Diversity of cause-of-death coding practices across Russian regions *

Inna Danilova[a,b], Vladimir M. Shkolnikov[a,c], Dmitri A. Jdanov[a,c], France Meslé[d], Jacques Vallin[d]

a *Max Planck Institute for Demographic Research (Germany)*
b *National Research University Higher School of Economics (Russia)*
c *New Economic School (Russia)*
d *Institut national d'études démographiques (France)*

Introduction

Data on mortality by causes of death are an important source that allow for monitoring epidemiological patterns and identifying public health problems. However, even though ICD manuals provide clear and detailed instructions on coding process, the validity and comparability of cause-specific data series is questionable. The Russian system of producing information on causes of death is highly decentralized, which causes potential hazards of discrepancies of coding practices within the country.

Here we present the first study that systematically addresses the problem of comparability of cause-specific mortality statistics across Russian regions. We examined the regional cause-specific data series as they are published in official statistics and tried to assess the presence and the level of regional discordances in propensity of coding certain diagnoses in Russia.

Data

Regional death counts and mid-year population estimates were obtained from the Russian Statistics Service. We use data for the period from 2002 to 2012 for a sub-sample of 52 regions. We have limited our analysis to 52 regions in order to get rid of the random fluctuations caused by small numbers of death events. In 2002-2012 these regions comprised 88.4% of the whole population and 88.5% of all deaths. For the same reason – to eliminate biases generated by small numbers – we grouped some items of Russian Abridged Classification of Causes of Death (RC-1999) into broader diagnostic groups. The final list of selected causes of death includes 70 items.

Method

To estimate the inter-regional variability of mortality from specific cause of death we used *the cause-specific share of the all-cause age-standardized death rate*:

$$S_{r,c,t} = \frac{SDR_{r,c,t}}{SDR_{r,t}} \cdot 100\% \ ,$$

where $SDR_{r,c,t}$ is age-standardized death rate for cause *c* in region *r* in year *t*, and $SDR_{r,t}$ is all-cause age-standardized death rate in region *r* in year *t*.

Next, for each possible combination region/cause we calculated the indicator measuring the deviation from the cross-regional mean (period average) - $V_{r,c}$.

Thus we got a data set of scores where each percentage score $V_{r,c}$ shows how much on average (with respect to time) the share of cause *c* in all-cause SDR of region *r* differs from an average inter-regional share of the same cause. The total size of the data set is equal to the number of regions multiplied by the number of causes of death.

To determine whether there is a certain regularity as to which causes and which regions are more likely to deviate from an average inter-regional level, we applied a least squares regression model with two sets of dummy variables for regions and for causes of death:

$$V_{r,c} = a + b_r I_r + d_c I_c + \varepsilon_{r,c} \ ,$$

where a is a constant term, I_r and I_c are independent regional and cause-specific dummy variables respectively, b_r and d_c are the coefficients on these variables, and $\varepsilon_{r,c}$ is an error term. We used "Kaluga Oblast" as the reference category for variable I_r and "Trachea, bronchus and lung cancers" for variable I_c.

Results

Among 69 causes of death that were assigned dummy variables, 45 causes showed a statistically significant ($p<0.05$) difference from the reference level, and 38 causes showed the difference at $p<0.01$. The highest regression coefficients d_c were found for dummy-variables corresponding to AIDS (+71.4%), senility (+70.9%), mental and behavioral disorders (+63.1%), atherosclerosis (+53.8%), hypertensive diseases (+51.4%). The highest consistency (the lowest regression coefficients d_c) was found for causes representing different groups of cancers (from +0.7% for stomach cancer to +11.8% for the item "cancers of other digestive organs"). Other causes with low deviations from the reference category are: nontraumatic intracranial hemorrhage (+4.3%) and transport accidents (+4.4%).

Panels of figure 1 present examples of distributions of causes of death that are very close to each other by the share in the overall mortality while the respective coefficients d_c returned by the regression model differ among these causes quite substantially.

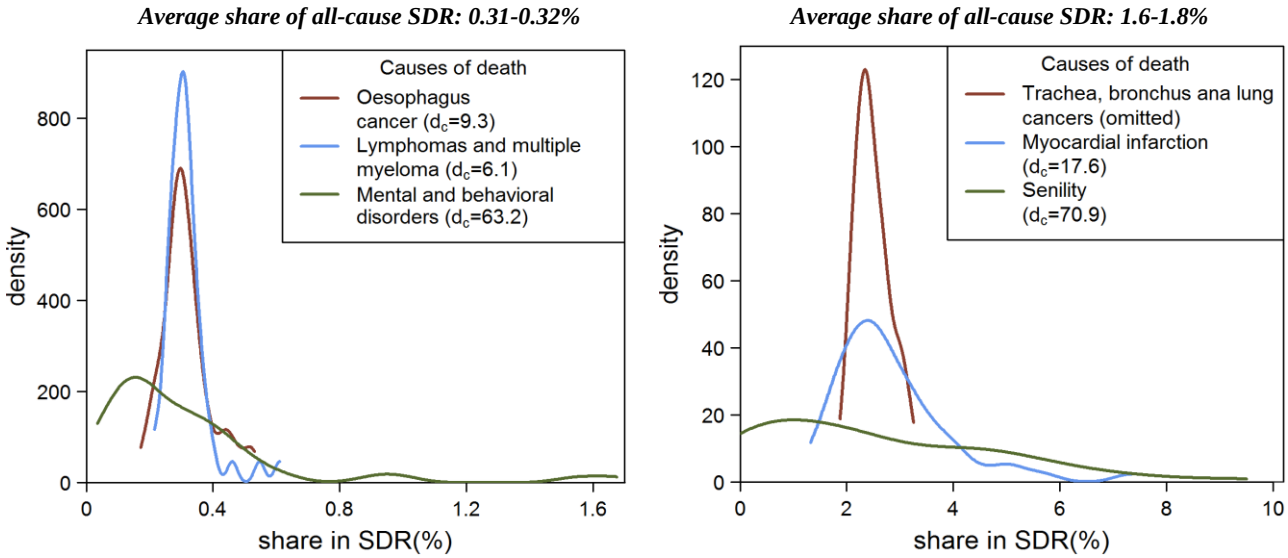

Figure 1. Examples of distributions of cause-specific shares of the all-cause SDR across 52 regions

In 13 regions the average of $V_{r,c}$ values predicted by equationwere statistically different from that in the reference (Kaluga) at $p<0.05$ level and in seven of them it was statistically different at $p<0.01$ level. The top scores were found for Dagestan (+32.6%), the city of Moscow (+29.8%), the city of Saint Petersburg (+19.9%).

To see whether the patterns of deviations are stable in the regions over an observation period, we inspected the regional time series. We have found a number of regional cause-specific series which were unexpectedly distorted during the period 2002-2012. These abrupt and/or unpredictably large changes of mortality levels from particular causes over time possibly indicate a modification of coding practices, due to which some number of deaths that had previously been coded to a certain item started to be coded to the other one (Figure 2). Interestingly, the breaks at regional time series occurred at different time points and the directions of these changes were sometimes even opposite in different regions.

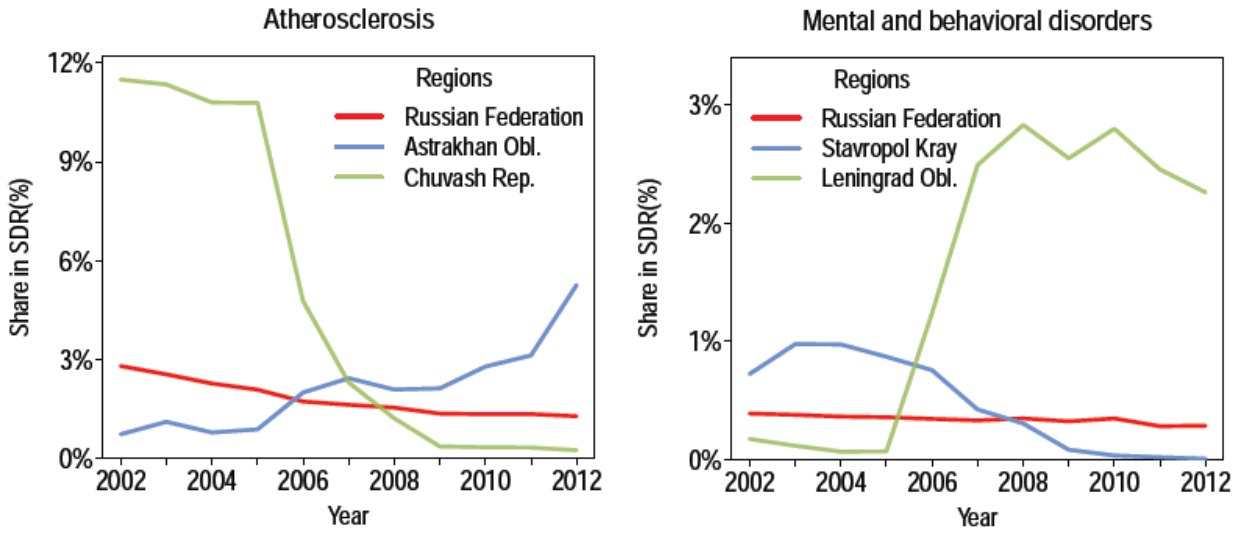

Figure 2. Examples of rapid and contrasting changes in regional cause-specific shares of all-cause SDR (both sexes combined). *The trend for Russia as a whole is provided for comparison.*

There is a meaningful association between the causes of death the highest spatial variation was found for and the causes with the highest number of breaks in regional time series.

Conclusion

The systematical analysis we performed showed that there is a high variance of coding practices in choosing some causes of death as underlying across Russian regions. We found that mortality statistics from some causes is more a reflection of the coding practices rather than about the real epidemiological situation. These problems of comparability can affect the validity and the generalizability of cause-specific mortality statistics. This possibility of biases should be taken into account while performing mortality analysis. There is an urgent need to provide the uniformity and stability of coding practices at the subnational level in Russia to strengthen the accuracy and the quality of mortality statistics.

* This study was conducted in the framework of the project "From disparities in mortality trends to future health challenges (DIMOCHA)" funded by Deutsche Forschungsgemeinschaft (DFG) (Germany) (JA 2302/1-1) and Agence nationale de la recherche (ANR) (France) (ANR-12-FRAL-0003-01). This study was also supported by the AXA Research Fund and the Fund "Dynasty" (Russian Federation).