Mortality Reconstruction by Sub-populations

László Németh, Domantas Jasilionis and Dmitri Jdanov

1 November 2019

1 Introduction

2 Data

The available mortality data is a longitudinal register-based data set for both sexes at least 30 years of age. The data covers more decades starting from year 1991 until the year 2015 aggregated into five-year-long periods: 1991-95, 1996-2000, ..., 2011-2015. The baseline always refers to 1 January and the follow-up stops on 31 December in a given period. Education is grouped into three categories based on the Internation Stardard Classification of Education: Low - Primary and lower secondary education (ISCED 1-2); Medium - Upper secondary education (ISCED 3-4); High - Tertiary education (ISCED 5-6). The data is grouped into 5-year-long age intervals: 30-34, 35-39, ..., 85-89, 90+. Note that information on education is not available for people more than 69 years old in 1991, more than 74 in 1996, more than 79 in 2001, more than 84 in 2006 and more than 89 in 2001 (see Figure 1). Also note that educational level is not systematically registered for immigrants and refugees.

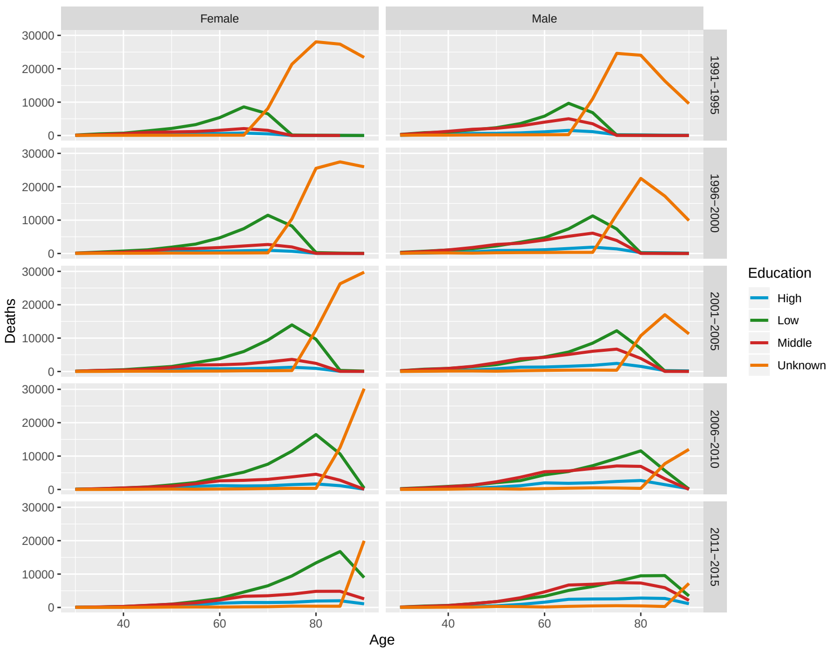

Figure 1: Distribution of deaths by educational groups for females and males in the fives periods.

Figure 2 shows that – even with the assumption that deaths corresponding to unknown educational level might be considered in the low education group – hazards are problematic, especially for earlier periods where the number of unknown cases are much higher than that in the latest period. Irregular pattern of mortality and crossovers between educational groups are present in the data set, therefore mortality at older ages must be reconstructed adequately.

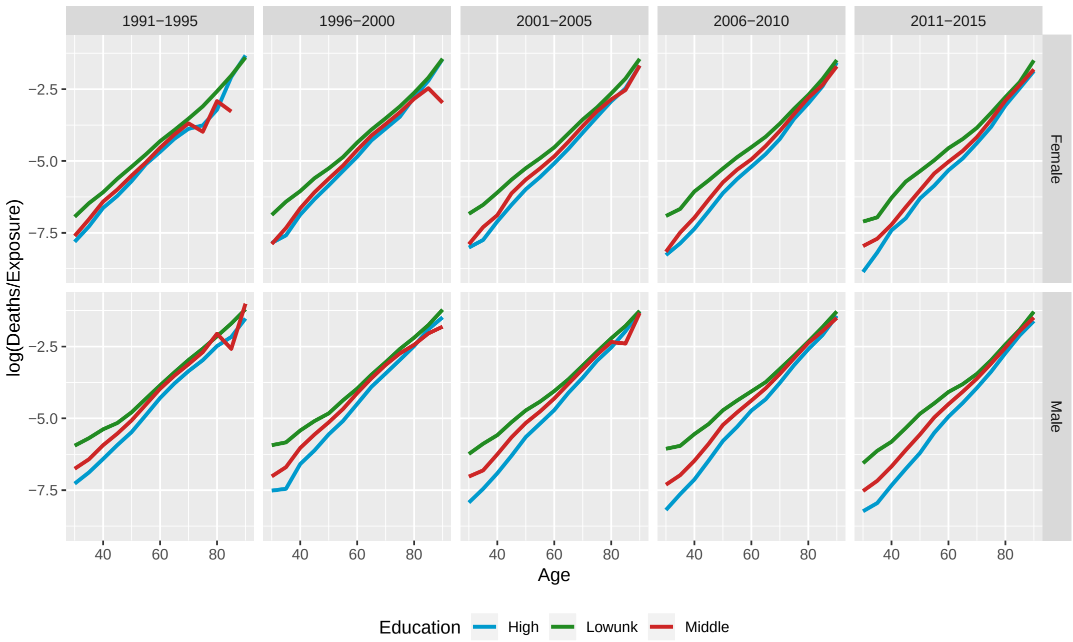

Figure 2: Distribution of deaths by educational groups for females and males in the fives periods.

3 Method

3.1 Parametric Approach

The standard approach to model old-age mortality stems from the observation that death rates at adult ages increase exponentially [2].

We assume that death counts at age x, $D(x)$ are Poisson-distributed [1]: $D(x) \sim Poisson(E(x)\mu(x))$, where E(x) denotes the corresponding exposure at age x and $\mu(x)$ is the hazard risk at this age. The hazard function for the gamma-Gompertz-Makeham model at age x is described by the following expression:

$$\mu(x) = \frac{ae^{bx}}{1 + \frac{a\gamma}{b}(e^{bx} - 1)} + c\,.$$

Parameter a denotes the level of senescent mortality at the starting age of analysis, b is the rate of individual aging, c is an age-independent external risk of death, and γ equals the squared coefficient of variation of the distribution of unobserved heterogeneity [5].

We estimate the parameters of the hazard by maximizing a Poisson log-likelihood in the form

$$\ln L = \sum_x \left[D\left(x\right)\ln\mu\left(x\right) - E\left(x\right)\mu\left(x\right)\right].$$

In case age-specific death counts and exposures are unknown and only death rates are available, we use the non-linear least square approach.

Optimization was carried out by by applying differential evolution [4] using the `DEoptim` R-package [3].

Figures 3 and 4 show the results of a γ-Gompertz and a γ-Gompertz-Makeham model fitting. The estimated models are not able to handle mortality crossovers in all periods but can be easily extended to single ages until a chosen maximum age. Consistency with the HMD total is not ensured.

3.2 Non-parametric Approach

Figure 5 compare the data grouped by educational level to the aggregated data available in the Human Mortality Database.

The non-parametric approach consist of the following steps:

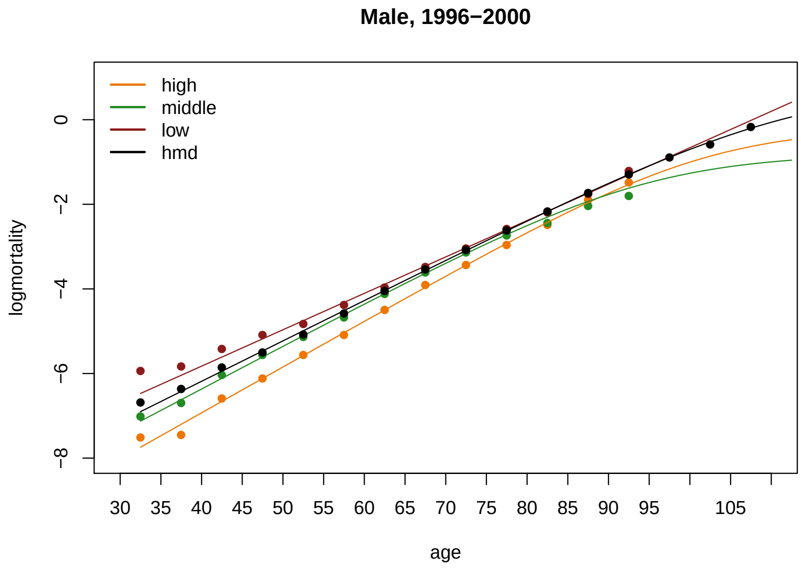

Figure 3: Results of a ΓG model fitting for males in 1996-2000

1. First, the mortality curves are extrapolated to older ages with the assumption that the proportion of a given educational level compared to the total mortality is converging to 1 ("everyone becomes similar at the end") after a specific age. This age is increasing over time as we have more data available and less unknown deaths.

   - Figure 6 shows the relationship between mortality of each group compared to the total mortality in the data set. In the first period the proportion at age 70 is linearly converging to 1 at age 100, in the second the linear change occurs between ages 75 and 100, and so on for later periods. This ensures that no mortality crossover occurs at older ages.

2. Second, weighing the mortality for each educational group with its population share should result in the total mortality for the whole population. The proportion difference is kept constant at later ages. In the first period the difference of the proportions of educational groups in the population between ages 70 and 65 are kept constant. This facilitates the assumption that the shares of educational groups are very rigid at older ages after people finish education at younger ages. Figure 7 shows an example of the shares of each educational group in the first period for females. In Figure 8 we can see how the difference between ages 65 and 70 are kept constant and the shares gradually increase or decrease after age 70. In the later periods the extrapolation of shares starts only at higher ages.

3. Third, total mortality is calculated from the extrapolated mortality curves and population shares.

4 Results

Life expectancy of males are lagging behind but the life expectancy gain achieved between 1991 and 2015 is higher for males than that of females.

Age	Female		Male	
	1991	2015	1991	2015
30	48.8	52.5	43.9	48.6
65	17.7	20.1	14.1	17.4

Table 1: Life expectancy values from the Human Mortality Database

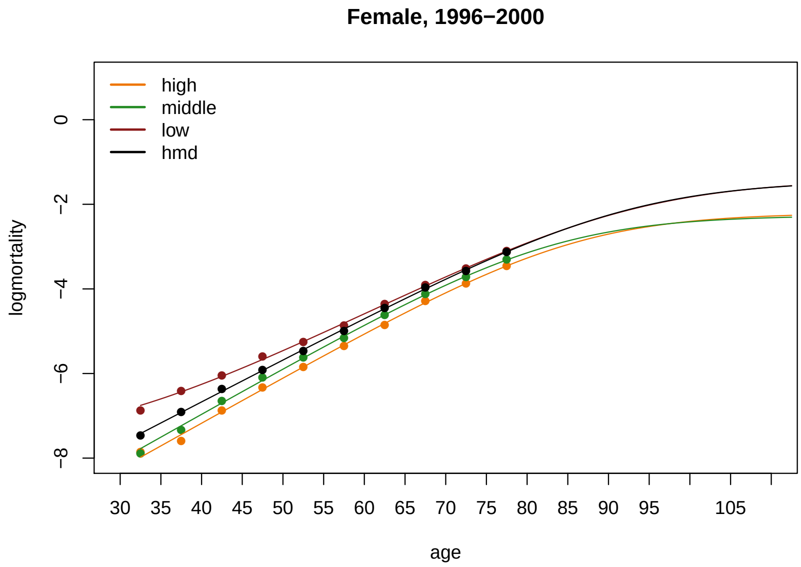

Figure 4: Results of a ΓGM model fitting for females in 1996-2000

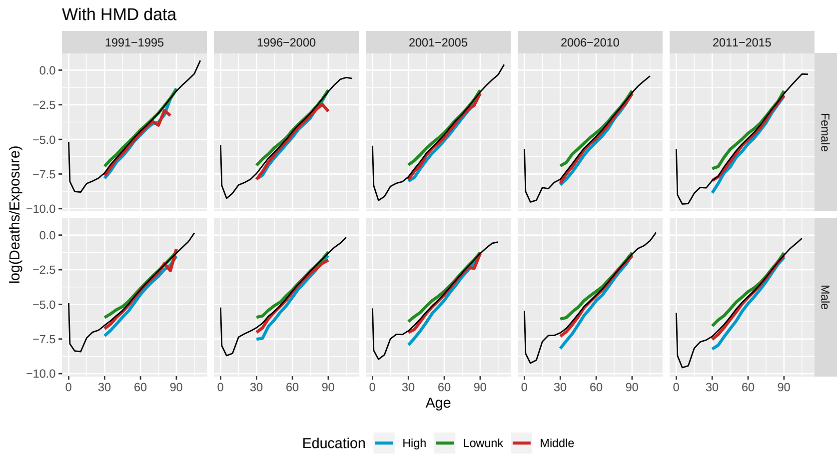

Figure 5: Hazards in comparison with aggregated HMD data

Life expectancy results for the different educational groups are summarized in Table 2. The highest gain was achieved by the highest educated males. On the whole, higher educated males have similar remaining life expectancy as lower or total female life expectancy. Figure 9 depicts the estimated life expectancy values by educational groups between 1991 and 2015.

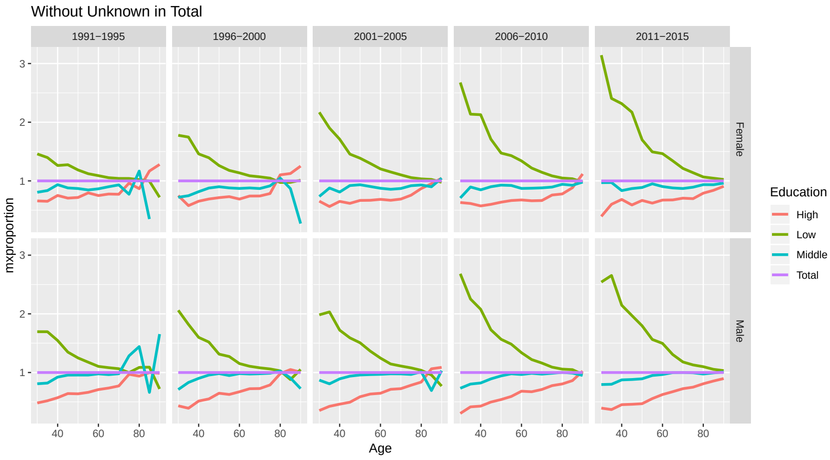

Figure 6: Relationship between mortality of each group compared to the total in the data set

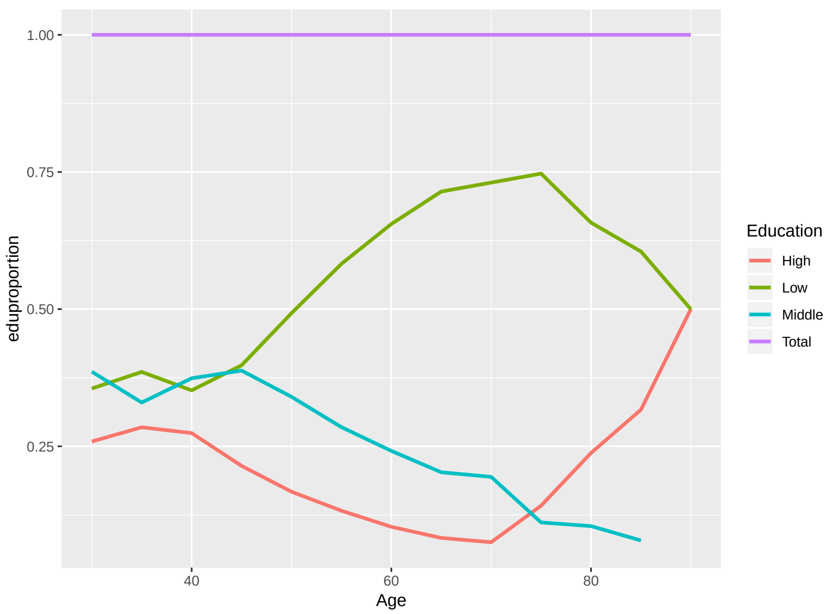

Figure 7: Example: shares of educational groups in the first period for females.

Education	Period	Female		Male	
		30	65	30	65
Low	1991-1995	47.9	17.5	42.1	13.7
	2011-2015	49.9	19.2	45.2	16.4
Middle	1991-1995	49.7	18.2	44.3	14.3
	2011-2015	53.4	20.8	49.0	17.5
High	1991-1995	51.1	19.2	47.1	15.8
	2011-2015	55.2	22.0	52.1	19.3

Table 2: Estimated life expectancy values for each educational group in the first and last periods at ages 30 and 65

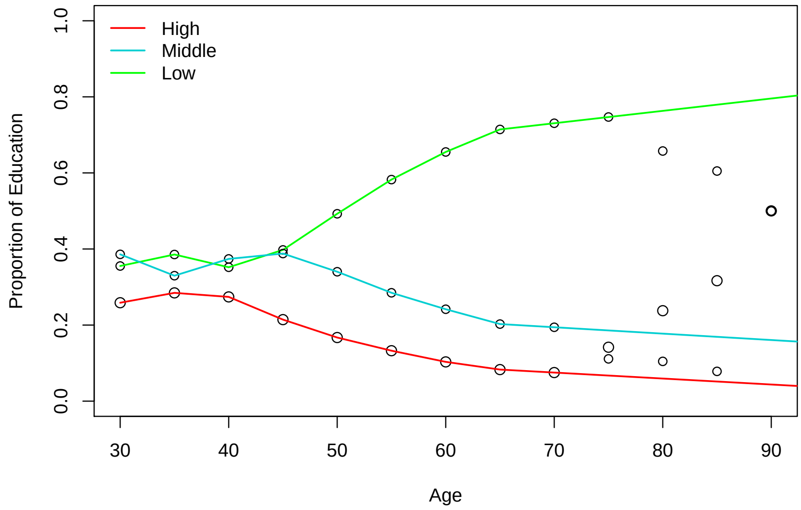

Figure 8: Example: extrapolating shares to higher ages

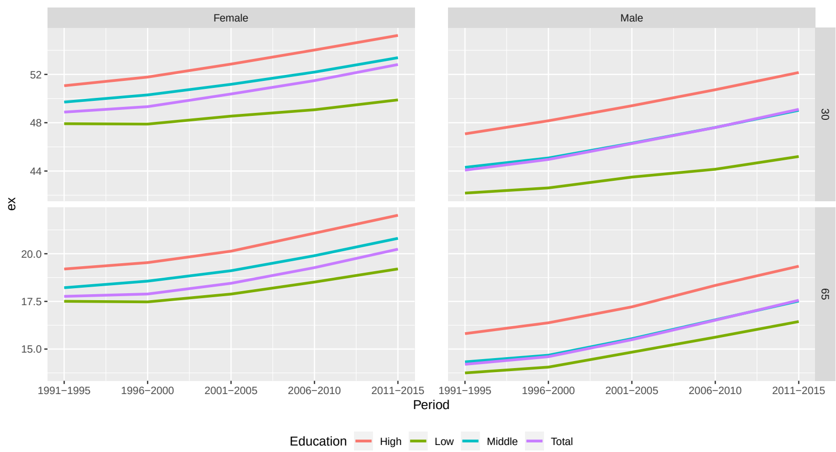

Figure 9: Estimated life expectancy values between 1991 and 2015 by educational groups

References

[1] D.R. Brillinger. The natural variability of vital rates and associated statistics. *Biometrics*, 42(4): 693–734, 1986.

[2] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–583, 1825.

[3] K.M. Mullen, D.L. Gil D. Ardia, D. Windover, and J. Cline. Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 2011.

[4] R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[5] J.W. Vaupel and T.I. Missov. Unobserved population heterogeneity: A review of formal relationships. *Demographic Research*, 31(22):659–686, 2014.