ABOUT MORTALITY DATA FOR RUSSIA

by Vladimir M. Shkolnikov and Dmitri Jdanov 15 September 2006

GENERAL

In Russia the most detailed data on deaths and population size by single age have never been published.

Between the 1950s and 1973, published statistical annuals on population included only aggregate indices like the total (all ages combined) number of deaths, births, and population, or sometimes disaggregated by very broad age categories (e.g., children, working ages, retired ages).

In 1974-1986, any publication of mortality data was forbidden in the Soviet Union for ideological reasons. In 1986-87, in a completely new era of M. Gorbachev's "glasnost", the Central Statistical Office resumed publication of mortality statistics in demographic yearbooks. These data were much more detailed than those of the 1960s since they included deaths and population by sex and ages 0, 1-4, 5-9, ... 85+. This practice continues now without significant changes.

Since 1946, the Soviet (Russian) statistical system has been producing annual tables (unpublished) of deaths and population by sex and single-year age group.

These statistical tables have been used for various statistical purposes (population projections, life tables etc.), but they have not been published. Demographers had (and have) access to these data by request from the Central Statistical Office of the Russian Federation (currently named “Goskomstat”).

Source of data

Most of the data included in the mortality database were received from E. Andreev and S. Zakharov. The population estimates for the period 1946-59 were made by E. Andreev (Andreev et al., 1998). Birth counts were published in the Demographic Yearbook of Russia (Gomkomstat, 2000). The Input Database (InputDB) death and population numbers cover each year since 1946.

TERRITORIAL COVERAGE

The following territorial changes have taken place in Russia during the period covered by the data:

- Crimea region was excluded from Russia and included in Ukraine in 1956
- Kaliningrad, Tuva, and Karel provinces were included in Russia after the end of the Great Patriotic war in 1945.

S. Zakharov (see his research note in Appendix 2) made adjustments for these territorial changes. All data in the Input Database are for Russia with

Crimea region excluded and with Kaliningrad, Tuva, and Karel provinces included.

In addition, there were some changes in the coverage of vital statistics although they were not official territorial changes (i.e., the national boundaries of Russia did not change). For these changes in population coverage, we use the methods described in Appendix D of the *Methods Protocol* to make the appropriate adjustments in the formulas that calculate population estimates and death rates:

Dates	Population coverage	Area Code†
1959-1992	Vital statistics and official population estimates cover the entire population of Russia.	1
1993-1994	Vital statistics exclude the Chechen-Ingush Republic. Although the official population estimates include this region, but we have adjusted these estimates to exclude the Chechen-Ingush Republic for the purposes of our calculations.	11
1995-2002	Vital statistics exclude the Chechen Republic. Again, the official population estimates include this region, but for the purposes of our calculations, we have adjusted these estimates to exclude the Chechen Republic.	12
2003-2005	Vital statistics and official population estimates cover the entire population of Russia.	1

† The area code is used in the raw data files (Input Database) to denote the geographic area covered by the data.

DEATH COUNT DATA

Coverage and Completeness

Registration of deaths in Russia by the centralised state civil registration system was organised in the 1920s. It included the Central Statistical Office, named TCUNHU SSSR/Central Office for the Statistics of the National Economy (later renamed TCSU SSSR/Central Statistical Office of the USSR and then again renamed Goskomstat/State Statistical Committee of the Russian Federation), regional ("oblast") statistical offices, and district statistical bureaus (ZAGS). This system is highly centralized. District statistical bureaus register vital events and send copies of individual records to the regional statistical offices. Regional statistical offices (with the help of standard computer programs since 1988) assemble individual death records and transform them into standard summary tables going to the Central statistical office in Moscow. The coverage of the whole population by this system became complete about the mid-1930s.

The annual statistical table named "forma 4" (computer table S41 since 1988) includes deaths by sex and single-year age group for Russia as a whole and for each of Russia's oblast-level regions.

For the period between the end of the Second World War and the mid-1960s, there has been a concern in the demographic literature about the quality of infant and old-age mortality values in Russia and other parts of the former Soviet Union (Anderson and Silver, 1986, Shkolnikov, Meslé, Vallin, 1996). For Central Asia and Transcaucasia, the problem was very significant

even at the end of the Soviet era in the late 1980s. However, for Russia, the problem was radically reduced by the end of the 1960s. There are some signals from statisticians and researchers suggesting that in the most recent years, the problem of death under-registration at very old ages might be growing. But (to our knowledge), there is still no scientific evidence on the issue.

A substantial proportion of births with a higher risk of infant death are excluded from live births according to a restricted definition of live birth which results in incomparability with countries using the WHO definition of live birth (see section "Birth Count Data" for more details).

Death statistics for Russia do not include deaths in Chechen-Ingush Republic in 1993-1994 and deaths in Chechen Republic in 1995-2002.

Specific Details

- There is a concern about age misreporting on death certificates mostly at ages 70, 80, and 90 before 1970 (see S. Zakharov's note below for more details). Significant age heaping and problems with erratic ages may exist even now at very old ages. Unfortunately, no systematic individual-level validation of the quality of death records for deaths at old ages has been done.
- S. Zakharov has detected a specific problem of death heaping at age 99 (see his note below for more details). We can only guess about the nature of the problem. It might be due to some computational procedures by the Goskomstat, which (perhaps) allow exchanges between age 99 and the next age group 100+. To avoid this problem in our estimates, we aggregated deaths in the open age interval 99+ for the period 1959-1989.

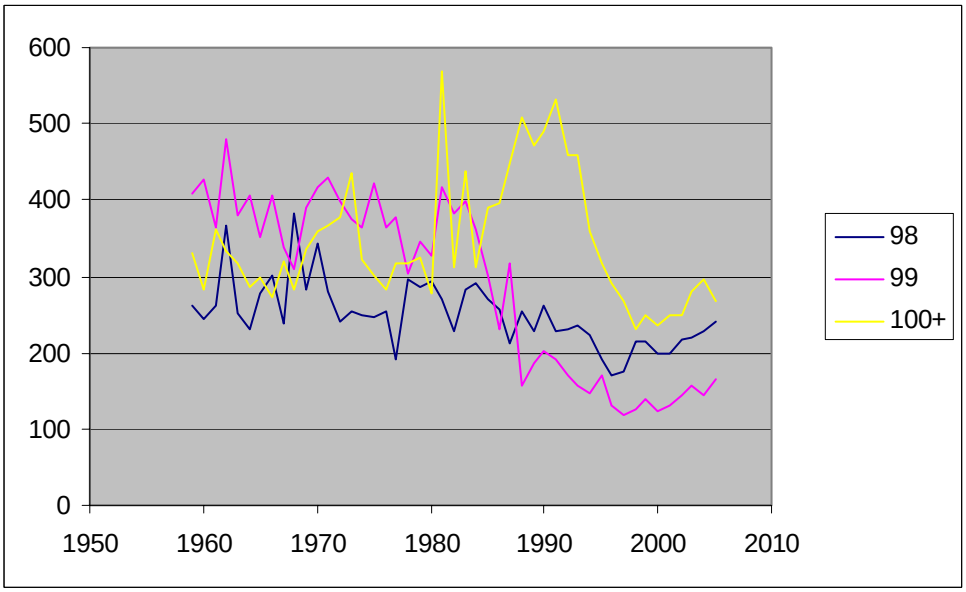


Figure 1. Figure 4. Number of deaths at ages 98, 99 and 100+, Russia, males, 1959-2005

POPULATION COUNT DATA

Coverage and completeness

Population estimates are based on the all-Soviet censuses of 1959, 1970, 1979 and 1989. Censuses were conducted on the 15th of January in 1959 and 1970, and on the 17th of January in 1979 and 1989. From the census counts, the Goskomstat produced the official population estimate as of January 1.

Specific Details

Before the first after-war census of 1959, very little was known about the Russian population. In 1998, Andreev et al. (1998) completed a reconstruction of population by age for the period 1946-58. These estimates are included in the "raw data" files, but they were not used for the construction of the HMD mortality estimates. In the 1960s, the Central Statistical Office did not produce official estimates of population by age. The all-Soviet census of 1959, the micro-census of 1964, and the census of 1970, along with (incomplete) current data on births, deaths, and migrations provided grounds for an estimation of population. Since the census of 1970, the Central Statistical Office has produced official estimates of population by age on an annual basis from census counts, deaths, births and migrations. After the censuses of 1979 and 1989, the population figures within respective inter-census periods have been re-calculated on the basis of the new census data.

For population counts at advanced ages, there is a problem of age heaping and age overstatement (see S. Zakharov's note below). In addition, there is a particular problem of over-registration of centenarians (Garson, 1991).

As noted in the section "Territorial Coverage", deaths in the Chechen Republic were excluded from official statistics of Russia in 1993-2002, although official population estimates are available for Russia as a whole (including the Chechen Republic). To avoid inconsistency between the numerator and denominator, we used population estimates in 1993-2002 excluding the Chechen-Ingush Republic (1993-1994) and the Chechen Republic (1995-2002). Thus, for this period our estimates do not refer to the whole of Russia. Although these changes in coverage do not represent official territorial changes, they are treated as such in order to make the appropriate adjustments to the formulas. We should note that the adjustment factor for the inclusion/exclusion of the Chechen Republic varies substantially across age. The Chechen Republic comprises less than 1% of the total population of Russia, but at very old ages (100+) the comparable percentage is more than 10%. This indicates a serious age overstatement problem in the Chechen Republic and (perhaps) some other Muslim regions of the North Caucasus (Coale and Kisker, 1986).

BIRTH COUNT DATA

Coverage and completeness

The registration of live births in Russia (like in many post-Soviet and former communist countries) differs from conventional western practices and the WHO recommendations. According to the Soviet definition of live birth (launched before the Second World War), a live birth is to be officially registered by the statistical system if the gestation period is 28 weeks or longer, the body mass at birth is 1000 g or higher, the body length is 35 cm or longer, and the newborn breaths. Such a restrictive rule leads to underestimation of births and population at age 0 and also to underestimation of neonatal mortality by about 50% and infant mortality by about 25% (Anderson and Silver, 1986, Blum and Monnier, 1989, Velkoff and Miller, 1995, Andreev, 1995).

DATA QUALITY ISSUES

The data prior to 1970 should be used with extra caution due to problems of data quality.

1. The problem of death age heaping is still one of the main problems of Russian data, especially for the period before 1970. We have not attempted to solve this problem, and leave it open to researchers who would particularly focus on it. Therefore, use of the Russian data on number of deaths and mortality rates by one-year age groups, should be done carefully.
2. There are serious fallacies in the official estimation of the population of elderly men and women reported by the RF GOSKOMSTAT, especially for persons over 80. As a result, elderly population changes and, correspondingly, mortality rates, become highly irregular and unjustified. The older people are, the more problematic the official data are. These fallacies could be of methodological character and, most probably, have to do with use of the wrong model of approximation of very old people's migration. These problems are partly solved in the HMD estimates by using extinct/almost extinct cohort method.

   The sex ratio at ages 90+ also suggests data quality problems with the population estimates at very old ages during recent years. For example, life expectancy[1] at age 90 in 2005 for males is higher than that for females (3.38 vs. 3.28 years, respectively), which seems implausible. For most population estimates at ages 80+, the HMD methodology uses the extinct cohort method, which may provide more accurate estimates than the official population estimates. Nevertheless, for more recent years when older cohorts are not yet “extinct”, our methods for estimating population rely on the official estimates. Thus, data quality problems with these

[1] Calculated by the HMD Methodology using official population data.

estimates influence the quality of HMD estimates for the previous 10-15 years (see the *Methods Protocol* for more details).

Our analyses of data suggest that the male population remains inflated even after excluding the Chechen Republic from the all-Russia population. Moreover, the problem can not be attributed to increasing numbers of deaths at unknown age in recent years. Thus, it appears to be a systematical error, which may also (to some degree) affect the data for females. To avoid this problem, we aggregated the official population estimates to make the open age interval 80+. This approach has been used by Statistical Offices in similar cases. For example, since 2001 Statistisches Bundesamt (West Germany) has provided population estimates with open age interval 90+ instead of 95+, and Statistics Lithuania has published data with open age interval 85+ instead of 100+ since 2003. We use the survival ratio method to split the population totals for ages 80+ instead of 90+ (see the *Methods Protocol* for details). Correspondingly, we applied extinct/almost extinct cohort methods for all ages above 80. Using an open age interval 80+ produces more plausible results (see Figure 2).

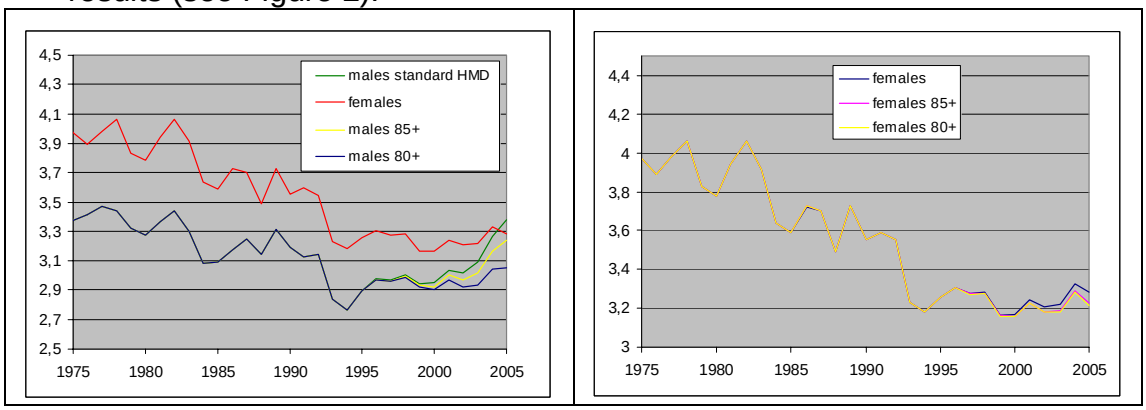


Figure2. Life expectancy at age 90 estimated based on various open age intervals in 1975-2005. [Note: "Standard HMD" refers to calculations using the SR(90+) method (i.e. open age interval 90+).]

For more details about data quality issues see Appendix 2.

INPUT DATABASE AND MORTALITY SURFACES OF THE HMD

Although all the Russian data included into InputDB may be of interest for users, we have decided to make further estimates of mortality surfaces only for the period since 1959 with the last age group 99+. Nevertheless, data for 1959-1969 should be used with caution due to the problems of data quality mentioned above (see also the note by S. Zakharov in Appendix 2).

REFERENCES

Anderson B.A., Silver B.D. (1986). "Infant mortality in the Soviet Union: regional differences and measurement issues." *Population and Development Review*, Vol. 12, No 4, pp. 705-738.

Andreev E. (1995). "Mladencheskaya smertnost v Rossii" (Infant mortality in Russia). *Vestnik Statistiki*, Vol. 5, pp. 66-71.

Blum A., Monnier A. (1989). "Recent mortality trends in the USSR: new evidence." *Population Studies*, Vol. 43, pp. 211-241.

Coale, A.J. and E.E.Kisker. 1986. Mortality crossovers: reality or bad data? *Population Studies*, 40: 389-401.

Garson L.K. (1991). "The centenarian question: old-age mortality in the Soviet Union, 1897 to 1970." *Population Studies*, Vol. 45, pp. 265-278.

Shkolnikov V.M, Meslé F., Vallin J. (1996). "Health crisis in Russia. I.Recent trends in life expectancy and causes of death from 1970 to 1993." *Population: An English Selection*, Vol. 8, pp. 123-154.

Velkoff V., Miller J.E. (1995). "Trends and differentials in infant mortality in the Soviet Union, 1970-90: How much is due to misreporting?" *Population Studies*, Vol. 49, pp. 241-258.

REFERENCES TO SOURCE OF DATA

Goskomstat. (2000). *Demograficheskiy ejegodnik Rossii* (The Demographic Yearbook of Russia). Moscow: Goskomstat.

Andreev E., Darskii L., Kharkova T. (1998). *Demograficheskaya istoria Rossii: 1927-1959* (Demographic History of Russia: 1927-1959). Moscow: Informatics.

APPENDIX 1:
DESCRIPTION OF DATA USED FOR LEXIS DATABASE

Deaths

Period	Type of Data	Age Grouping	Comments	RefCode†
1959-2005	Annual number of deaths by sex and age	0,1,2,3,4,5,…,99,100+	Unpublished data; for the purposes of our calculations, we aggregate deaths for age 99+ during 1959-1989 (see "Specific Details" under "Death Count Data").	01, 11, 22

Population

Period	Type of Data	Age Grouping	Comments	RefCode†
1959, 1970	Population estimates as of January 1st	0,1,2,…,100+	Census Data, estimated population as of January 1	03
1971-1978	Population estimates as of January 1st	0, 1, 2, …, 85+		06
1979	Population estimates as of January 1st	0,1,2,…,100+	Census Data, estimated population as of January 1	03
1980-2006	Population estimates as of January 1st	0,1,2,…,100+	For the purposes of our calculations, we aggregate population estimates for ages 80+ (see "Data Quality Issues").	08, 12, 21

BIRTHS

Period	Type of Data	Comments	RefCode†
1959-2005	Annual counts of births by sex		10, 11

† The reference code is used in the raw data files (Input Database) to link data with sources

APPENDIX 2:
DESCRIPTION OF DATA USED IN THE INPUT DATABASE

Research Note by Sergei Zakharov[2] April, 2001

Deaths

Type of data: Annual officially registered number of deaths by age compiled from the death certificates.

Age grouping: single age groups (0,1…99, 100+)

Period covered: 1946 – 1999

Comments:

1. No adjustment for age heaping and for underestimation has been made.
2. Slight adjustment for territory compatibility has been made:

- Crimea region was subtracted by direct method (1946-1956);
- Data for a few regions—namely Kaliningrad, Tuva, and Karel provinces—are not available or incomplete for some years before 1956. Based on fragmentary data available, we decided to include some extra deaths. About 6,500 deaths for 1946 and 1953-1955, and about 900 for 1948 and 1949 were added and distributed by age proportionately.

Population

Type of data: Permanently Resident population on 1st of January since 1946. Different sources for different calendar periods and different age groups have been used.

Age grouping: single age groups (0,1…99, 100+)

Period covered: 1946 – 1999

Comments

1. Official estimates as of the 1st of January are based on the 1959, 1970, 1979, 1989 Census Data.

   For the 1959 Census, only an "Actually present population" ("Nalichnoe naselenie") distributed by age is available, because only this category of population was processed according to the programme of data processing adopted for this Census.

2. Official (the Russian Federation State Committee on Statistics - GOSKOMSTAT) estimates for 1979–1999 are based on 1979 and 1989 Census data, annually registered number of births, age-classified number of deaths, and estimated number of net migrants by age. For migration estimates, a special procedure was applied. This procedure was developed in TCSU in 1978, and then modified several times. Currently, GOSKOMSTAT estimates for migration are based on registration information of the Ministry of Internal Affairs (registration of in- and out-

[2] Centre for Demography and Human Ecology (Moscow)

migrants) and the State Committee on Migration (registration of forced migrants, refugees etc.). The procedure allows for some adjustments for underestimation and non-formalized expert estimates as well.

3. Official (the USSR Central Statistical Board - TCSU) estimates for 1970-1978 are supplemented by E. Andreev's[3] interpolations for people aged 85 and over.

   Before 1979, the TCSU used only an "Actually present population" ("Nalichnoe naselenie") category rather than a "Permanently resident population" ("Postoyannoe naselenie") category for one-year age distribution of people aged 85+. We have included E. Andreev's estimates since they are the only estimates known to us, and are more or less reliable. However, they are necessarily very smoothed and have to be used carefully.

4. Estimates for the 1959-1970 intercensal period were made by E.Andreev. In the 1990s, E. Andreev made estimates for this period using conventional methods. No official distributions of population by one-year age groups exist for 1959-1970.

5. Backward extrapolation for the period of 1946-1959 based on the 1959 Census data was done by E. Andreev in the 1990s (smoothed data). There are no official distributions of population by one-year age groups for this period.

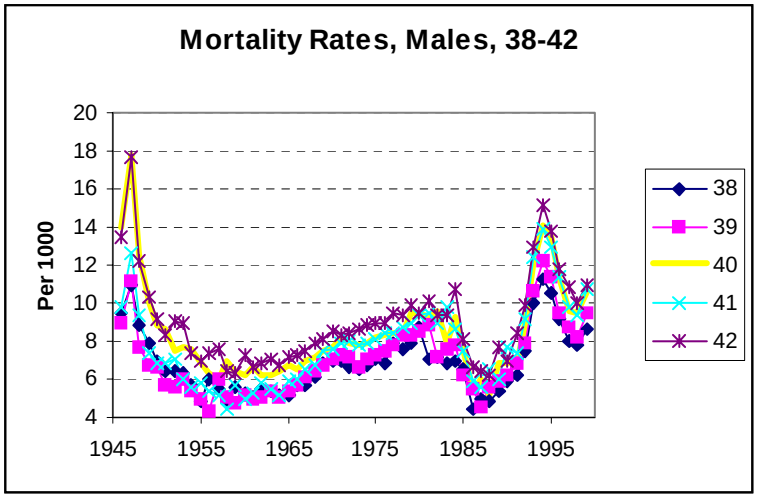
6. For 1959, we have two population distributions: a smoothed one as a base for the 1946-1959 backward extrapolation mentioned above, and a second one estimated as of the 1st of January based on the Official Census Data.

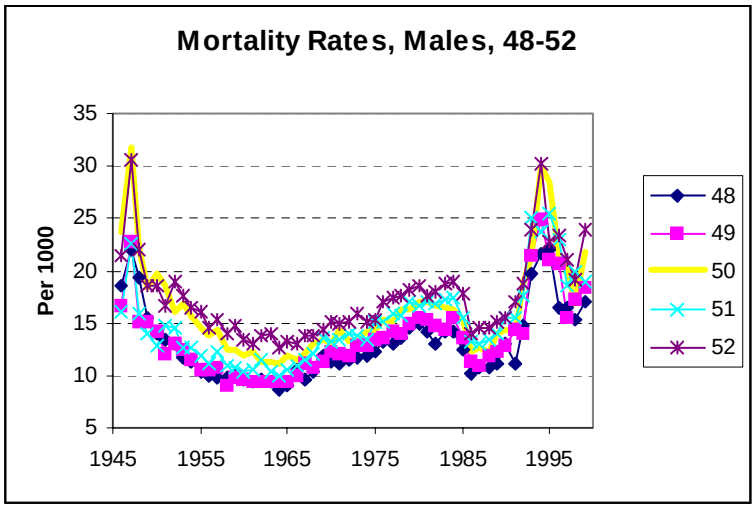
SOME DATA PROBLEMS REVEALED

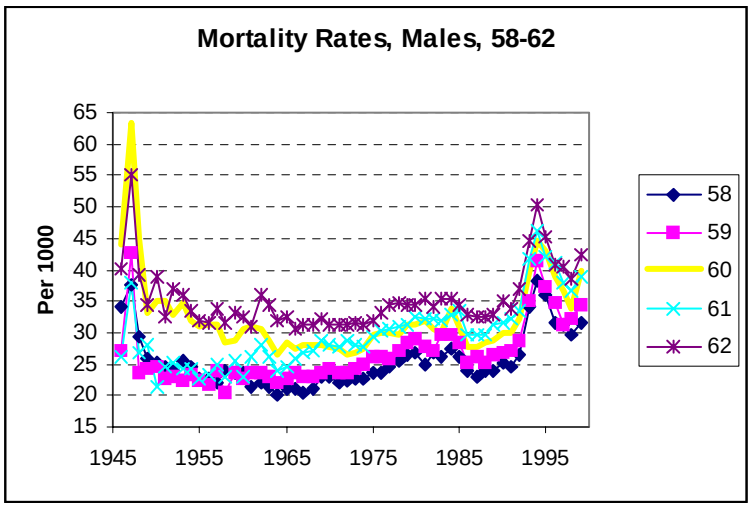
1. Age-heaping effect.

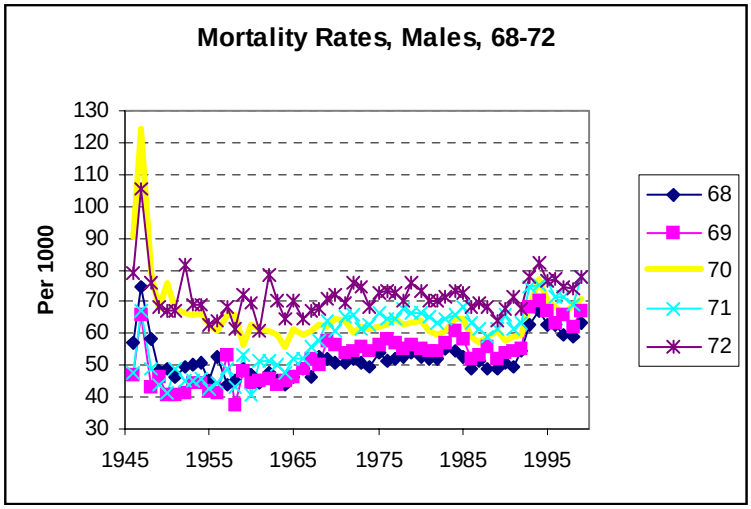
The most serious problem with the Russian mortality data is pronounced age heaping on deaths at ages ending in "0" and "5". For all years prior to 1970, it is clear that age heaping in number of deaths results in abnormally high mortality rates at every age ending in "5" and "0", and, consequently lower rates between these ages. At older ages, the age-heaping effect is more pronounced. It is also evident that by applying a smoothed distribution of population by age to calculate mortality rates (e.g., 1946-1959 period), we just reinforce the age-heaping effect.
Figures below demonstrate the problem for the male population. The same is true for females (not presented here).

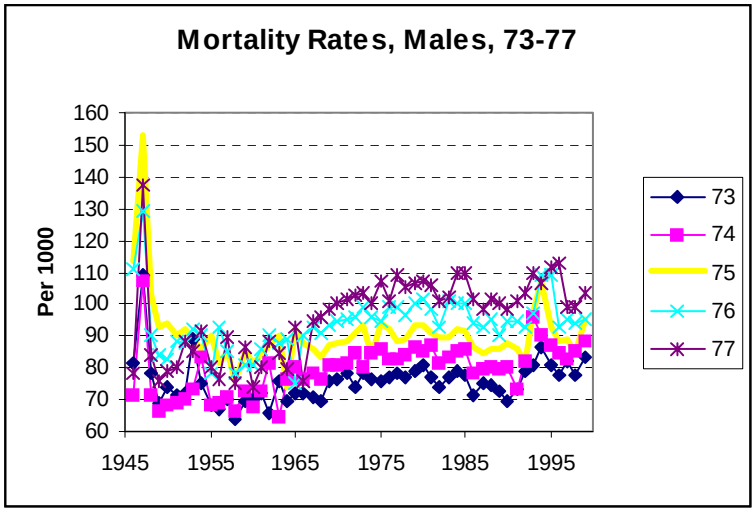
[3] Andreev, Evgenij, Ph.D., Leader of Expert Group on Methodology and Population Estimates in the Russian Federation State Committee on Statistics. He is also a Head of the Laboratory in the Center for Demography and Human Ecology, Russian Academy of Sciences Institute for Economic Forecasting.

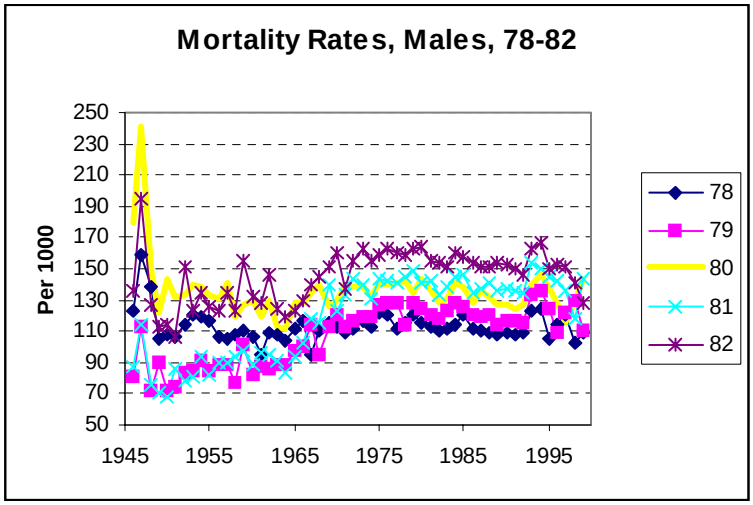

Mortality Rates, Males, 38-42
Per 1000
20
18
16
14
12
10
8
6
4
1945
1955
1965
1975
1985
1995
38
39
40
41
42

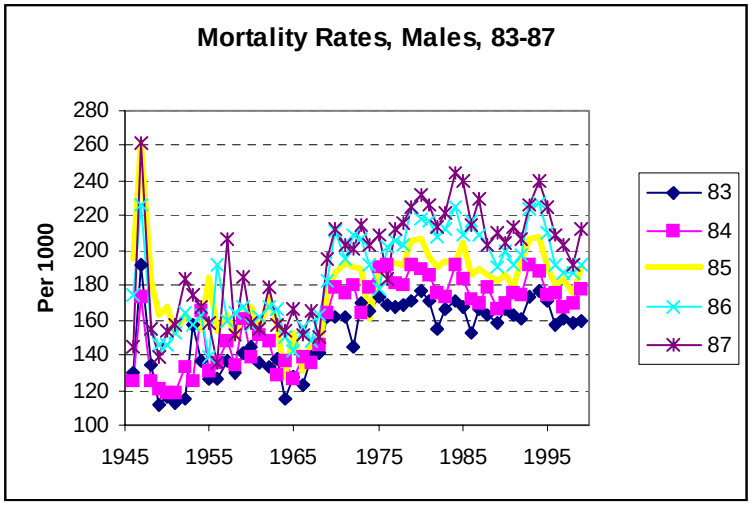

Mortality Rates, Males, 48-52
Per 1000
35
30
25
20
15
10
5
1945
1955
1965
1975
1985
1995
48
49
50
51
52


Mortality Rates, Males, 58-62
Per 1000
65
60
55
50
45
40
35
30
25
20
15
1945
1955
1965
1975
1985
1995
58
59
60
61
62


Mortality Rates, Males, 68-72
Per 1000
130
120
110
100
90
80
70
60
50
40
30
1945
1955
1965
1975
1985
1995
68
69
70
71
72


Mortality Rates, Males, 73-77
Per 1000
160
150
140
130
120
110
100
90
80
70
60
1945
1955
1965
1975
1985
1995
73
74
75
76
77


Mortality Rates, Males, 78-82
Per 1000
250
230
210
190
170
150
130
110
90
70
50
1945
1955
1965
1975
1985
1995
78
79
80
81
82



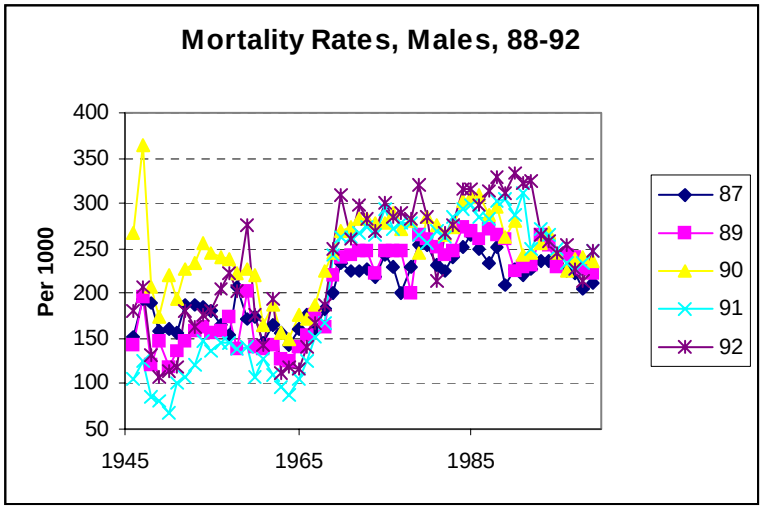


Fig. 1 - 8. Mortality Rates at selected ages, Russia, Males: 1946-1999.

Taking into account the serious age heaping problem, we decided to include in the Human Mortality Database not the whole Russian Data Set, but only series for the period since1970.

2. Unnatural trend in number of deaths at age 99.

In the Russian data for the whole period under study, age of 99 is the last closed age interval, and 100+ is an open-ended age interval. It is clear that the number of deaths in 1958-1987 is too high relative to deaths at nearby ages such as 97, 98, and 100+ (see the ratios presented in fig. 9). Given the situation of age heaping at other ages, we would expect the opposite effect for an age ending in “9” (i.e., artificially lower number of deaths). Thus, we are dealing with a systemic error of another nature, and most likely associated with some rules or special instructions adopted for a registration procedure.

One feasible explanation could be the following:

- The age of people who die as indicated on the death certificate is a result of simple subtraction of the year of birth (e.g., specified in the passport of

the deceased person) from the known year of death. It means that a big portion of deaths (more than 50%) was wrongly classified by age in a cohort-wise manner.

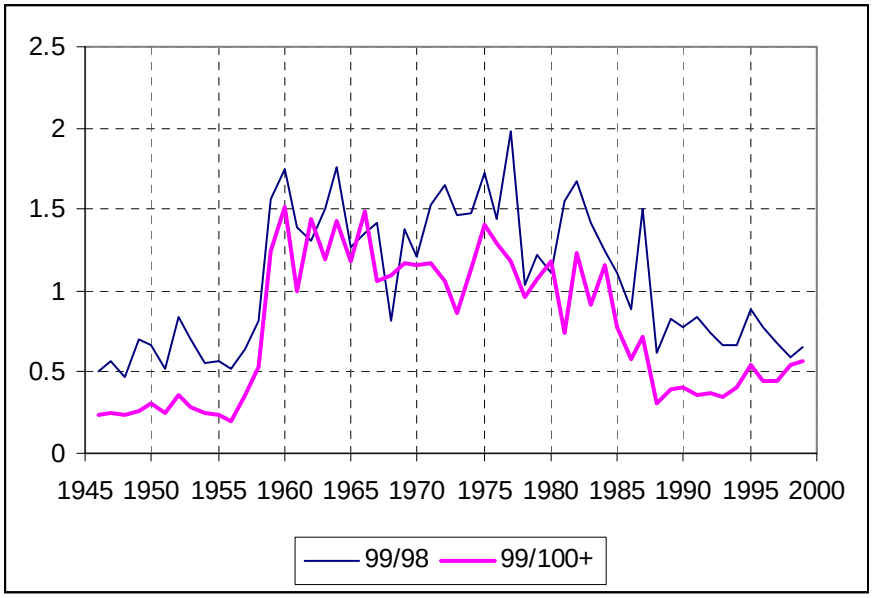


Fig. 9. Ratio of number of deaths at age 99 to number of deaths at age 98, and 100&over, Russia, Males: 1946-1999.

Taking into account the latter problem, we decided to use 99+ as an open-ended age interval for the period 1970-1988 instead of the 100+ age group that is officially used.

CONCLUSION

1. There are serious fallacies in the official estimation of the population of elderly men and women reported by the RF GOSKOMSTAT, especially for persons over 80. As a result, elderly population changes and, correspondingly, mortality rates, become highly irregular and unjustified. The older people are, the greater the error in the official data. These fallacies are of a methodological character and, most probably, have to do with use of the wrong model of approximation for very old people's migration activity.
2. Application of the extinct cohort method gave good results. New estimates of the population over 80 and age mortality rates re-calculated with the new denominator are much better than corresponding official data with respect to their dynamics over time. At least, the unjustified irregularity of population changes and mortality rates has been attenuated. Inconsistency between age-specific mortality rates has also been reduced.
3. Nonetheless, the problem of death age heaping is still one of the main problems of Russian data. We did not intend to solve this problem, and leave it open to researchers who would particularly focus on it. We have tried to mitigate this problem, having limited the data matrix to the period after 1970. Therefore, use of the Russian data on number of deaths and

mortality rates by one-year age groups included into Human Mortality Database (HMD), should be done carefully.