Ethnic mortality differentials in Lithuania:
Conflicting evidence from longitudinal census-linked and unlinked data

Domantas Jasilionis, Dmitri A. Jdanov, Vladimir M. Shkolnikov

Max Planck Institute for Demographic Research
Konrad Zuse Str. 1, 18057 Rostock, Germany
email: jasilionis@demogr.mpg.de

Introduction

Studies on post-Soviet countries suggest about importance of the widening mortality inequalities. However, the majority of the studies still rely on cross-sectional (census-unlinked) data which are affected by so called numerator-denominator bias (Vallin, 1979). The bias originates from possible differences between self-reported information from the census and death certificate information provided by proxy informants. The most reliable data on group-specific mortality are provided by longitudinal studies linking individual deaths and census records (Valkonen, 1993).

The magnitude of the numerator-denominator bias in group-specific mortality has been estimated in several studies. Comparisons of self-reported and reported (by proxy informants) education, race, and occupation, and an examination of the reliability of information indicated in death records were performed in the USA and Great Britain. The results are contradictory, however. Two studies have shown misreporting in death certificates that was relatively insignificant and thus could not have produced a substantial effect on the group-specific mortality estimates (Rosamond, Tyroler, Chambless et al., 1997; Goldblatt, 1989; Marmot & McDowel, 1986). Other matching studies, however, have found significant disagreement (up to 30-40%) between the self-reported data in the survey and the follow-up death registry data (Shai & Rosenwaike, 1989; Sorlie & Johnson, 1996). Our own previous study on Lithuania also found very notable disagreements between the census-based and death records-based data on education and marital status (Shkolnikov et al., 2007).

This study uses a unique data set from Lithuania containing aggregated mortality data based on the individual linkage of deaths in 2001-2004 to the population census of 2001. The study systematically examines disagreements between the census and the death registry information in reporting the ethnicity of the deceased. In addition, the paper also provides higher quality census-linked data on age- and cause-specific patterns of ethnic mortality differentials in Lithuania.

Data and methods

This study uses a data set provided by Statistics Lithuania. It is based on all records from the 2001 Population and housing census and all death and emigration records for the period between July 1, 2001 and December 31, 2004. Information about emigrations was drawn from the Population Register database.

A functioning system of personal identification numbers (PINs) allowed us to link about 95% of the death records to the corresponding census records. Person-years at risk were estimated by adding up the years of persons living in Lithuania during July, 1 2001 and December 31, 2004. For individuals who died or emigrated, the exposure time was censored at the date of death or emigration. The data include 3.2 and 4.1 million person-years of population exposure and 72.5 and 65.9 thousand deaths for males and females, respectively.

To estimate the impact of various factors on the disagreement between the census and death record based information on ethnicity, we applied Poisson

regression with a misreporting rate as the dependent variable. This analysis was performed separately for each sex and socio-demographic variable. Age was present in all regressions, whereas separate age-adjusted models were estimated to measure the independent effects of causes of death, education, marital status, and the place of residence (urban/rural).

Relative mortality differences for ages 30+ were obtained from the Poisson regressions with mortality as the dependent variable. Separate models were estimated to control for age and all other variables under study (age, education, marital status, place of residence). All results produced by Poisson regressions are expressed in terms of mortality rate ratios and their 95% confidence intervals.

Results

Table 1 gives absolute numbers and agreement rates (in percent) between the data on ethnicity obtained from the death and census records. The lowest agreement rate (84.1%) was found for Other ethnicity. 8% and 5% of the deceased declared as Other ethnicity in the death records reported themselves as being Poles and Russians in the census. The agreement rate for the Russian ethnicity was higher (90%). There was an interchange between the Russian and Other ethnicities (4.4% of Russians reported in the death certificates were classified as Other ethnicity in the census). The highest agreement rates were found for Lithuanian and Polish ethnic groups (94%). The most significant interchange for the Polish ethnicity was observed with the Lithuanian ethnicity (4.4%), whereas for the Lithuanian ethnicity it was with the Russian ethnic group (2.7%). The overall number of excess deaths for the Lithuanian ethnic group gained due to the reporting bias was very significant (almost 7 thou. deaths).

Table 1. *Agreement rate between the data on ethnicity obtained from the death and census records, in percent*

Ethnicity in the census records	**Lithuanian *(death record)***		**Russian *(death record)***		**Polish *(death record)***		**Other *(death record)***	
	Abs. number	%	Abs. number	%	Abs. number	%	Abs. number	%
Lithuanian	**111509**	***94.2%***	236	*2.8%*	401	*4.4%*	79	*2.3%*
Russian	3205	*2.7%*	**7564**	***89.9%***	55	*0.6%*	186	*5.4%*
Polish	2216	*1.9%*	184	*2.2%*	**8568**	***93.9%***	273	*7.9%*
Other	1319	*1.1%*	369	*4.4%*	75	*0.8%*	***2906***	***84.1%***
Unknown	152	*0.1%*	58	*0.7%*	23	*0.3%*	12	*0.3%*

Annex 1 provides data about the relationships between selected socio-demographic variables and misreporting of ethnicity. With the exception of males at ages 80+, who show a lower risk of misreporting, the reporting bias for ethnicity seems to be independent of age effects. Dying from neoplasms (females) and from external causes of death was related to lower risk of misreporting. The highest risk of misreporting was for the deceased dying from ill-defined causes. The effects of education on misreporting ethnicity are relatively small. Secondary education has a small but statistically significant impact on misreporting male and female ethnicity.

Surprisingly, the lowest educational category shows exactly the opposite relationship and is associated with better agreement between the linked and the unlinked information on ethnicity. Finally, there is a notably lower risk of misreporting of ethnicity among rural residents than among urban residents.

The relationships between life expectancy and ethnicity take different directions in the linked compared to the unlinked estimates. According to the unlinked data, the lowest life expectancy, at age 30, is observed among Polish and Lithuanian males and Lithuanian females, whereas the highest longevity is found in the groups of Other ethnicity and Russians (Table 3). Conversely, the census-linked life expectancies reveal an expected advantage of the Lithuanian group experiencing about 2 years higher longevity than that of the Russians and about 3 years higher longevity as compared to Polish males and females.

Table 3. *Life expectancy at age 30 calculated according to the census-linked and the unlinked data.*

	Males		**Females**	
	Census-linked	**Unlinked**	**Census-linked**	**Unlinked**
Lithuanian	**39.00** 38.88-39.11	**38.28** 38.17-38.39	**49.18** 49.08-49.28	**48.57** 48.48-48.66
Russian	**37.18** 36.81-37.55	**40.16** 39.76-40.55	**47.29** 46.97-47.61	**49.81** 49.49-50.13
Polish	**35.85** 35.48-36.23	**38.21** 37.83-38.60	**46.84** 46.52-47.16	**48.87** 48.55-49.18
Other	**38.98** 38.44-39.52	**43.48** 42.92-44.04	**47.66** 47.19-48.13	**50.19** 49.74-50.65
Highest-Lowest	**3.15**	**5.27**	**2.34**	**1.62**

The mortality rate ratios calculated from the two sources also provide exactly opposite evidence about the ethnic differentials (Table 4). The census-based mortality estimates suggest mortality risks 1.1-1.3 times higher for the Russian and Polish ethnicities compared to the Lithuanian ethnic group (the reference category). Conversely, the unlinked mortality estimates suggest an implausible health disadvantage of the Lithuanian ethnic group (Table 4).

Table 4. *Poisson regression mortality rate ratios for ages 30+ calculated from the census linked and unlinked mortality data.*

	Poisson regression mortality rate ratios, age 30+			
	Males		**Females**	
	Census-linked	**Unlinked**	**Census-linked**	**Unlinked**
Ethnicity				
Lithuanian	1	1	1	1
Russian	**1.14***** 1.07-1.14	**0.83***** 0.80-0.85	**1.17***** 1.13-1.21	**0.84***** 0.81-0.87
Polish	**1.26***** 1.23-1.29	0.98 0.95-1.01	**1.21***** 1.18-1.24	**0.92***** 0.89-0.95
Other	0.99 0.97-1.06	**0.64***** 0.61-0.67	**1.15***** 1.09-1.21	**0.84***** 0.80-0.88

All models are adjusted for age; *** $p \leq 0.001$; ** $p \leq 0.01$

The observed mortality inequalities by ethnicity can be related to notable compositional differences by education and urban-rural place of residence. However, our findings suggest that even when ethnicity is measured simultaneously controlling for educational level and place of residence, Russians and Poles have excess mortality with respect to Lithuanians of, respectively, 1.23 and 1.19 for the men and 1.23 and 1.17 for the women (Table 5). This means that the difference in life expectancy between Russian and Other ethnic groups, on one hand, and Lithuanians, on another hand, is narrowed by the higher educational level and more frequent urban residence of the Russian and Other ethnic groups compared with the Lithuanians. Conversely, between Poles and Lithuanians the difference in life expectancy would be lower if the compositional differences (especially by education and urban-rural residence) would be eliminated.

Table 5. *Poisson regression mortality rate ratios for ages 30+ by ethnic group: outcomes from two models*

	Poisson regression mortality rate ratios, age 30+			
	Males		**Females**	
	Model 1 *(adjusted for age only)*	**Model 2** *(adjusted for age, education, marital status, and urban-rural residence)*	**Model 1** *(adjusted for age only)*	**Model 2** *(adjusted for age, education, marital status, and urban-rural residence)*
Ethnicity				
Lithuanian	1	1	1	1
Russian	**1.14***** 1.07-1.14	**1.23***** 1.20-1.27	**1.17***** 1.13-1.21	**1.23***** 1.20-1.27
Polish	**1.26***** 1.23-1.29	**1.19***** 1.16-1.22	**1.21***** 1.18-1.24	**1.17***** 1.14-1.20
Other	0.99 0.97-1.06	**1.10***** 1.06-1.15	**1.15***** 1.09-1.21	**1.20***** 1.15-1.26

All models are adjusted for age; *** $p \leq 0.001$; ** $p \leq 0.01$

Concluding remarks

More reliable census-linked data-based estimates confirm that there is a moderate health advantage of the Lithuanian ethnic group against the remaining ethnic groups (with exception of the Other ethnicity for males) in Lithuania. The results of this study together with the fact that the issue of mortality differentials is central to public health call for producing better data for the assessment of differential mortality in Eastern Europe.

References

Goldblatt, P. (1989). Mortality by social class, 1971-85. *Population Trends,* 56, 6-15.

Marmot, M. G., & McDowall, M. E. (1986). Mortality decline and widening social inequalities. *Lancet,* 2, 274-276.

Rosamond, W.D., Tyroler, H.A., Chambless, L.E., Folsom, A.R., Cooper, L., Conwill, D. (1997). Educational Achievement Recorded on Certificates of Death Compared with Self-Report. *Epidemiology,* 8(2), 202-204.

Shai, D., & Rosenwaike, I. (1989). Errors in reporting education on the death certificate: some findings for older male decedents from New York State and Utah. *American Journal of Epidemiology*, 130(1), 188-192.

Shkolnikov, V.M., Jasilionis, D., Andreev, E.M., Jdanov, D.A., Stankuniene, V., Ambrozaitiene, D. (2007). Linked versus unlinked estimates of mortality and length of life by education and marital status: evidence from the first record linkage study in Lithuania. *Social Science and Medicine*, 64(7), 1392-1406.

Sorlie, P.D. & Johnson, N.J. (1996). Validity of Education Information on the Death Certificate. *Epidemiology*, 7(4), 437-439.

Valkonen, T. (1993). Problems in the measurement and international comparisons of socio-economic differences in mortality. *Social Science and Medicine,* 36(4), 409-418.

Vallin, J. (1979). Socioeconomic determinants of mortality in industrialized countries, *Readings in Population Research Methodology 2*, 957-971.

Annex 1. *Relationships between misreporting of ethnicity and selected socio-demographic factors. Outcomes of Poisson regressions.*

	Males	**Females**
Age group		
30-39	1	1
40-49	1.04 0.89-1.22	1.21 0.90-1.62
50-59	1.08 0.94-1.25	1.30 0.99-1.71
60-69	0.91 0.79-1.05	1.17 0.90-1.52
70-79	0.87 0.76-1.00	1.18 0.91-1.53
80+	**0.68*** 0.59-0.79	0.91 0.70-1.18
Cause of death		
Cardiovascular syst. dis.	1	1
Neoplasms	0.92 0.85-1.00	**0.86*** 0.79-0.93
External causes	**0.89*** 0.80-0.98	1.06 0.92-1.22
Alcohol-related deaths	1.09 0.94-1.28	1.04 0.84-1.30
Infectious & respiratory system diseases	0.93 0.82-1.06	0.88 0.73-1.05
Ill-defined causes	**1.70*** 1.33-2.19	**2.25*** 1.66-3.05
All other causes	0.96 0.83-1.12	1.06 0.94-1.20
Education		
Higher	1	1
Secondary	**1.12*** 1.00-1.26	**1.15*** 1.00-1.32
Lower than secondary	**0.75*** 0.67-0.84	**0.79*** 0.69-0.91
Unknown	0.90 0.77-1.04	**0.79*** 0.68-0.92
Marital status		
Married	1	1
Never Married	0.98 0.88-1.10	**0.78*** 0.69-0.88
Divorced	0.98 0.89-1.08	**1.22*** 1.09-1.37
Widowed	1.08 0.98-1.18	0.97 0.89-1.04
Unknown	1.04 0.26-4.17	2.01 0.95-4.22
Place of residence		
Urban	1	1
Rural	**0.58*** 0.54-0.62	**0.60*** 0.56-0.64

All models are adjusted for age; * $p \leq 0.05$.