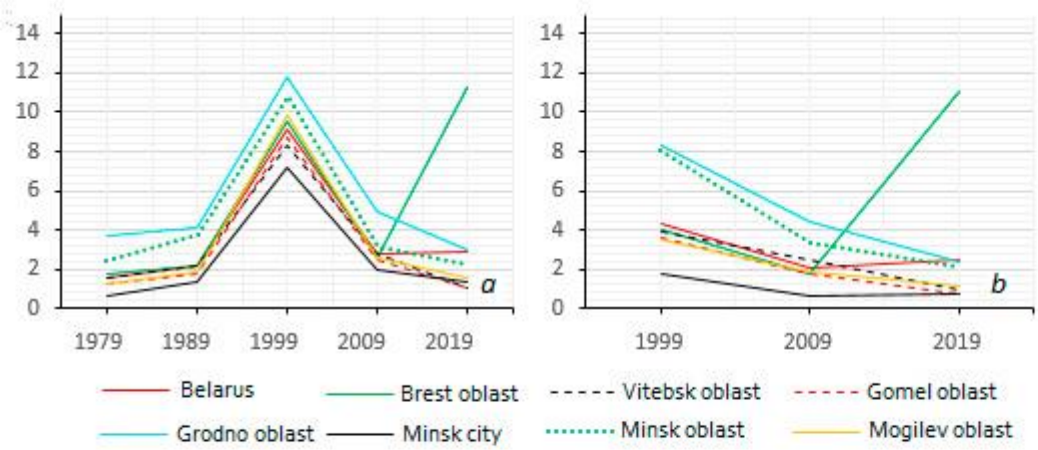
1 - more than +40; 2 - from +20 to +40;
3 - from +5 to +10; 4 - from -5 to +5;
5 – from –5 to –20; 6 - less than -20

б

1 - more than +20; 2 - from +10 to +20;
3 - from 0 to +10; 4 – from –10 to 0;
5 – from –10 to –20; 6 - less than -20

Source: Author's calculations based on Belstat data.

Figure 5. Dynamics of the share of Russians calling Belarusian their native (a) and home (b) language



Source: Author's calculations based on Belstat data.

For Russians, an increase in the share of Belarusian as both their native and home language was noted only in the Brest oblast. At the same time, we see a rather interesting situation in which, in the Brest oblast, the share of Belarusians calling Belarusian their home language in 2009-2019 decreased more than in any other region - 12.8%, while among Russians it increased by 6.2 times. In the oblast's center - the city of Brest - these figures are even more striking: the share of Russians indicating Belarusian as their native language increased from 1.5 to 14.1%, and as their home language - from 0.3 to 22.9%. The results of the analysis of the dynamics of these indicators over several decades (Figure 5) also show that until 2019, no sharp and opposite changes in direction to other regions were observed in any region.

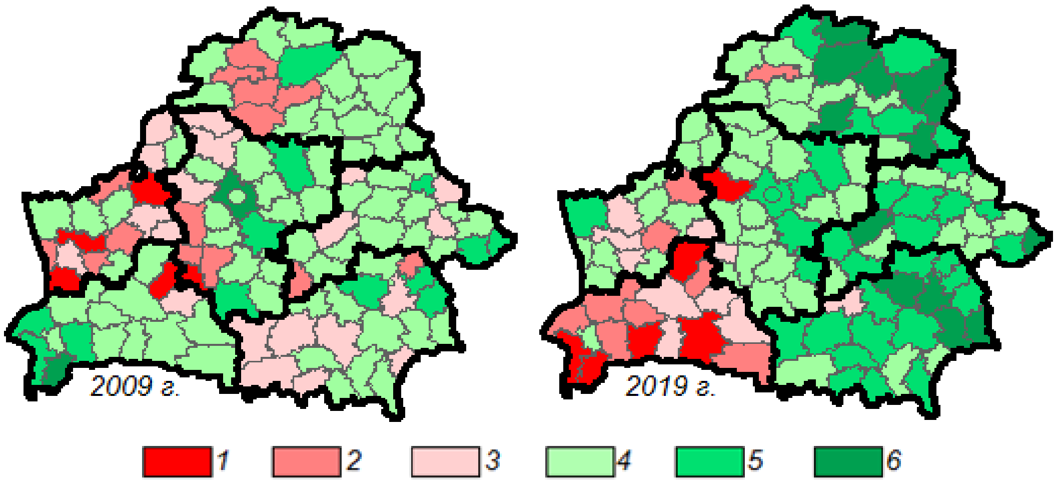
Table 10. The share of Belarusian as native and home language for Russians with different levels of education, %

Education level	Region							
	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast	Belarus
Native language								
Primary	2.7	2.6	1.9	9.3	0.9	3.6	2.3	3.1
Upper primary	9.1	2.0	1.4	4.9	1.8	3.1	2.4	3.4
Secondary	6.8	1.6	1.7	4.6	1.4	1.7	1.0	2.3
Vocational	2.3	1.5	0.8	2.2	0.8	2.4	2.0	1.5
Specialized secondary	8.4	1.0	1.0	3.3	1.7	2.7	1.9	2.6
Higher and postgraduate	13.8	0.6	0.8	2.1	1.2	1.6	1.2	3.1
Home language								
Primary	4.4	3.2	2.5	9.5	1.6	6.5	2.6	4.2
Upper primary	10.6	2.5	1.6	5.0	1.0	3.9	3.0	3.9
Secondary	4.9	1.9	1.3	4.4	0.5	1.9	1.2	2.0
Vocational	2.2	1.7	0.6	2.2	0.5	2.9	1.5	1.4
Specialized vocational	8.4	0.8	0.5	2.5	0.8	2.5	1.3	2.2
Higher and postgraduate	11.6	0.4	0.5	1.3	0.7	1.2	0.6	2.4

Source: Author's calculations based on Belstat data.

The anomalous growth in the share of Russians calling Belarusian their native and home language is reflected in other characteristics as well. For example, the 2019 census data on the distribution of the Russian population with Belarusian as their native and home language by educational level in the regions of Belarus (Table 10) show a huge difference in this indicator between the Brest oblast and other regions, especially for the population with higher education. Moreover, while in all other regions the proportion of people calling Belarusian their native and home language is generally lower among people with higher education than among those with any other level of education, in the Brest oblast the opposite is true.

Figure 6. Share of Russians calling Belarusian their native language (urban population), 2009 and 2019, %

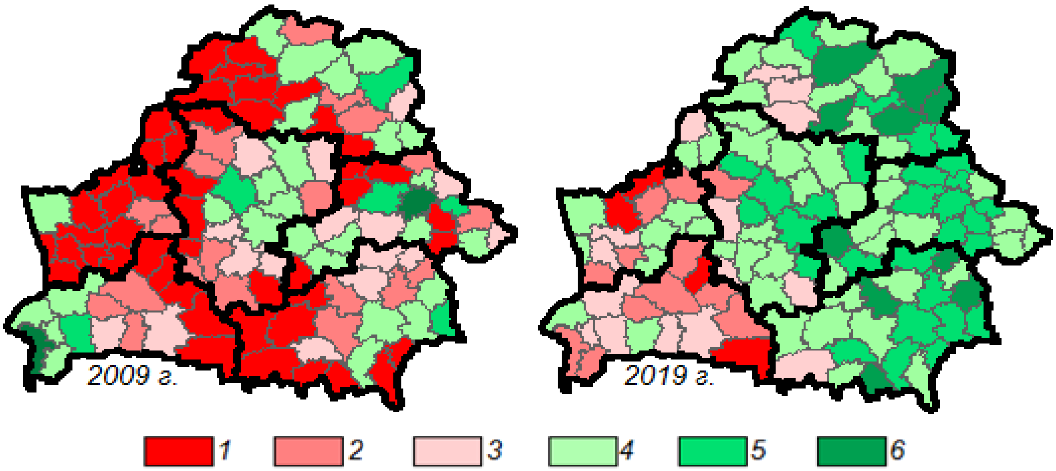


1 - more than 10; 2 - 7-10; 3 - 5-7; 4 - 2-5; 5 - 1-2; 6 - less than 1

Source: Author's calculations based on Belstat data.

In terms of administrative districts and types of locality (Figures 6-8), among Russians the situation with the share of Belarusian as a native language and its dynamics in 2009-2019 is generally similar to that among Belarusians: an increase in this share is found in almost all administrative districts of the Brest oblast and in almost none outside the region.

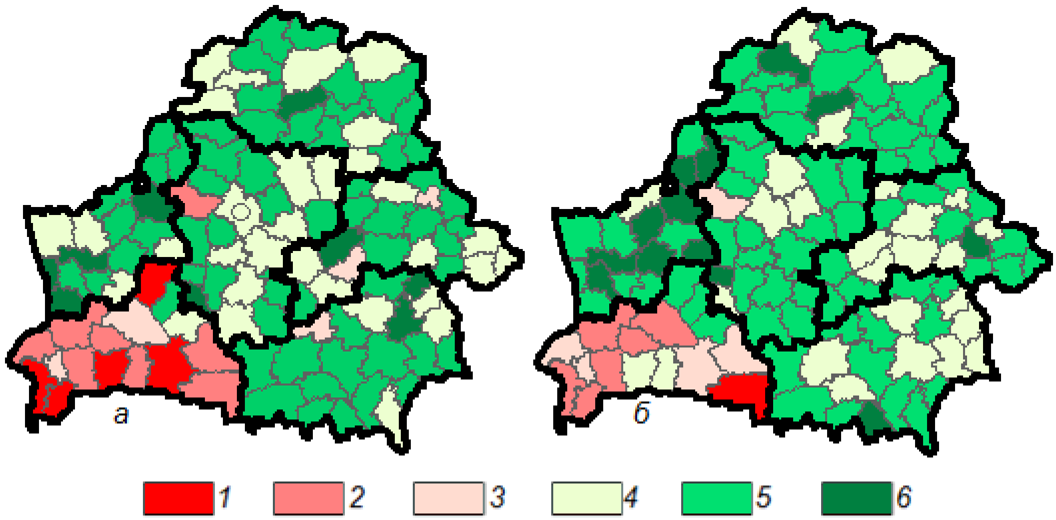
Figure 7. Share of Russians calling Belarusian their native language (rural population), 2009 and 2019, %



1 - more than 10; 2 - 7-10; 3 - 5-7; 4 - 2-5; 5 - 1-2; 6 - less than 1

Source: Author's calculations based on Belstat data.

Figure 8. Change in the share of Russians calling Belarusian their native language among the urban (a) and rural (b) population, 2009-2019, %



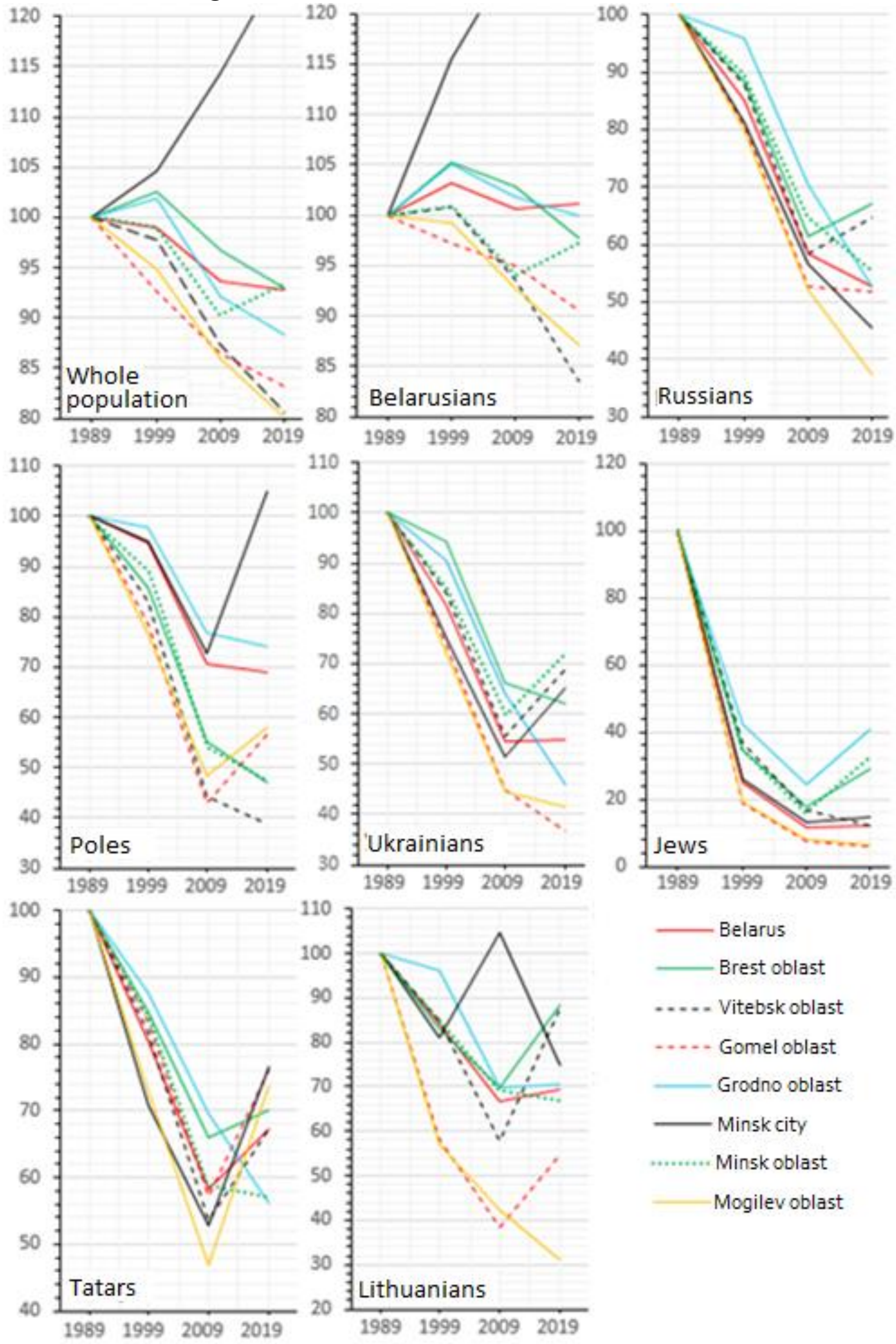
1 - more than +8; 2 - from +3 to +8; 3 - from +1 to +3;
4 – from –1 to +1; 5 – from –1 to –5; 6 - less than -5

Source: Author's calculations based on Belstat data.

Ultimately, it was the anomalous changes in some regions (1-2 regions for each indicator) that ensured the growth in the share of Belarusian both as a native and as a home language for the population of Belarus as a whole, as well as for the Belarusian and Russian populations separately.

The changes in the size of the Russian and Belarusian population of the regions and the number of those calling Belarusian their native and home language are revealing (Figure 9; Table 11).

Figure 9. Base population growth rates of individual nationalities in Belarus and its regions, 1999-2019, % relative to the level of 1989



Source: Author's calculations based on Belstat data.

As can be seen in Figure 9, the most recent census showed for the first time very different changes in the populations of the nationalities in question by region compared to previous censuses, which showed largely synchronized changes.

For example, the population of Russians, which has been continuously and significantly declining in all regions, unexpectedly increased in the Brest and Vitebsk oblasts in 2019, while in both these regions the total population, as well as the number of Belarusians, decreased. Moreover, in the Vitebsk oblast the decrease in the size of the Belarusian population and of the entire population as a whole turned out to be the greatest of all regions, as did the increase in size of the Russian population. On the other hand, in Minsk the population of all major nationalities increased, with the exception of Russians, whose population noticeably decreased, although it is in Minsk that the share of Russians relative to the entire Russian population in Belarus is the highest - more than 20%. At the same time, there are no objective factors capable of explaining the growth in the size of the Russian population in the Brest and Vitebsk oblasts, given the decrease in the total population in them and the significant decrease in the absolute number of Russians in other regions.

Among Belarusians, a significant increase in their numbers was recorded in Minsk, which accounts for the small increase in persons of Belarusian nationality in the country as a whole. It is precisely for Minsk that the largest increase in the population is accompanied by the largest increase in those calling Belarusian their native and home language (Table 11). For Russians, the population growth is accompanied by an increase in the number of those calling Belarusian their native and home language only in the Brest oblast, while for the Vitebsk oblast the situation is the opposite.

Traditionally, population growth in Minsk is explained by migration, since Minsk has always had the lowest population replacement rates among all regions, which, moreover, have been continuously decreasing since 2014. Nevertheless, it was precisely Minsk that between the 2009 and 2019 censuses had the highest absolute increase in the population of Belarusians below working age among all regions (38.4% versus 5.2% for all other regions as a whole) and one of the highest increases in its share among the entire population. Compared with the dynamics of this indicator between the 1999 and 2009 censuses, for Minsk in 2009 the change in the number of Belarusians below working age was -22.8%, which was slightly higher than the national average (-26.2%) and corresponded to the indicators for other regions (from -22.2 to -31.4%, and in general for all regions, except for the city of Minsk, -26.8%). After the census, in 2020 and 2021, the population of Minsk, according to current figures, steadily decreased.

For Russians, the difference in the growth of the population below working age varies dramatically by region, with a maximum value in the Brest oblast of + 76.9% and also very high numbers in the Vitebsk and Gomel oblasts. In other oblasts, this indicator falls by tens of percent. Between the 1999 and 2009 censuses there was, as for the Belarusians, no such difference between the regions, with the value of this indicator ranging from -64.3% in Minsk to -71.6% in the Gomel oblast.

In the Brest and Gomel oblasts, the child-woman ratio (the ratio of children aged 0-9 years to the number of women of reproductive age of the corresponding nationality) more than doubled, and in the Vitebsk oblast is also noticeably higher than the national average (Table 12).

Based on these indicators and their dynamics, it can be assumed that there is a connection between the anomalous changes in language characteristics and the anomalous increase in the population, especially in the group below working age. Thus, in Minsk, which is the leader in terms of population growth of Belarusians in this group, there is an abnormally high increase in the number of those calling Belarusian their native and home language. The Mogilev oblast, where the increase in the proportion of the population of Belarusians below working age in 2009-

2019 was the highest among all regions of Belarus and the overall population growth one of the lowest, shows an anomalous increase in the number of people identifying Belarusian as their home language. In the Brest oblast, where the Russian population registered an increase in the size of the population under working age much higher than in other regions, there occurred at the same time an anomalous increase in the number of Russians calling Belarusian their native and home language. The same patterns were recorded for other nationalities, as will be shown below.

Table 11. Changes in the population size and the number of those calling Belarusian their native and home language among the Belarusian and Russian population of the regions of Belarus

Oblasts and Minsk city	Population size			Including those under working age			Belarussian native			Belarussian home		
	2009	2019	+/–, %	2009	2019	+/–, %	2009	2019	+/–, %	2009	2019	+/–, %
Belarussians												
Brest oblast	1 233 385	1 171 835	–5.0	235 059	241 747	+2.8	735 662	1 009 981	+37.3	362 320	195 083	–46.2
Vitebsk oblast	1 047 979	934 925	–10.8	167 269	158 930	–5.0	634 931	504 027	–20.6	266 235	136 977	–48.6
Gomel oblast	1 271 022	1 211 234	–4.7	224 750	238 530	+6.1	778 740	648 302	–16.7	321 295	198 394	–38.3
Grodno oblast	715 251	701 190	–2.0	130 172	142 392	+9.4	492 182	430 443	–12.5	274 547	276 472	+0.7
Minsk city	1 455 875	1 753 122	+20.4	228 127	315 667	+38.4	633 119	968 556	+53.0	102 212	674 762	+560.2
Minsk oblast	1 258 672	1 302 780	+3.5	215 324	247 400	+14.9	965 908	862 223	–10.7	534 560	538 146	+0.7
Mogilev oblast	975 148	915 633	–6.1	165 641	168 575	+1.8	600 845	469 607	–21.8	212 735	255 409	+20.1
Russians												
Brest oblast	89 685	97 936	+9.2	7 019	12 418	+76.9	2 299	11 107	+383.1	1 610	10 831	+572.7
Vitebsk oblast	124 958	138 075	+10.5	9 079	14 452	+59.2	3 462	1 324	–61.8	3 084	1 196	–61.2
Gomel oblast	111 085	108 712	–2.1	7 583	10 508	+38.6	2 751	1 137	–58.7	2 025	729	–64.0
Grodno oblast	87 451	65 550	–25.0	9 013	6 565	–27.2	4 291	1 980	–53.9	3 847	1 576	–59.0
Minsk city	184 070	148 079	–19.6	11 849	10 180	–14.1	3 700	1 905	–48.5	1 123	963	–14.2
Minsk oblast	101 579	86 408	–14.9	7 051	6 316	–10.4	3 172	1 863	–41.3	3 314	1 787	–46.1
Mogilev oblast	86 256	62 232	–27.9	6 556	4 248	–35.2	2 281	940	–58.8	1 650	699	–57.6

Source: Author's calculations based on Belstat data.

Pearson's linear correlation coefficient between the increase in the population of Belarusians by regions of Belarus and the increase in the number of Belarusians calling Belarusian their home language is r = +0.93, p <0.01.

The author can assume that the anomalous increase in the population, especially of those below working age, in certain regions for certain nationalities, together with a completely different dynamics of this indicator in other regions and an equally anomalous increase in the share of the Belarusian language (in some cases, a statistically significant correlation is observed between these indicators), is the result of a deliberate distortion of the census results (primarily by overstating the population size below working age.). Moreover, for the vast majority of the populations recording anomalous (distorted) growth, a single language was reported as both native and home: Belarusian as native and home for the Belarusians of Minsk, as well as for the Russians of the Brest oblast, Belarusian as the home language for the Belarusians of the Mogilev oblast, Russian as native and home for Russians in the Vitebsk and Gomel oblasts, Belarusian as native and home for Poles in the Gomel and Mogilev oblasts and Minsk, Belarusian as native for Ukrainians in the Vitebsk oblast, Ukrainian as native for Ukrainians in the Mogilev oblast, etc.

It is for this reason that the child-woman ratio skyrocketed over 10 years, and all this against the backdrop of a decline in fertility, when the crude birth rate decreased from 11.5 in 2009 to 9.9 in 2019, and the total fertility rate from 1.40 to 1.27. It is possible that the overstatement of the number of persons below working age eliminates the need to correct such census results as belonging to the labor force, education, marital status, etc., categories which here are not relevant, and in general creates a more favorable demographic picture against the backdrop of the sharp decline in fertility that began in 2017. At the same time, since 2019, Belstat has completely stopped publication of any data at all on the natural movement of the population and of the Demographic Yearbook and corresponding bulletins, and has excluded thematic sections from all other statistical collections. Among nationalities with a small population, such an overstatement is more pronounced compared to Belarusians and Russians and leads to such results as when, for example, in 2009 in the Minsk oblast there was 1 child per 5 women among Jews, and in 2019 18 children per the same 5 women (Table 12).

Table 12. Change in the child-woman ratio by regions and nationalities

Region		Whole population	Belarusians	Russians	Poles	Ukrainians	Jews	Tatars	Lithuanians
Belarus	2009	0.38	0.40	0.18	0.29	0.08	0.24	0.14	0.20
	2019	0.49	0.51	0.29	0.47	0.25	1.50	0.42	0.30
Brest oblast	2009	0.43	0.46	0.18	0.19	0.09	0.30	0.10	0.06
	2019	0.56	0.58	0.37	0.36	0.22	3.97	0.40	0.22
Vitebsk oblast	2009	0.35	0.38	0.17	0.16	0.06	0.21	0.07	0.10
	2019	0.45	0.47	0.32	0.30	0.41	0.97	0.37	0.13
Gomel oblast	2009	0.39	0.42	0.15	0.07	0.06	0.20	0.07	0.14
	2019	0.51	0.54	0.30	0.38	0.18	1.25	0.45	0.24
Grodno oblast	2009	0.39	0.44	0.26	0.32	0.10	0.43	0.26	0.26
	2019	0.55	0.58	0.36	0.50	0.26	2.87	0.45	0.34
Minsk city	2009	0.32	0.34	0.17	0.14	0.06	0.24	0.16	0.22
	2019	0.42	0.44	0.23	0.36	0.26	1.20	0.39	0.41
Minsk oblast	2009	0.40	0.42	0.16	0.18	0.08	0.20	0.10	0.17
	2019	0.52	0.54	0.24	0.33	0.28	3.57	0.42	0.42
Mogilev oblast	2009	0.38	0.40	0.18	0.10	0.06	0.29	0.06	0.14
	2019	0.49	0.51	0.23	0.38	0.28	0.95	0.55	0.51

Note: Cases of an increase in the indicator by more than 2 times have been identified.

Source: Author's calculations based on Belstat data.

It should also be noted that for the majority of the population below working age, the answers to the census questions (including about their native and home language) are provided by parents. Parents also choose the language of instruction for children in preschool and general secondary education. Therefore, it is necessary to analyze the data on the languages of instruction in these institutions by regions of Belarus (Table 13). In Minsk, the proportion of students whose schooling is in the Belarusian language is noticeably lower than in all other regions; in the Mogilev oblast this proportion is also lower than the national average and at the same time is trending downward, although it is precisely these regions that had the maximum increase in the absolute number of Belarusians identifying Belarusian as their home language and in its share among the entire population (Tables 6; 11).

Table 13. Belarusian as the language of instruction in preschool and general secondary school, 2012-2018, %

Region	Preschool		General secondary	
	2012	2018	2012	2018
Brest oblast	15.7	13.1	22.9	17.4
Vitebsk oblast	10.8	6.6	14.9	7.6
Gomel oblast	7.5	6.8	13.8	9.8
Grodno oblast	18.5	15.7	19.6	12.5
Minsk city	3.1	3.5	2.0	2.1
Minsk oblast	20.5	15.4	28.9	20.1
Mogilev oblast	9.0	6.7	14.8	9.8
Belarus	11.4	9.1	16.5	11.1

Source: (Belstat 2019).

In this regard, there is a paradoxical situation: parents are responsible for their children, for example, using Belarusian at home, yet at the same time they choose Russian as their language of instruction, and the share of Belarusian in educational institutions is constantly and steadily declining. In addition, the figures given in Table 13 are provided mainly by rural schools, in most of which the language of instruction is Belarusian, so that parents have no actual choice, since there are no other schools within reach and there is only one group for each grade. At the same time, there are cases when parents, dissatisfied with this state of affairs, have been seeking for years the possibility of having their children schooled in Russian[1]. In cities (where 77.5% of the population lives), in some cases the number of schoolchildren studying in the Belarusian language is a mere handful[2],[3]. A vivid example is the only completely "urban" region - the city of Minsk, where 2.1% of school students study in the Belarusian language, while in Minsk, according to this census, the proportion of those calling Belarusian their home language "soared" from 5.8% in 2009 to 34.1% in 2019, and in absolute numbers from 102,212 to 674,762 people - that is, this figure increased by 560% (!).

Standing out against the general background is the Brest oblast, where anomalous changes have occurred in almost all administrative districts: out of 16 districts in 2019, the size of the Russian population increased in 11, the share of Russians calling Belarusian their native language in 13 and their home language in 11, and the share of Belarusians calling Belarusian their native language in 15, while the share of Belarusians calling Belarusian their home language decreased in absolutely all regions. Between the increase in the number of Russians in the administrative districts of the Brest oblast and the increase in the number of Russians calling Belarusian their home language, there is a significant rank correlation: Spearman coefficient $R = +0.52$, $p < 0.05$. Also, for the districts of the Brest oblast, a fairly high linear correlation was established between the increase in the number of Belarusians and of Russians both calling Belarusian their native tongue ($r = 0.81$, $p < 0.001$).

These features in the Brest oblast differ sharply both from all other regions and from the results of previous censuses, when they were not observed. So, for example, for the share of

[1] In Zhirovichi, they fought for 10 years to change schooling into Russian. https://sputnik.by/20211018/v-zhirovichakh-10-let-dobivalis-chtoby-perevesti-obuchenie-v-shkole-na-russkiy-yazyk-1057278559.html
[2] Another Belarusian-speaking classroom opened in Belarus. https://news.rambler.ru/cis/34613020-v-belorussii-otkrylsya-esche-odin-belorusskoyazychnyy-klass/
[3] Class for one student in a school with a swimming pool.: https://ru.hrodna.life/2019/08/16/pervoklassniki-grodno-po-belorusski/

those who called Russian their native language among the entire urban population, only in the Brest oblast did the indicator of variation of this value by administrative districts sharply decrease. While in 2009 the standard deviation of this indicator for the districts of the Brest oblast was 16.6, in 2019 it was 4.0 (thus decreasing by 4.2 times), whereas for all other regions of Belarus this value in 2019 was 7.5-13.6, and its maximum decrease in 2009-2019 1.8 times. As a result, while in 2009 the difference between the minimum and maximum values of the indicator for the urban population of the districts of the Brest oblast was very significant (57.9%), in 2019 it was only 13.2%; all values were around the 20% mark, deviating from it by an average of 3.4% and a maximum of 7.1%. It can be assumed that it was the figure of 20% that became the value that was initially administratively set as a benchmark to which it was necessary to “adjust” the results.

“Adjustment” of the results to the same initially given value for all districts led to an extremely high Pearson linear correlation coefficient (r = +0.98, p <0.01) between the proportion of the urban population who called Russian their native language in 2009 and the modulus of change of this indicator. For other regions of Belarus, the situation is diametrically opposite, and this coefficient has a negative value (r = -0.52, p <0.01), with similar values shown by the calculation of this indicator for each region separately.

The statistical significance of the differences between the Brest oblast and other regions in terms of the share of Russian as a native language is also confirmed by the calculation of Student's t-test. Thus, its calculation for the indicator of the change in the share of Russian as a native language for two populations (districts of the Brest oblast and all other districts of Belarus) showed its value t = 12.3 (with $t_{critical}$ = 2.61 for p <0.01). The calculation of the t-test for 5 pairs of populations of districts of the Brest oblast and each of the other oblasts also showed a high value from 8.1 to 10.4. At the same time, there were no significant differences in the t-test for any of the pairs of other regions. A similar result was shown by calculations of the t-test for the rural population.

Thus, the anomalous increase in the population, especially in working age, together with other ways of distorting the census results that are not related to population registrations (for example, a sharp increase in the share of Belarusian as a native language for Belarusians in the Brest oblast is not accompanied by an anomalous increase in the share of people below working age), made it possible to achieve "ideal" results - an increase by a small amount in the share of Russian and Belarusian both as native and as home languages for both Russians and Belarusians, as well as for the entire population as a whole.

Change in the population size and language structure of other nationalities

The third largest nationality in Belarus are the *Poles*, more than three-quarters of whom live in the Grodno oblast and who in 2009 had the country’s lowest level of urbanization (61.4%). Just as for all other major nationalities, the 2019 census for the first time showed multidirectional dynamics of the Polish population in different regions (Figure 9; Table 14).

At the same time, attention should be paid to the fact that the population growth, quite noticeable, occurred only in regions that historically were not the traditional place of residence of the Poles: in the Mogilev and Gomel oblasts and especially in Minsk, where over 10 years their number increased by 44 .5% - the largest relative increase in the number of all nationalities in Minsk for the period 2009-2019. (Largely due to this, the urbanization coefficient for Poles increased by 9.9%, which is a record among all nationalities - more than 3 times higher than the increase in this coefficient on the national average). In these regions, the growth of the Polish population was ensured primarily by the population below working age, for which the growth rate is more than an order of magnitude higher than its national average, and the birth rate increased by 2.6-5.4 times (Table 12). And it was precisely in these regions that the number and proportion of the population calling Belarusian their native and home language grew, also significantly (Tables 3; 6; 15).

Table 14. Growth of the Polish population in the oblasts and in Minsk

Regions	Population size			Including those under working age		
	2009	2019	+/–, %	2009	2019	+/–, %
Belarus	294 549	287 693	-2.3	36 766	44 692	+21.6
Brest oblast	17 539	14 893	-15.1	1 264	1 594	+26.1
Vitebsk oblast	11 141	9 806	-12.0	589	873	+48.2
Gomel oblast	1 958	2 572	+31.4	57	348	+510.5
Grodno oblast	230 810	223 119	-3.3	32 602	37 334	+14.5
Minsk city	13 420	19 397	+44.5	838	2 749	+228.0
Minsk oblast	17 908	15 785	-11.9	1 344	1 541	+14.7
Mogilev oblast	1 773	2 121	+19.6	72	253	+251.4

Source: Author's calculations based on Belstat data.

Table 15. Growth in the number of those calling Belarusian their native and home language among the Polish population of the oblasts and Minsk

Region	Belarusian, native			Belarusian, home		
	2009	2019	+/–, %	2009	2019	+/–, %
Belarus	171 287	156 650	-8.5	120 378	132 366	+10.0
Brest oblast	9 673	6 518	-32.6	6 480	4 273	-34.1
Vitebsk oblast	6 195	4 284	-30.8	5 366	3 668	-31.6
Gomel oblast	860	1 176	+36.7	464	1 035	+123.1
Grodno oblast	133 920	125 113	-6.6	93 995	109 210	+16.2
Minsk city	5 491	8 370	+52.4	1 684	5 081	+201.7
Minsk oblast	14 553	10 277	-29.4	12 168	8 425	-30.8
Mogilev oblast	595	912	+53.3	221	674	+205.0

Source: Author's calculations based on Belstat data.

The coefficient of linear correlation between the population growth in the regions and the increase in the proportion of those calling Belarusian their native language of the whole Polish population is quite high (r = 0.79, p <0.05), and the increase in the proportion of those calling it their home language is even higher (r = 0.84, p <0.05).

In regions where the Polish population size is decreasing, the relative value of the decrease in those calling Belarusian their mother tongue is 2 or more times higher than the decrease in the total size of the Polish population. In regions with an increasing number of Poles, on the contrary, the relative increase in the number of people with Belarusian as their native and, especially, home language, exceeds the overall relative increase. Exclusively in the Brest oblast, the proportion of Poles calling Polish their home language has noticeably increased - up to 7.8%, which is 6.5 times higher than the national average; in the Grodno oblast, where the total number of Poles decreased by 7.7 thousand people, the number of those calling Belarusian their native language also decreased by 8.8 thousand people, but the number of those calling it their home language increased by 15.2 thousand people.

For *Ukrainians*, who reside mainly in the Brest oblast (about a quarter of all Ukrainians), the peculiarities of changes in language characteristics are also closely connected to administrative districts. Just as for the Poles, a significant increase in the Ukrainian population, and in particular an increase of hundreds of percent of the population under working age, was observed only in regions where a relatively small number of Ukrainians live, while in the Brest and Gomel oblasts, which in 2009 had the largest number of Ukrainians, it decreased (Table 16). However, in the end, all the same, in the country as a whole the results of the census showed a slight increase in the Ukrainian population, even though all previous censuses had shown a synchronous decrease in the number of Ukrainians across all regions (Figure 9).

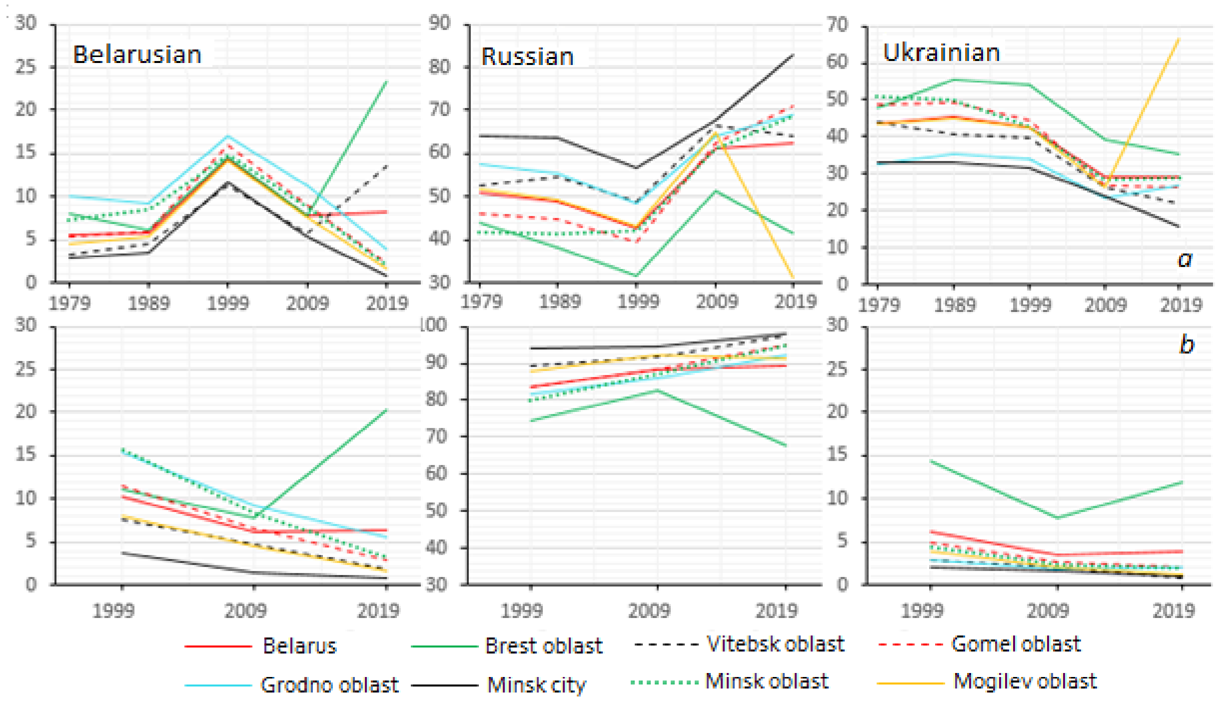
Table 16. Growth of the Ukrainian population in the oblasts and Minsk

Region	Population size			Including those below working age		
	2009	2019	+/–, %	2009	2019	+/–, %
Belarus	158 723	159 656	+0.6	5 986	13 884	+131.9
Brest oblast	40 046	37 648	–6.0	2 365	3 138	+32.7
Vitebsk oblast	14 557	17 993	+23.6	403	2 341	+480.9
Gomel oblast	30 920	25 085	–18.9	967	1 480	+53.1
Grodno oblast	14 983	10 767	–28.1	654	814	+24.5
Minsk city	27 362	34 662	+26.7	610	2 993	+390.7
Minsk oblast	17 745	21 273	+19.9	618	1 994	+222.7
Mogilev oblast	13 110	12 228	–6.7	369	1 124	+204.6

Source: Author's calculations based on Belstat data.

The share of Belarusian as a native language among Ukrainians increased sharply only in the Brest and Vitebsk oblasts (by 2.9 and 2.3 times, respectively). In all other regions this indicator continued a noticeable decline, which was typical for all regions in 1999-2009. (Figure 10). The share of Belarusian as a home language increased sharply (2.6 times) only in the Brest oblast, and its dynamics in the Vitebsk oblast did not differ from the dynamics in other regions.

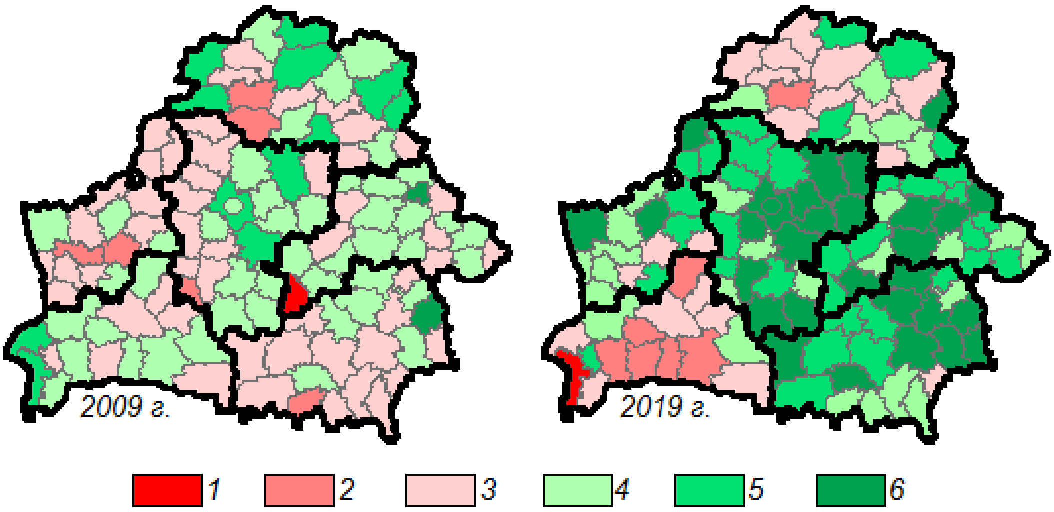
Figure 10. Dynamics of the share of Belarusian, Russian and Ukrainian as native (a) and domestic (b) language for the Ukrainian population, %



Source: Author's calculations based on Belstat data.

Figure 11 shows that an increased (more than 10%) share of Belarusian as a native language among the urban population was observed in 2009 in 57 out of 118 administrative districts of Belarus. Of these districts, only 26% were in the Brest and Vitebsk oblasts. In 2019, such values were observed only in 28 districts, and already 86% of them are districts of the Brest and Vitebsk oblasts, and outside them they were recorded only in 4 districts. Among the rural population (Figure 12), the picture is similar - in 2019, out of 26 districts where the share of Belarusian as a native language exceeds 10%, 20 districts are located in the Brest and Vitebsk oblasts.

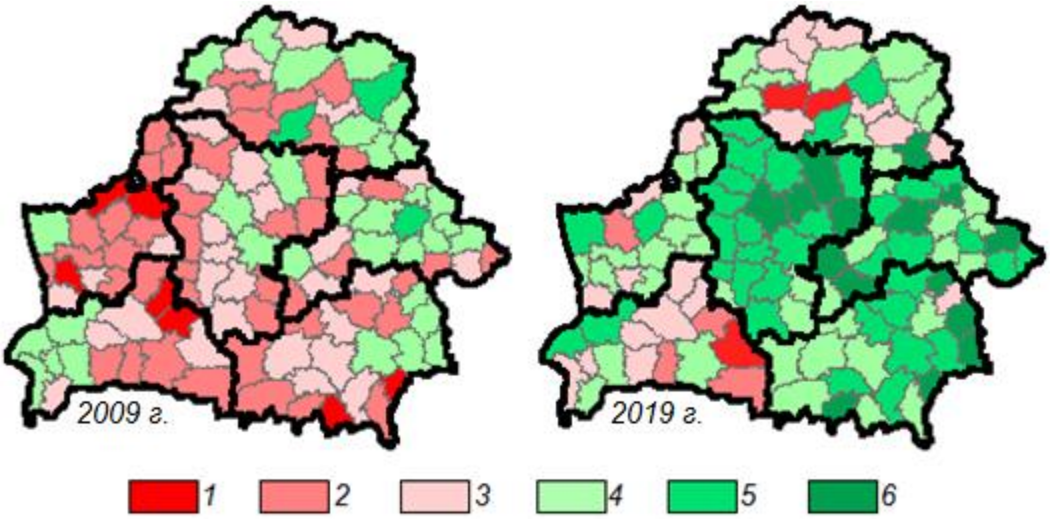
Figure 11. Share of Ukrainians calling Belarusian their native language (urban population), 2009 and 2019, %



1 - more than 30; 2 - 20-30; 3 - 10-20; 4 - 5-10; 5 - 2-5; 6 - less than 2

Source: Author's calculations based on Belstat data.

Figure 12. Share of Ukrainians calling Belarusian their native language (rural population), 2009 and 2019, %

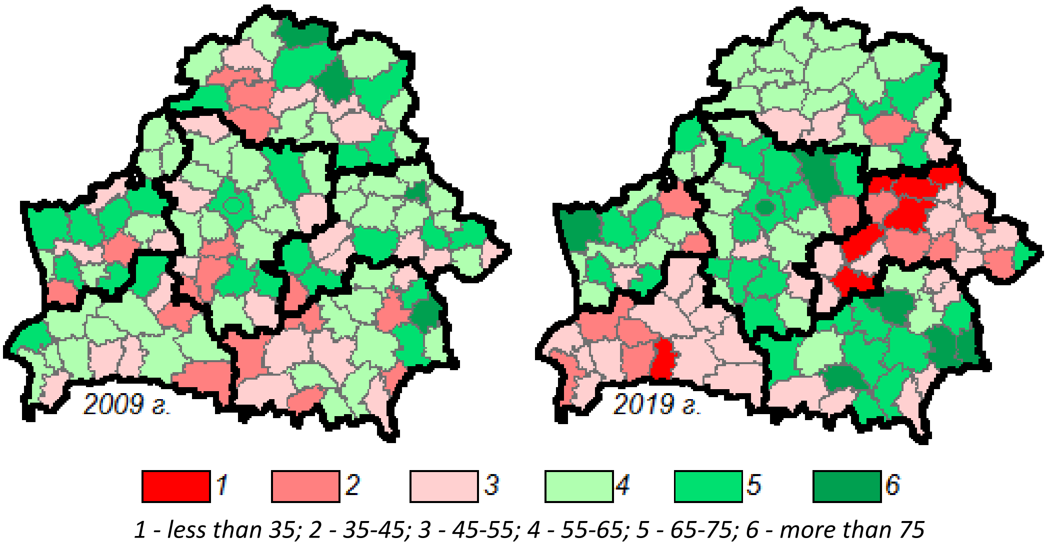


1 - more than 25; 2 - 15-25; 3 - 10-15; 4 -5-10; 5 - 2-5; 6 - less than 2

Source: Author's calculations based on Belstat data.

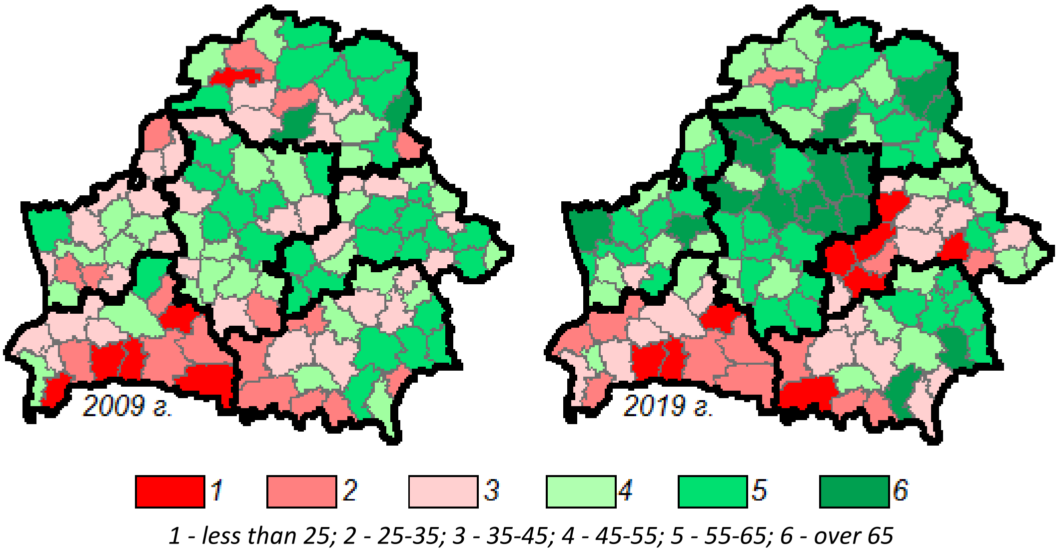
In terms of the decrease in the share of Russian as a native language for the urban population (Figure 13), the Brest and Mogilev oblasts clearly stand out, where in absolutely all districts, except for one district of the Mogilev oblast, it fell below 55% in 2019. The Mogilev oblast was also the only one in Belarus to see a noticeable decrease in the share of the Russian language among the rural population in most districts (Figure 14).

Figure 13. Share of Ukrainians calling Russian their native language (urban population), 2009 and 2019, %



Source: Author's calculations based on Belstat data.

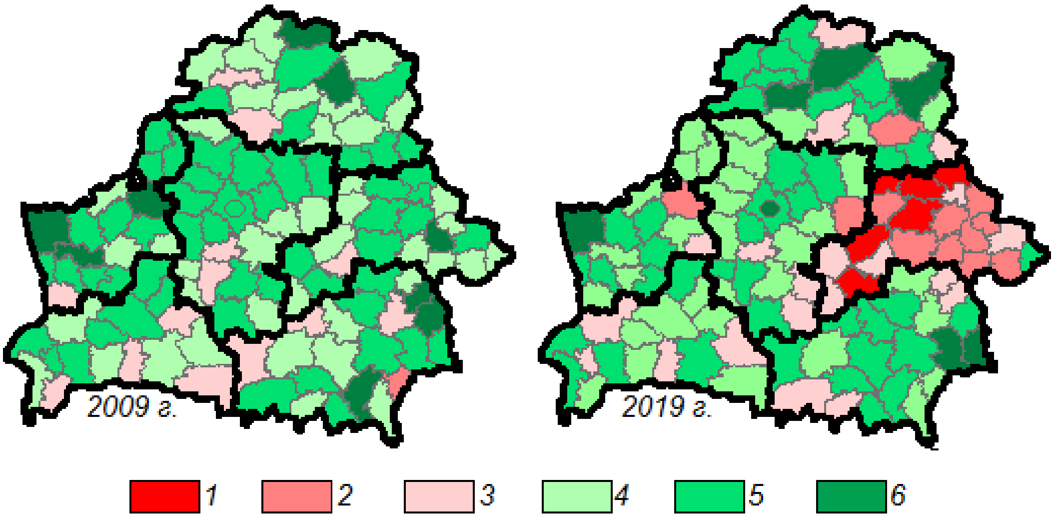
Figure 14. Share of Ukrainians calling Russian their native language (rural population), 2009 and 2019, %



Source: Author's calculations based on Belstat data.

The urban population of the districts of the Gomel and Brest oblasts has no fundamental differences in these indicators from other regions. This is clearly seen in Figures 15 and 16, which show that the situation with Ukrainian as a native tongue did not change significantly between 2009 and 2019, with the exception of anomalous results in the Mogilev oblast, which, despite never having had a high level of Ukrainian, became the only oblast to see a dramatic increase in its share as a native language, moreover, in 20 out of 21 districts at once.

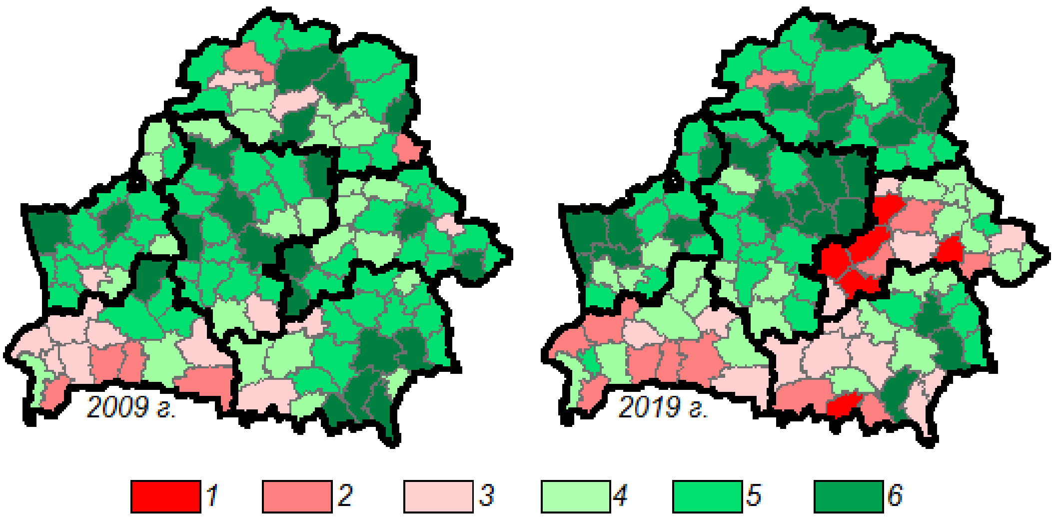
Figure 15. Share of Ukrainians calling Ukrainian their native language (urban population), 2009 and 2019, %



1 - more than 60; 2 - 60-50; 3 - 50-40; 4 - 40-30; 5 - 30-20; 6 - less than 20

Source: Author's calculations based on Belstat data.

Figure 16. Share of Ukrainians calling Ukrainian their native language (rural population), 2009 and 2019, %



1 - more than 70; 2 - 60-70; 3 - 50-60; 4 - 40-50; 5 - 30-40; 6 - less than 30

Source: Author's calculations based on Belstat data.

Thus, the dynamics of native languages for Ukrainians have pronounced individual features in different regions (Table 17).

As for the dynamics of the share of home languages, only the Brest oblast saw an increase in the shares of Belarusian (by 2.6 times) and Ukrainian (by 1.5 times) and a corresponding decrease in the share of Russian, while in other regions it is the other way around. In the Mogilev oblast, the share of all three languages is falling. This became possible due to a sharp increase in the proportion of people who did not indicate their home language during the census.

Table 17. Changes in the share of languages as native among Ukrainians by regions

Native language	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast
Belarusian	green	green	red	red	red	red	red
Russian	red	yellow	green	green	green	green	red
Ukrainian	red	red	yellow	yellow	red	yellow	green

Note: red – *decrease over 4%;* yellow – *magnitude of change ± 4%;* green – *increase over 4%.*

Source: Compiled by the author.

In general, this indicator is an excellent illustration of the strict correlation of the dynamics of language characteristics to administrative units and individual nationalities, which cannot be the result of natural causes, but indicates a deliberate distortion of the census results (Table 18).

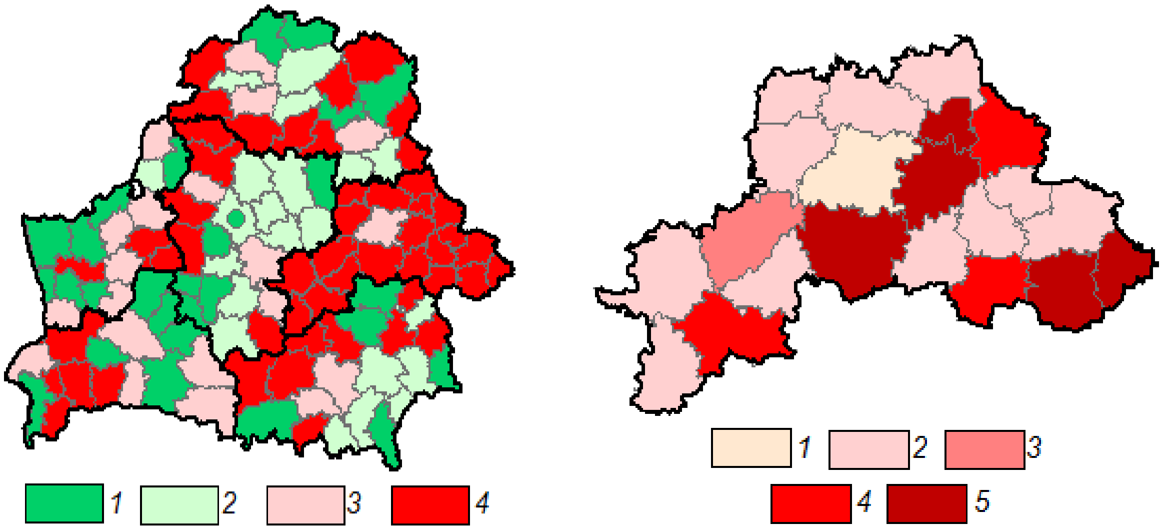
Table 18. Dynamics of the share of persons not indicating their native and home languages, by nationalities and regions, 2009-2019

Nationality		Native language							Home language						
		Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast	Brest oblast	Vitebsk oblast	Gomel oblast	Grodno oblast	Minsk city	Minsk oblast	Mogilev oblast
Belarusians	2009	0.5	0.4	0.5	0.5	1.1	0.7	0.4	0.8	0.7	0.9	1.0	1.4	1.3	0.7
	2019	0.0	0.0	0.0	0.0	0.0	0.1	4.2	0.0	0.0	0.0	0.0	0.0	0.0	4.4
Russians	2009	0.3	0.3	0.3	0.3	0.4	0.4	0.3	0.5	0.5	0.6	0.6	0.6	0.6	0.4
	2019	0.4	0.5	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7
Poles	2009	0.4	0.1	0.3	0.4	0.9	0.3	1.6	0.5	0.2	0.3	0.7	1.1	0.5	1.7
	2019	0.0	0.1	0.3	0.3	0.3	0.6	4.3	0.0	0.0	0.0	0.1	1.1	1.3	6.7
Ukrainians	2009	0.4	0.4	0.5	0.3	0.7	0.6	0.5	0.5	0.5	0.5	0.3	0.8	0.8	0.5
	2019	0.0	0.6	0.2	0.2	0.2	0.5	0.7	0.0	0.0	0.0	0.0	0.0	0.0	5.7
Jews	2009	1.6	1.0	1.5	0.4	0.7	0.9	1.4	1.4	0.7	0.8	0.2	0.4	0.1	0.3
	2019	1.2	0.6	0.7	0.4	0.2	3.7	6.1	0.0	0.0	0.1	0.0	0.0	0.0	10.0
Tatars	2009	0.6	0.7	1.7	0.7	0.7	0.5	0.2	0.3	0.1	0.8	0.5	0.3	0.4	0.0
	2019	0.5	0.9	0.4	0.1	0.4	4.6	11.5	0.1	0.0	0.0	0.0	0.0	0.0	7.4
Lithuanians	2009	1.4	1.9	1.1	0.5	1.2	1.1	1.5	0.0	1.1	1.1	0.5	1.1	0.4	0.7
	2019	0.9	1.0	1.0	0.9	0.0	2.8	5.0	0.0	0.0	0.0	0.0	0.0	1.5	6.5

Source: Author's calculations based on Belstat data.

The Mogilev oblast is the only one where the proportion of people who did not indicate their native and home languages increases many times among all nationalities and becomes much larger than in all other regions. This is especially noticeable for the home language, where in other regions there is practically no population that did not indicate the home language, while in the Mogilev oblast, among all nationalities, the proportion of such people is very significant and reaches 10% (for Jews). For the native language, this indicator is also clearly differentiated by nationality. For example, in the Mogilev oblast, among all Russians only 2 people in 2019 did not indicate their native language, and among Ukrainians, too, the change compared to 2009 is also insignificant. However, a very sharp increase is observed for all other nationalities, reaching a maximum among the Tatars, where the proportion of those who did not indicate their native language increased by almost 60 times. It is obvious that the question why, out of all the regions, only in Mogilev a significant proportion of the population of the main nationalities decided to hide their native and home languages from the census takers, while the Russian and Ukrainian population, unlike other nationalities, hid only their home languages, is rhetorical in nature, and these results (as well as other anomalies localized within administrative units) allow us to draw a conclusion only about the quality of the census and the level of reliability of its results.

Figure 17. Change in the number of people not indicating their native language, by regions of Belarus (a) and the Mogilev oblast (b), 2009-2019



а

1 - the indicator decreased;
the indicator increased by:
2 - from 0 to 100%; 3 - from 100 to 200%;
4 - more than 200%

б

the indicator increased: 1 - less than 3 times;
2 - 3-7 times; 3 - 7-11 times; 4 - 11-15 times;
5 - more than 15 times

Source: Author's calculations based on Belstat data.

It must be said here that the proportion of people who did not indicate their languages increased in many districts not only of the Mogilev oblast (Figure 17a); however, in all other regions there is no pattern in the dynamics of this indicator by district, and in different districts there are different directions and degrees of change. And only in the Mogilev oblast did all districts without exception show an increase in this indicator, moreover severalfold, almost everywhere by more than 3 times (Figure 17b). Among the entire population of the oblast, the proportion of those who did not indicate their native language increased in 2009-2019 from 1.7 to 6.2%, while in all other regions of Belarus this share did not exceed 2.4% in 2019. Despite the fact that in the Mogilev district of the Mogilev oblast, the increase in the proportion of the population that did not indicate their native language is the smallest of all districts of the oblast (due to the initially high value of the indicator in 2009), this share in it is 8.0%, which is the highest value among all regions of the country. Overall, 4 out of 5 regions of Belarus where this indicator exceeds 5% are located in the Mogilev oblast.

In 1989-2009 the number of Jews in all regions was declining at the fastest rate of all the nationalities considered (Figure 8) and in Belarus it decreased by 75%, including in 1999-2009 by 53%. In 2009-2019 the country recorded an increase in the number of Jews, which (with the exception of Minsk) was noted only in regions where it was minimal in 2009 (and this increase is very significant - 65.4-106.5%) (Table 19). At the same time, a huge increase, reaching 1.000% in the Minsk oblast, was recorded for the population below working age. The proportion of people below working age in Belarus as a whole increased by more than 3 times, and in the Minsk oblast by more than 5 times. The child-woman ratio for Jews in Belarus as a whole increased by 6.2 times, which is 4.7 times higher than its growth for the entire population. In some regions, this value is even higher, for example, in the Brest oblast, where it increased by 13.4 times from 0.30 in 2009 to 3.97 in 2019 (Table 12). As a result, in 2019 the child-woman ratio for Jews became the highest of all nationalities, exceeding by 3.1 times its value for Belarus as a whole, despite the fact that in 2009 it was below the national average.

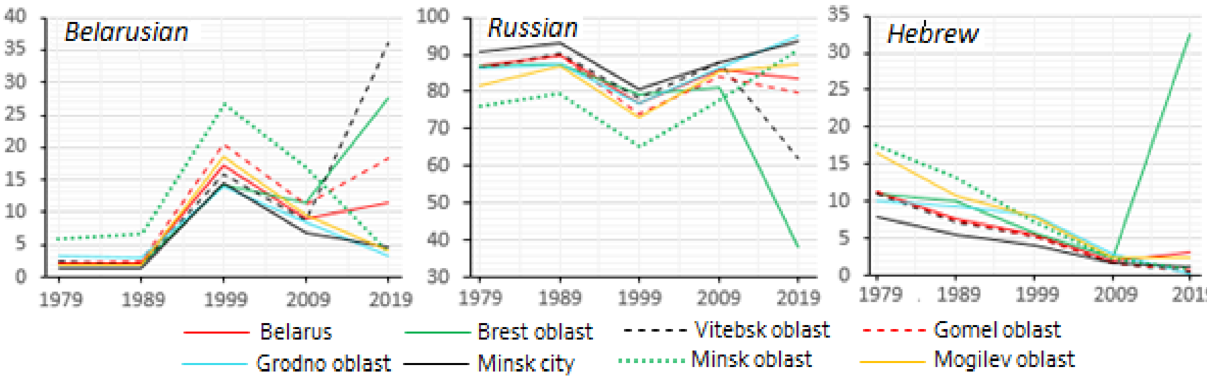
Such drastic changes in population size were accompanied by drastic changes in language characteristics, with, for the first time in 2019, a multidirectional character in different regions (Figure 18).

Table 19. Increase in the number of Jews by region

Region	Population size			Including those below working age		
	2009	2019	+/–, %	2009	2019	+/–, %
Belarus	12 935	13 705	+6.0	545	1 767	+224.2
Brest oblast	570	943	+65.4	32	173	+440.6
Vitebsk oblast	2 127	1 561	–26.6	70	133	+90.0
Gomel Oblast	2 341	1 962	–16.2	89	239	+168.5
Grodno oblast	538	905	+68.2	34	167	+391.2
Minsk city	5 194	5 699	+9.7	225	677	+200.9
Minsk oblast	703	1 452	+106.5	23	253	+1000.0
Mogilev oblast	1 462	1 183	–19.1	72	125	+73.6

Source: Author's calculations based on Belstat data.

Figure 18. Share of Belarusian, Russian and Hebrew as native language for the Jewish population in Belarus and its regions, %



Source: Author's calculations based on Belstat data.

An enormous increase in the number of people calling Belarusian their native language is clearly visible in the cities and rural areas of the Brest, Vitebsk and Gomel oblasts (in some cases by hundreds of percent), as well as Hebrew exclusively in the Brest oblast (more than 10 times). According to the results of the 2009 census, there were no Jews among the rural population in any district of the Brest oblast who called Hebrew their native language. According to the 2019 census, they appeared in 12 out of 16 districts, despite the fact that among all other districts of Belarus, the census showed their presence only in 2. As for the languages of home communication, only the Brest oblast also stands out here, with a 2.9-fold increase in the share of the Belarusian language (Table 6).

Table 20. Administrative units with the largest increase in the urban population of Jews calling Belarusian their native language, pers.

Region	Population size		Belarusian native		Belarusian home		Russian native		Russian home		Hebrew native	
	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019
Brest oblast												
Brest	234	328	24	68	6	42	183	169	215	282	8	86
Baranovichi	116	267	16	85	3	13	100	75	113	254	0	104
Pinsk	146	177	15	44	1	0	120	66	138	176	5	63
Vitebsk oblast												
Orshansky district	272	188	34	61	4	1	231	127	266	187	2	0
Polotsky district (without Novopolotsk)	238	157	13	67	1	0	211	88	233	157	8	0
Vitebsk	1 315	912	89	303	11	5	1 194	595	1 292	903	12	4
Novopolotsk	108	130	4	49	0	3	98	75	105	127	4	2
Gomel oblast												
Gomel	1 684	1 281	135	182	26	9	1 463	1 083	1 623	1 267	27	10
Mozyr	167	215	28	66	4	1	135	140	161	211	1	2

Source: (Belstat 2020).

Table 21. Change in the population size of Jews and in the number of those calling Russian, Belarusian and Hebrew their native and domestic language among the rural population, pers.

Region	Population size		Belarusian native		Belarusian home		Russian native		Russian home		Hebrew native	
	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019
Brest oblast	18	86	3	31	1	5	15	21	16	81	0	33
Vitebsk oblast	63	89	9	50	5	3	51	38	56	85	0	0
Gomel oblast	54	103	8	31	4	4	43	65	48	97	1	0
Grodno oblast	15	99	4	3	4	2	11	95	11	97	0	0
Minsk oblast	131	622	20	19	9	11	102	581	120	610	4	4
Mogilev oblast	34	46	7	1	4	1	23	41	29	22	0	0
Belarus	315	1 045	51	135	27	26	245	841	280	992	5	37

Source: (Belstat 2020).

The number of *Tatars* shows an abrupt increase of 15.0% between the 2009 and 2019 censuses (in 1989-2009 it decreased by 42.1%), largely due to regions where Belarusian Tatars have not historically lived, while in the Grodno oblast, which is just such a region, their number decreased by the maximum value (Figure 8; Table 22). In the Mogilev oblast, where the share of Tatars in 2009 was minimal, the increase in their number in 2019 was maximum. Also, the number and proportion of the population younger than working age and the child-woman ratio increased many times over, and a close negative correlation was recorded between the population younger than working age in 2009 and its relative increase in 2019 (r = -0.77, p <0.05).

Table 22. Increase in the number of Tatars by region

Region	Population size			Including those below working age		
	2009	2019	+/–, %	2009	2019	+/–, %
Belarus	7 329	8 445	+15.2	370	1 132	205.9
Brest oblast	725	773	+6.6	31	92	196.8
Vitebsk oblast	822	1 033	+25.7	23	137	495.7
Gomel oblast	776	1 030	+32.7	27	149	451.9
Grodno oblast	1 710	1 385	–19.0	160	168	5.0
Minsk	1 558	2 240	+43.8	70	309	341.4
Minsk oblast	1 239	1 201	–3.1	44	154	250
Mogilev oblast	499	783	+56.9	15	123	720

Source: Author's calculations based on Belstat data.

Comparison of the dynamics of the urban population of the Tatars and the share of the Belarusian language as native language and home language showed clearly individual regional features of changes in this set of indicators. The Brest and Vitebsk oblasts are characterized by an increase by tens of percent of the population of Tatars with a simultaneous increase by hundreds and thousands of percent of the share of Tatars calling Belarusian their native and home language. The Gomel oblast differs from them only in the stabilization of the population in large regional centers with a sharp increase in its population in the regional center and the absence of an increase in the share of the Belarusian language as a home language. At the same time, in a number of districts with a decrease or slight increase in the urban population of the Tatars the share of the urban population that called Belarusian their native language increased sharply: in the Zhlobin district from 0 to 21.1%, in the Mozyr district from 0 to 22.2%, and in the Rechitsa district from 1.5 to 21.8%.

In the Grodno and Minsk oblasts, in not a single district with a population of more than 15 Tatars does the share of Belarusian increase either as a native or as a home language, while the number of Tatars everywhere decreases or is stable. In Minsk, the number of Tatars increases noticeably, with a significant decrease in the share of the Belarusian language. In the Mogilev oblast, the number of Tatars increases by tens of percent and the shares of the Belarusian language continue to remain near zero, with a sharp increase in the number of Tatars who did not indicate their native language during the census. Also, the Mogilev oblast is the only one where the share of the population that named Tatar as their home language increased, and by 4.6 times at once.

A more detailed picture for the urban population is provided by Table 23, the data of which show sharp differences in the dynamics of the share of the Belarusian language in various oblasts and huge figures for the increase in the share of the Belarusian language in large cities of the Brest, Vitebsk and Gomel oblasts.

The cities listed accounted for 20% of the urban population of Tatars in 2009; in 2019 this figure increased to 60%. Excluding the city of Minsk, these values were 9 and 30%, respectively.

Features of the dynamics of the number and share of the Belarusian language for the rural population of Tatars by region (Table 24) have some similarities with the urban ones: in the Brest, Vitebsk and Gomel oblasts, the share of Belarusian as a native language is increasing, while the population is decreasing; in the Grodno oblast, both of these indicators are declining, and in the

Minsk and Mogilev oblasts, the increase in population is accompanied by a decrease in the share of the Belarusian language. The share of the Tatar language as a native and home language continues to be insignificant.

Table 23. Administrative units with the largest growth of the Tatar population for the urban population (only those where the number in 2009 exceeded 15 people were taken into account), pers.

Region	Population size		Belarusian native		Belarusian home		Russian native		Russian home		Tatar native		Tatar home	
	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019
Brest oblast														
Brest	225	271	4	120	0	119	178	101	222	151	37	43	0	1
Baranovichi	101	129	7	59	4	53	73	54	95	76	19	16	0	0
Pinsk	38	47	3	25	0	25	26	14	38	22	9	8	0	0
Vitebsk oblast														
Orshansky district	54	144	1	80	1	50	39	20	53	94	12	41	0	0
Polotsky district (without Novopolotsk)	86	109	1	56	2	30	68	30	86	79	16	22	0	0
Vitebsk	201	408	6	224	3	117	144	107	197	289	45	67	0	0
Novopolotsk	83	72	1	32	1	19	64	27	80	53	16	12	1	0
Gomel oblast														
Gomel	235	514	7	216	0	1	180	236	225	511	41	61	2	0
Minsk oblast														
Minsk	1558	2240	201	138	62	61	1 057	1 812	1 442	2 174	238	266	2	2
Mogilev oblast														
Mogilev	127	306	2	0	2	0	91	213	125	258	33	79	0	22
Bobruysk	147	237	7	2	2	0	97	122	140	203	40	51	5	16

Source: (Belstat 2020).

Table 24. Change in the population size of Tatars and of the number of those calling Russian and Belarusian their native and home language among the rural population, pers.

Region	Population size		Belarusian native		Belarusian home		Russian native		Russian home	
	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019
Brest oblast	167	163	9	35	12	37	128	104	153	125
Vitebsk oblast	181	151	39	51	41	37	103	71	136	113
Gomel oblast	198	170	8	29	8	3	139	111	185	167
Grodno oblast	195	128	52	20	50	13	113	86	129	116
Minsk oblast	407	507	63	29	56	27	266	366	345	479
Mogilev oblast	107	117	4	2	3	2	73	84	103	111
Belarus	1255	1236	175	166	170	119	822	822	1 051	1 111

Source: (Belstat 2020).

The number of *Lithuanians* in Belarus decreased from 1989 to 2009 by 33.1%. The 2019 census recorded its growth from 2009 to 2019 at 3.9%. At the same time, the change in numbers differs sharply by region and age group (Table 25). The traditional region of residence of Lithuanians is the Grodno oblast, where a little less than half of the entire Lithuanian population of Belarus is located. In it, the number of Lithuanians practically did not change. A sharp increase occurred only in the Brest, Vitebsk and Gomel oblasts, where the number of Lithuanians below working age more than doubled.

The calculation of the correlation coefficients showed a very close relationship between the growth of the total size of the Lithuanian population by region and the growth in the proportion of Lithuanians calling Belarusian their native language ($r = 0.93$, $p < 0.05$). A positive relationship is also observed between the growth of the Lithuanian population below working age and the proportion of Lithuanians calling Belarusian their language of home communication ($r=0.71$, $p<0.05$). The share of those who named Lithuanian as their native language fell significantly in all regions, except for the city of Minsk, where it changed insignificantly.

Table 25. Increase in the number of Lithuanians by region

Region	Population size			Including those below working age		
	2009	2019	+/–, %	2009	2019	+/–, %
Belarus	5087	5287	+3.9	426	623	+46.2
Brest oblast	355	450	+26.8	11	38	+245.5
Vitebsk oblast	624	944	+51.3	27	60	+122.2
Gomel oblast	271	388	+43.2	18	43	+138.9
Grodno oblast	2153	2174	+1.0	245	291	+18.8
Minsk city	935	669	–28.4	74	101	+36.5
Minsk oblast	476	462	–2.9	31	62	+100.0
Mogilev oblast	273	200	–26.7	20	28	+40.0

Source: Author's calculations based on Belstat data.

As for those calling Lithuanian their home language, a sharp increase in their number occurred only in the Brest oblast (from 2 to 101 people). In other regions, except for the Grodno oblast, such persons are almost entirely absent both in 2009 and in 2019. In the Grodno oblast, where they represent a significant share of the Lithuanian population, their number nonetheless decreased twofold from 2009 to 2019.

If we analyze the change in the number and language characteristics of Lithuanians at a lower level of the administrative division (among the urban and rural population of the districts), we see that a maximum increase in their number is characteristic of the urban population of several administrative units listed in Table 26. The total number in these administrative units increased by 85.7% (for the whole of Belarus, by 3.9%, and for the entire urban population, by 13.2%). While in 2009 the urban population of the 26 regions indicated in the table was 13.7% of the total number and 21.2% of the urban population of Lithuanians, in 2019 it was 24.5 and 37.7%, respectively, and most of these regions did not belong to the areas of their historical residence.

Also revealing is the increase in these administrative units of the number of people calling Belarusian their native and home language. These indicators, with the exception of the regions of the Grodno oblast, increased by hundreds and thousands of percent. Among the urban population of a number of other regions (districts with large regional centers), these indicators also increased sharply. Thus, in the city of Pinsk, the number of Lithuanians calling Belarusian their native language increased from 4 to 16, and of those calling it their home language, from 1 to 12. Elsewhere, the corresponding figures were: in Novopolotsk, from 5 to 38 and from 0 to 13, respectively; in the Orsha district, from 5 to 46 and from 2 to 14; in Grodno, from 53 to 164 and from 21 to 60; in the Borisov district, from 12 to 27 and from 4 to 21; in Mogilev, from 19 to 29 and from 7 to 29; in Bobruysk, from 7 to 21 and from 4 to 18. A similar picture is observed among the rural population of the Minsk district (where a significant part of the rural settlements is actually suburbs of Minsk), from 8 to 44 and from 8 to 32.

The total urban population of Lithuanians in the regions considered in Table 26 who called Belarusian their native language increased by 353%, and of those calling it their home language, by 283%. (For comparison: among the Lithuanians of Belarus as a whole, these figures increased by 86.0 and 13.1% respectively, and among the urban population, by 116 and 78%). The share of this population among all Lithuanians of Belarus calling Belarusian their native language and home language increased from 20.0 to 48.7% and from 11.1 to 37.4%, respectively, and among the urban population alone, from 32.5 to 68.1% and from 28.6 to 61.5%, respectively.

Table 26. Administrative units with the largest increase in Lithuanians for the urban population (only those where the number in 2009 exceeded 15 people were taken into account), pers.

	Population size		Belarusian native		Belarusian home		Russian native		Russian home		Lithuanian native		Lithuanian home	
	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019	2009	2019
Brest oblast														
Brest	105	166	18	101	8	60	61	45	95	64	20	17	0	42
Baranovichi	75	97	12	53	10	29	36	25	62	45	21	18	1	23
Vitebsk oblast														
Vitebsk	120	366	18	284	7	93	61	61	109	273	38	20	1	0
Polotsky district (without Novopolotsky)	39	90	2	64	1	22	24	16	37	67	10	6	0	1
Postavsky district	43	62	3	29	5	15	20	23	33	47	18	10	1	0
Gomel oblast														
Gomel	101	172	13	115	0	35	61	41	98	136	24	14	0	1
Mozyr	28	39	4	26	1	11	17	9	25	28	6	3	0	0
Zhlobinsky district	17	37	4	26	0	11	10	7	17	26	3	4	0	0
Grodno oblast														
Voronovsky district	75	133	45	51	39	33	18	70	32	99	10	10	0	0
Oshmyany	30	49	11	12	7	4	14	30	23	45	4	7	0	0
Smorgon	64	83	19	41	12	13	24	29	41	69	19	13	3	1

Source: (Belstat 2020).

Conclusions

An analysis of the dynamics of the ethnolinguistic structure of the population of Belarus as a whole and by region shows a number of anomalous changes which ultimately lead to an overestimation of the population size and the proportion of the population calling Belarusian their native and home language. What is striking about these anomalous changes is the pronounced multidirectional dynamics of the indicators of the ethnolinguistic structure in the regions of Belarus (which was not observed according to the results of previous censuses) and the fact that the changes are found in most administrative districts of the corresponding regions.

In all regions for all nationalities where anomalous population growth was observed, the number of persons below working age was simultaneously increasing, by many times more than the total population in these regions and than the population below working age in all other regions.

Abnormally abrupt changes in the characteristics of the ethnolinguistic structure in 1-2 regions provide a small change in the same direction of these characteristics throughout the country as a whole. For example, an increase in the share of Belarusian as a native language among the entire population and among ethnic Belarusians only in the Brest oblast (respectively +23.3 and +26.6%) and the city of Minsk (+13.5 and +11.7%) ensured its slight increase among similar categories of the population of Belarus as a whole (+0.9 and +0.4%, respectively).

The increase in the share of Russians calling Belarusian their native language, seen only in the Brest oblast (by 8.7%), also ensured a slight increase in the country as a whole of the share of Russians calling Belarusian native (by 0.1%). An increase in the proportion of Ukrainians calling Belarusian native in only the Brest (by 15.4%) and Vitebsk (by 7.8%) oblasts led to an increase in this indicator for the country as a whole of 0.3%. The countrywide increase in the share of Belarusian as their native language among the Jews and Tatars (by 2.4 and 1.9%, respectively) resulted from a significant (up to 27.3% for Jews and up to 33.0% for Tatars) increase in this indicator in the Brest, Vitebsk and Gomel oblasts. A very significant increase in the proportion of Ukrainians calling Ukrainian their native language in the Mogilev oblast (+40.1%) ensured that the value of this indicator among the entire Ukrainian population of Belarus remained almost at the same level (it decreased by only 0.1%). The above-listed sharp increases in the share of the Belarusian and Ukrainian languages in these regions, accompanied by an equally sharp decrease in the share of the Russian language, led to the fact that in the country as a whole the share of the Russian language as a native language among the entire population, Belarusians, Russians and Ukrainians, increased slightly (by 0.8, 1.1, 0.5 and 1.2%, respectively), although in other regions this increase was in some cases an order of magnitude higher. The share of Hebrew as a native language for the Jewish population increased in Belarus by 1.3% solely due to its increase by 30.1% in the Brest oblast.

The increase in the share of Belarusian as a home language among the entire population of the country (+2.6%) and among ethnic Belarusians (+2.4%) resulted from a sharp increase (5.9 and 5.5 times, respectively) in this share among the population of Minsk, while the increase among Russians and Ukrainians (+0.4 and +0.3%) was due to the Brest oblast alone (+9.3 and +12.4% respectively). A similar picture is observed for Jews and Tatars: slight changes in the share of Belarusian as a home language (for Jews +0.1%, for Tatars -0.9%) are due to a significant increase in this share in the Brest oblast (+4.3 and +12.0 %). For Russian, as both home and native language, there is a slight increase in its share for the entire population and individual nationalities in Belarus as a whole due to a sharp decrease in the share in some regions, while in other regions its increase is many times greater. The share of other ethnic languages as home languages also changed by an insignificant amount (by no more than 0.8%) in the country as a whole, despite sharp changes in certain regions. For example, among Lithuanians, the share of Lithuanian as a home language decreased by 0.8% in the country as a whole, with an increase in the Brest oblast of 21.8%, while among the Tatars the share of the Tatar language as a home language increased by 0.2% nationwide with an increase of 5.1% in the Mogilev oblast alone, where the number of Tatars was the highest of all regions - +56.9%.

Along with a 0.95% decrease in the total population of the country, the number of ethnic Belarusians increased by 0.42% due to the increased number of Belarusians below working age in Minsk, an increase of 38.4% compared to 2009, 7.4 times higher than the value for all other regions as a whole (5.2%). And it was in Minsk, as mentioned above, that there was a sharp and simultaneous increase in the share of Belarusian as both native and home language.

A characteristic feature of the anomalous changes in ethnolinguistic characteristics in individual regions is that they also occurred in most or all areas of the respective regions. Thus, in the Brest oblast, the number of districts where the share of Belarusian as a native language among the urban population of Belarusians increased by more than 5% from 2009 to 2019 is 14 out of 16, while in the Vitebsk and Gomel oblasts there are 2 such districts each, and in the Mogilev, Minsk and Grodno oblasts - not a single one. Among the rural population, the number of districts where, in principle, there was an increase in the share of Belarusian as a native language is 13 out of 16, while among all other districts of Belarus - only 4 out of 102. Among the urban population of Ukrainians, the number of districts with a share of the population calling Belarusian native greater than 10% is 26, of which 22 are located within the Brest and Vitebsk oblasts (despite the fact that in 2009 there were only 15 such districts out of a total of 59). Among the rural population in 2019, there were 20 out of 26 such districts in these regions, and in 2009 - 21 out of 81. Among the urban population of Ukrainians, the number of districts with a share of the population calling Ukrainian native greater than 50% is 17, of which 15 are in the Mogilev oblast (in which there are 21 districts in total), whereas in 2009 there was not a single such district in the Mogilev oblast and only 1 in the entire republic.

The 2009 census showed neither abnormally sharp changes in ethnolinguistic indicators in several regions opposite to those in other regions, nor a strict administrative localization of these changes in a few districts that differed greatly from most other districts. In our view, it is not possible to explain this situation by any factors other than administrative influences (for example, migration, inter-ethnic marriages, a change of ethnic identity, the "salmon effect", etc.) affecting the census results. Unlike in large or mountainous countries and regions, there are no fundamental ethno-cultural differences between the regions of the small, flat Belarus, and besides, the anomalous changes are localized within the boundaries of administrative rather than ethno-cultural regions. The factor of changing ethnic self-identification cannot explain why there is a noticeable increase in the number of Russians only in the Brest and Vitebsk oblasts when it is decreasing in the country as a whole. (Besides, in Belarus this factor causes the change of self-identification of Russians to Belarusians, and not vice versa, as in Russia). Moreover, the number of Russians in the Brest oblast indicating Belarusian as their home language increased from 2009 to 2019 by 573% (!), while in the Vitebsk oblast it decreased by 61% (whereas the number of Belarusians in the Brest oblast indicating Belarusian as their home language decreased by 46%). Nor can this factor explain the noticeable increase in the number of Poles in the Gomel and Mogilev oblasts when it is decreasing in the country as a whole (and especially the 510.5% growth in the Gomel oblast of the Polish population below working age), since such processes can occur mainly in the Grodno oblast bordering Poland, where 77.6% of Belarusian Poles live and where their number has decreased; and for Lithuanians, whose number has increased, especially in the Gomel and Vitebsk oblasts, a change of ethnic self-identification is not typical at all. Nor did these regions see any significant migration of Poles and Lithuanians from either Poland and Lithuania or from the Grodno oblast. Neither has any migration of Jews to Belarus been recorded, making it difficult to explain a noticeable increase in their numbers in several regions (especially of those below working age, for example in the Minsk oblast, where their number increased ten-fold) and an increase, moreover by 13 times, in the share of Hebrew as a native language exclusively in the Brest oblast; even among the rural population of the Brest oblast, cases of Hebrew being indicated as their native language were noted in 12 out of 16 districts (in 2009, such a population was completely absent), despite the fact that in the rest of Belarus this was recorded only in 2 districts. Migration of Russians and Belarusians from other CIS countries to Belarus took place mainly in the 1990s and 2000s. In the 2010s, not only was its scale markedly smaller, but it was directed to all regions; it is unlikely that a significant part of such a population would indicate its home language as Belarusian. And most importantly, when considering factors that might explain the unusual census results, it should be borne in mind that they would have had to start appearing only between 2009 and 2019 in order to have such a significant impact and explain the results of the census only in 2019.

On the whole, the scale of false attributions of those calling Belarusian their home language is higher than of those calling it native, since the real drop in the proportion of the population calling Belarusian their home language is much more pronounced, and thus to ensure the growth of the share of Belarusian as both a native and home language, the latter requires a greater degree of inflation, which is clearly seen in the example of Minsk.

The increase in the population size and in the number of those calling Belarusian both their native and home language, caused by an artificial inflation of these data, is much more noticeable for nationalities that are characterized in Belarus or in certain regions by a relatively small number. It is for such groups that the anomalous increase (positive or negative), absent in previous censuses, reaches hundreds of percent and leads to such results as, for example, in the Gomel oblast, where, in 2009, among the Polish population an average of 14 women of childbearing age accounted for 1 child aged 0-9 years, while in 2019 those same 14 women already accounted for 5 children, or in the Vitebsk oblast, where the increase in the number of Lithuanians of working age in 2019 in absolute terms is 10 times higher than their number below working age in 2009. Although such changes for ethnic groups which represent an extremely small proportion of the entire population of Belarus do not particularly affect the overall picture, they are, firstly, an indicator of the presence of administrative distortion of the results, and secondly, can have socio-political significance, demonstrating that ethnic diversity is preserved and that small ethnic groups also consider the Belarusian language valuable and use it in everyday life, that they are an integrated part of the Belarusian civil nation who have nonetheless not lost their national languages in favor of Russian.

Thus, we can conclude that the results of the census do not show real changes in the number and distribution of the languages of the population of Belarus and of individual nationalities, but rather are the result of distortions aimed at obtaining a politically desired result.

References

Anarbek N. (2019). Sociocultural mechanisms for the reproduction of society's spiritual values. *Sociological Almanac*, 10, 106-116. https://socio.bas-net.by/wp-content/uploads/2020/05/Sotsiol_Almanah_Vyp-10.pdf

Bahlova O.V., Bahlov I.V. (2020). Identity politics in the context of nation-building and integration interaction (on the example of the Union State of Belarus and Russia). *Regionology*, 4, 723-753. https://doi.org/10.15507/2413-1407.113.028.202004.723-753

Belstat (2019). *Education in the Republic of Belarus*. https://www.belstat.gov.by/upload/iblock/02f/02f0dcce5ea8e20041bca7728366684c.pdf

Belstat (2020). *Final data of population censuses of the Republic of Belarus.* https://census.belstat.gov.by

Goritskaya O.S. (2021). *Language and boundaries: Lexical and grammatical specifics of the Russian language in Belarus*. Minsk: MSLU. https://e-lib.mslu.by/bitstream/edoc/8798/1 Горицкая_Язык%20и%20границы%20_2021.pdf

Linchenko A.A., Belyaeva E.V. (2020). The Union State - a common memory? Ethical foundations of the historical policy of Russia and Belarus in the XXI century. *Discourse-Pi*, 2(30), 97-111. https://doi.org/10.24411/1817-9568-2020-10206.

Mechkovskaya N.B. (2003). *The Belarusian language: sociolinguistic essays*. Munich: Verlag Otto Sagner. https://library.oapen.org/bitstream/id/f54402c6-d761-487e-a54e-fbc5a4b570de/1003951.pdf

Minstat (2001). The national composition of the population of the Republic of Belarus and the prevalence of languages. Volume 1. Minsk: Min. stat. and analysis.

Rakachev V.N., Khalafyan A.A. (2012). Estimation of the population of the Kuban in the 1930s-1940s using statistical modeling methods. *Historical and socio-educational thought*, 5, 150-156. https://elibrary.ru/item.asp?id=18244715

Manakov A.G., Sokolov A.S., Suvorkov P.E. (2022). Formation of the ethnic and linguistic structure of the population of the Republic of Belarus and an ethnographic forecast up to 2069. *The Pskov Regional Journal*, 1, 35-55. https://elib.gsu.by/handle/123456789/38766

Sokolov A.S. (2021). Regional aspects of the dynamics of the use of the Russian language by the population of Belarus in 2009–2019. *Demographic review*, 8 (1), 127-149. https://doi.org/10.17323/demreview.v8i1.12396

Khamutovskaya S.V. (2021). Russian and Belarusian languages in the "mirror" of public opinion of the inhabitants of the Belarusian region of the Russian-Belarusian border area. In M.E. Lustenkov (Ed.), *The Russian-Belarusian borderland: linguistic and socio-cultural aspects of interaction* (pp. 207-209). Mogilev: BRU.http://e.biblio.bru.by/handle/1212121212/18178

Hentschel G. (2017) Eleven questions and answers about the Belarusian “trasyanka”. *The Russian language in a scientific light,* 1, 209-250. http://rjano.ruslang.ru/ru/archive/2017-1/209-250

Khentschel G., Kittel B. (2011) The Language Situation in Belarus: Belarusians' Opinion on the Prevalence of Languages in the Country. *Sociology*, 4, 62-78. https://elib.bsu.by/bitstream/123456789/39989/1/62-78.pdf

Shimov V.V. (2020) The language issue as a factor of potential political instability in Belarus. *Journal of Belarusian State University*. Sociology, 1, 67-70. https://journals.bsu.by/index.php/sociology/article/view/2903/2864