Applying frailty models to analyze educational differences in mortality in Turin

Virginia Zarulli*, Chiara Marinacci†, Graziella Caselli‡

Extended Abstract

Background According to the literature on *frailty models* every individual has a specific level of unobserved frailty, z, that defines the individual hazard in a context of proportional hazard models.

Assuming that frailty follows a Gamma distribution, the population hazard $\bar{\mu}(x)$ at any age x is expressed as a mixture of individual hazards, $\mu(x)$, with the following relationship:

$$\bar{\mu}(x) = \frac{\mu(x)}{1 + \sigma^2 \int_0^x \mu(t)dt} \qquad (1)$$

where σ^2 is the variance of frailty distribution with mean 1 at the initial age and $\mu(x)$ is the baseline hazard experienced by the standard individual with frailty 1.

The individual hazard increases faster than the population one because frailer individuals die faster than the more robust ones. At old ages, this causes the population hazard to level off (Vaupel et al. 1979; Caselli et al. 2000; **?**; Vaupel 2010).

When calculating the effect of observed covariates on survival in regression models, not taking into account the unobserved heterogeneity biases the shape of the hazard even if the omitted source of variation is uncorrelated with the predictors included in the model (Trussell and Rodriguez 1990; Rodriguez 1994). Moreover, ignoring this component may lead to biased estimates of the effect of the covariates (Gail et al. 1984; Lin et al. 1998; Betensky et al. 2002; Aalen 1994; Aalen et al. 2008). Finally, because of the selection process, unobserved heterogeneity may be the cause of converging mortality hazards at old ages by population subgroups (male and female, education groups, income level groups and so on) (Vaupel and Yashin 1985; Beckett 2000; Hoffmann 2005; Zajacova et al. 2009).

In the analysis of socio-economic differential mortality such converging patterns at old ages are often observed. Two main theoretical frameworks address this question. The *age as leveler* perspective attributes this pattern to factors like governmental support to elderly and biological frailty at old ages that levels socio-economic differences off (House et al. 1990, 1994; Markides and Black 1996).

*Department of Statistical Sciences, Sapienza University of Rome (Italy) & Max Planck Institute for Demographic Research, Rostock (Germany); virginia.zarulli@uniroma1.it - zarulli@demogr.mpg.de

†ASL TO3, S.C. a D.U., Public Health School - Turin (Italy)

‡Department of Statistical Sciences, Sapienza University of Rome (Italy)

The *cumulative advantage* perspective, instead, states that at the individual level the effect of the socio-economic position on health remains stable or even increases during the life course, and the convergence at the population level is the result of selection processes (Ross and Wu 1996; Lauderdale 2001; Lynch 2003; Dupre 2007).

Research question We investigate the presence of selection processes in the mortality patterns by socio-economic condition in the Italian city of Turin.

Our aim is to assess whether and how the estimates of the mortality differentials are affected by taking into account unobserved heterogeneity of frailty. Moreover, we want to test the hypothesis that decreasing effect of socio-economic variables with age is an artifact of selection at the population level, while the effect of such covariates at the individual level remains stable.

To our knowledge this is the first study that addresses to these questions using the analytical framework of the frailty models in Italy.

Data and methods We use individual longitudinal histories from the Turin Longitudinal Study (TLS). TLS is a rich study that contains census data linked with municipality population register updated to the end of July 2007 and with health information systems. Deaths that occurred in the population of Turin (nearly 900,000 inhabitants) are linked to the information from the last four decennial censuses (1971-1981-1991-2001) via record linkage techniques, providing rich and reliable demographic and socioeconomic information connected to mortality data.

The result is a follow up of the individuals ever registered in Turin during the period 24/10/1971 - date of the first census - and 31/07/2007: 1655327 individuals (847330 women and 807997 men).

Information is available on education level, sex, macro region of birth, date of birth, date of death or emigration from Turin, and date of immigration.

Missing dates of birth, death, emigration and immigration pertain to only 0.1 % of the total population. Missing data about education level among the adult population from age 30 on is only 0.2%.

The study focuses on socioeconomic differential mortality from late adult ages on. As a proxy for the socioeconomic status we use education level consistently with the literature on the topic (Doblhammer et al. 2009).

We selected the cohorts aged 50 and more at the beginning of the follow-up (area A in fig. 1) and the younger cohorts after they reached aged 50 during the period of analysis (area B in fig. 1). These include 391170 men and 456216 women observed over 36 years.

We use survival analysis considering that our data is both right censored and left truncated.

If for each individual i, y_i is entry time, x_i is the exit time, δ_i is the status (1=dead, 0=right censored), u_i is the covariate profile, $\mu(\cdot)$ denotes the hazard and $S(\cdot)$ the survival function, the likelihood function is:

$$L(\beta,\theta) = \prod_{i=1}^{n} \frac{(\mu(x_i,\theta)e^{u_i\beta})^{\delta_i}(S(x_i,\theta))^{e^{u_i\beta}}}{(S(y_i,\theta))^{e^{u_i\beta}}} \quad (2)$$

where θ is the vector of parameters of the hazard function. When accounting for unobserved heterogeneity, assuming a gamma distributed frailty the hazard takes the form of (1).

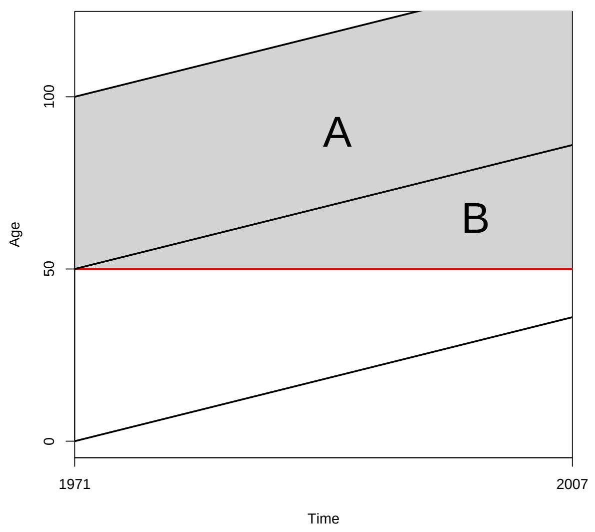

Figure 1: Stylized representation of the data

Some descriptive results The investigation of selection processes requires a cohort perspective. Our data allow us to follow a fully longitudinal approach that we will combine with survival analysis techniques for individual level data.

A fist glance at the empirical cohort death rates by education in figure 2 reveals that the usual pattern of converging hazards at old ages is more pronounced among women than among men. Men, surprisingly, show less marked educational differences, while the literature on the topic usually finds wider mortality gradients among men because women's education level is often less reflected in their socio-economic status and more likely to be influenced by the partner's one.

Figure 2: Log death rates for the 10 years birth cohort aged 50-60 at the beginning of the follow-up (census 1971)

We plan to investigate this phenomenon more deeply, taking into account that Turin is one of the most important industrial cities in Italy that has experienced vast migration flows in the past decades. Most of the immigrants came from the poorer southern regions of the country. Migrants are supposed to be healthier and, in the

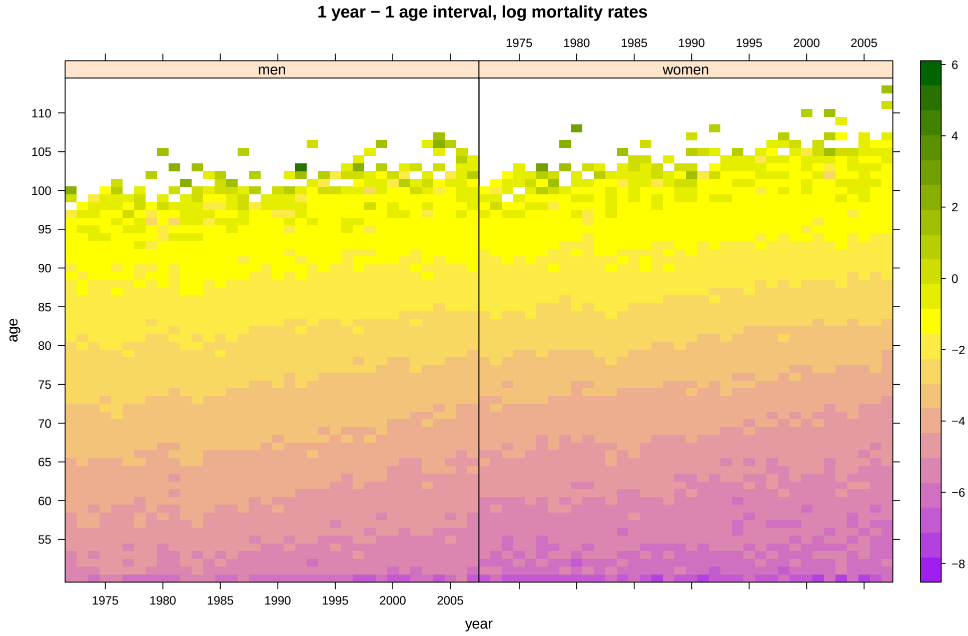

Figure 3: Log death rates by 1 age interval and 1 year

case of Turin, the big majority of them were manual workers with low education seeking a job in the car factories of the city. Therefore the massive migratory flow could have caused a reduction of the educational gradient. Selection among women, instead, might have been less pronounced because they were likely to be passive actors in the migratory decision. Many of them often followed the husband only after some time.

Therefore we plan to analyze the role of the unobserved heterogeneity component in the educational gradient and how and if this component will change if we control for a substantial source of variability like the migratory status and the macro-region of birth.

Geographical gradients are very important in Italy as well. We expect to find differences according to the region of birth. Men from the southern regions are known to perform better in survival and this could be another explanation for the reduced educational gradient found in this first analysis.

The Gompertz pattern is found to fit adequately human mortality data at late adult and old ages. This helps the choice of the functional form for the hazard. We plan to use mainly parametric models in survival analysis and to investigate primarily the Gompertz form and the Gamma-Gompertz one (the correspondent of Gompertz for heterogeneous populations). However, we will explore also other models like, for example, the Gompertz-Makeham and its version that accounts for heterogeneity.

Figure 3 shows the log death rates by 1 age interval and 1 calendar year. The different birth cohorts are represented along the diagonals. Period mortality improvements are clearly visible. Year by year, the same level of mortality shifts up to higher and higher ages. Significant acceleration of the improvements can be observed especially after 1985. The 80's were the years of the cardiovascular revolution in Italy, that started first for the women and some year later for the men. In our model, that describes the age trajectory of mortality of different cohorts, we will have to consider that the cohorts live through different periods and that periods undergo mortality improvement processes.

Acknowledgments

We would like to thank Jutta Gampe for her suggestions and technical advice.

References

Aalen, O. 1994. "Effects of frailty in survival analysis." *Statistical Methods in Medical Research* 3(3):227.

Aalen, O., Ø. Borgan and H. Gjessing. 2008. *Survival and Event History Analysis*, chap. 6: Unobserved heterogeneity: The odd effects of frailty, pp. 231–270. Springer.

Beckett, M. 2000. "Converging Health Inequalities in Later Life–an Artifact of Mortality Selection?" *Journal of Health and Social Behavior* 41(1):106–119.

Betensky, R., D. Louis and J. Gregory Cairncross. 2002. "Influence of unrecognized molecular heterogeneity on randomized clinical trials." *Journal of clinical oncology* 20(10):2495.

Caselli, G., J. Vaupel and A. Yashin. 2000. "Longevity, heterogeneity and selection." *Atti della XL Riunione Scientifica della Società Italiana di Statistica* pp. 49–72.

Doblhammer, G., R. Hoffmann, E. Muth, C. Westphal and A. Kruse. 2009. "A systematic literature review of studies analyzing the effect of sex, age, education, marital status, obesity, and smoking on health transitions." *Demographic Research* 20(5):37–64.

Dupre, M. 2007. "Educational differences in age-related patterns of disease: Reconsidering the cumulative disadvantage and age-as-leveler hypotheses." *Journal of Health and Social Behavior* 48(1):1.

Gail, M., S. Wieand and S. Piantadosi. 1984. "Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates." *Biometrika* 71(3):431.

Hoffmann, R. 2005. "Do socioeconomic mortality differences decrease with rising age?" *Demographic Research* 13(2):35–62.

House, J., R. Kessler and A. Herzog. 1990. "Age, socioeconomic status, and health." *The Milbank Quarterly* 68(3):383–411.

House, J., J. Lepkowski, A. Kinney, R. Mero, R. Kessler and A. Herzog. 1994. "The social stratification of aging and health." *Journal of Health and Social Behavior* 35(3):213–234.

Lauderdale, D. 2001. "Education and survival: Birth cohort, period, and age effects." *Demography* 38(4):551–561.

Lin, D., B. Psaty and R. Kronmal. 1998. "Assessing the sensitivity of regression results to unmeasured confounders in observational studies." *Biometrics* pp. 948–963.

Lynch, S. 2003. "Cohort and life-course patterns in the relationship between education and health: A hierarchical approach." *Demography* 40(2):309–331.

Markides, K. and S. Black. 1996. "Race, ethnicity, and aging: The impact of inequality." In *Handbook of Aging and the Social Sciences*, Academic Press.

Rodriguez, G. 1994. "Statistical issues in the analysis of reproductive histories using hazard models." *Annals of the New York academy of sciences* 709(1):266–279.

Ross, C. and C. Wu. 1996. "Education, age, and the cumulative advantage in health." *Journal of Health and Social Behavior* 37(1):104–120.

Trussell, J. and G. Rodriguez. 1990. "Heterogeneity in demographic research." In *Convergent issues in genetics and demography*, Oxford University Press, USA.

Vaupel, J. 2010. "Biodemography of human ageing." *Nature* 464(7288):536–542.

Vaupel, J., K. Manton and E. Stallard. 1979. "The impact of heterogeneity in individual frailty on the dynamics of mortality." *Demography* 16(3):439–454.

Vaupel, J. and A. Yashin. 1985. "Heterogeneity's ruses: some surprising effects of selection on population dynamics." *The American Statistician* 39(3):176–185.

Zajacova, A., N. Goldman and G. Rodriguez. 2009. "Unobserved heterogeneity can confound the effect of education on mortality." *Mathematical Population Studies* 16(2):153–173.