BMJ open

Mortality by education level at late-adult ages in Turin: a survival analysis using frailty models with period and cohort approaches

Virginia Zarulli,[1,2] Chiara Marinacci,[3] Giuseppe Costa,[4] Graziella Caselli[5]

To cite: Zarulli V, Marinacci C, Costa G, *et al*. Mortality by education level at late-adult ages in Turin: a survival analysis using frailty models with period and cohort approaches. *BMJ Open* 2013;**3**:e002841. doi:10.1136/bmjopen-2013-002841

▸ Prepublication history and additional material for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2013-002841).

Received 6 March 2013
Revised 28 May 2013
Accepted 29 May 2013

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 3.0 Licence; see http://bmjopen.bmj.com

For numbered affiliations see end of article.

Correspondence to
Dr Virginia Zarulli,
vzarulli@health.sdu.dk

ABSTRACT

Objectives: Neglecting the presence of unobserved heterogeneity in survival analysis models has been showed to potentially lead to underestimating the effect of the covariates included in the analysis. This study aimed to investigate the role of unobserved heterogeneity of frailty on the estimation of mortality differentials from age 50 on by education level.

Design: Longitudinal mortality follow-up of the census-based Turin population linked with the city registry office.

Setting: Italian North-Western city of Turin, observation window 1971–2007.

Population: 391 170 men and 456 216 women followed from age 50.

Primary outcome measures: Mortality rate ratios obtained from survival analysis regression. Models were estimated with and without the component of unobserved heterogeneity of frailty and controlling for mortality improvement over time from both cohort and period perspectives.

Results: In the majority of cases, the models without frailty estimated a smaller educational gradient than the models with frailty.

Conclusions: The results draw the attention of the potential underestimation of the mortality inequalities by socioeconomic levels in survival analysis models when not controlling for unobserved heterogeneity of frailty.

ARTICLE SUMMARY

Article focus

- Neglecting the presence of unobserved heterogeneity in survival analysis models has been shown to potentially lead to underestimating the effect of the covariates included in the analysis.
- Although frailty models have been widely developed to account for unobserved heterogeneity, in differential mortality analyses this source of variation is seldom controlled for. This study has applied these models to a longitudinal mortality analysis by education level.

Key messages

- Mortality differentials by education (or by any other variable used as a proxy of socioeconomic status) could be larger than those estimated with standard survival analysis approaches that do not control for unobserved heterogeneity.

Strengths and limitations of this study

- The strength of this study lies in the population-based longitudinal data. The long observational time (36 years) for more than 847 000 individuals gives a solid base for statistical power and detection of trends.
- The limitation consists in the lack of individual information on lifestyle factors and health events, which could certainly help to better model the concept of unobserved individual frailty by uncovering a part of it.

INTRODUCTION

An extensive body of literature shows significant differential mortality by socioeconomic condition.[1–3] The elderly show decreasing relative social inequalities in general mortality with increasing age.[4–8] The age-as-leveller hypothesis attributes this to factors that contribute to the levelling-off of differences at old ages: governmental support to the elderly,[9–11] disengagement from systems of social stratification[12] and general vulnerability.[13 14] However, this phenomenon could also be an artefact of selection due to the unobserved characteristics of the individuals: selective effects of earlier higher mortality, experienced by the disadvantaged group, would leave more robust individuals at old ages, causing the convergence with the risk of the lower mortality group that is subject to weaker selection.[15–18] Neglecting these hidden differences in survival chances (called unobserved frailty) has been shown to lead to biased estimates of the mortality hazard and of the effect of the covariates on the survival probability.[19–25]

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

In longitudinal analyses on differential mortality, it is important to control for hidden frailty because not controlling for it, in models of survival analysis, could lead to biased estimates of the effect of social position on mortality risk. The statistical literature shows that the bias is towards zero.[24–26] This would lead to an underestimation of the relative differences in mortality risks by socioeconomic group. Frailty models have been developed to control for unobserved frailty and to evaluate its impact on the observed mortality dynamics.[27] For more detailed explanations of the frailty models and how they relate to differential mortality analyses, please see online supplementary appendix A.

This study investigated the presence of selection processes in the mortality patterns of the Turin population (North-West Italy) from age 50 on. Adopting a longitudinal perspective, this study aimed to investigate if the estimates of the mortality differentials are affected by the introduction of the unobserved heterogeneity component into the models.

DATA AND METHODS

We used high quality census-linked data from the Turin Longitudinal Study (TLS), which includes 1971, 1981, 1991 and 2001 census data for the Turin population. TLS records the individual census sociodemographic information and, through record linkage with the local population registry and other local-health information systems, collects information on vital status, cause of death and other health indicators.[28 29]

For this study, the individuals registered in Turin during at least one of the four censuses were selected. Data on their migration and vital status were followed up until the end of July 2007. The result is an observation window of 36 years (from 24 October 1971, official date of the census, to the end of July 2007, end of the linkage) during which the individuals were followed up until death, emigration from the city or end of the observation period. The follow-up started at age 50. The study population contains 391 170 men and 456 216 women.

Study information includes an individual's date of birth, date of exit from the study, cause of exit (death or emigration), sex, macroregion of birth and education level.

Consistent with the literature[30–33] education level was used as an indicator of social position.

The study also controlled for the individual macro region of birth, as Turin is characterised by a history of immigration from other regions of the country.[34]

To facilitate comparison over a long follow-up and different cohorts, we created three broad educational groups: high (high school diploma or higher), medium (junior high school) and low (primary school or lower).

We estimated parametric survival models stratified by gender and as a function of macroregion of birth and education level, with and without a parameter for the unobserved heterogeneity component. The choice of using parametric models, rather than semi parametric or non parametric ones, is justified by the wide demographic literature showing that human adult mortality can be accurately described by a Gompertz function[35] or by some Gompertz-like variants, like Makeham. To identify the best functional form for the baseline, we compared the models with Akaike Information Criterion (AIC).[36]

The data are both right censored (due to emigration or end of follow-up) and left truncated (due to the different age at entry in the study of individuals).

The study includes many cohorts, each passing through 36 years of observation at different ages. However, from 1971 to 2007, a significant mortality improvement occurred and younger cohorts experienced lower age-specific mortality than older cohorts.

Time is a complex variable including three dimensions: age, period and cohort. Controlling adequately for the effect of time would require simultaneous assessment of the three components, but such models have been not identifiable for a long time because of the linear dependence between the three dimensions.[37–39] Recently, it has been shown that, through the introduction of the generalised linear mixed models framework, new estimation methods and model specifications can be used to tackle the identification problem.[40] However, this goes beyond the scope of our study.

We adopted two approaches for the control of time, corresponding to an age-cohort approach and an age-period approach, being aware that they represent two different dimensions of time.

First, we regarded the improvement as a cohort phenomenon, including a covariate for the cohort to which the individuals belong. In this setting, controlling for unobserved heterogeneity was implemented with univariate frailty models, which estimate the baseline parameters, the coefficients of the covariates and the variance of frailty (assumed to follow a γ distribution with mean 1 and variance σ^2 to be estimated).

We then considered the improvement as a period phenomenon and split the time into several calendar period covariates, as well as the survival spell of the individuals, according to which period they were passing through. This implied organising the data into clusters, where each cluster represents one individual's survival spells. In this setting, to control for unobserved heterogeneity shared frailty models are needed, where the spells in each cluster pertain to the same individual and share the same hidden frailty. For computational reasons, the estimation of these highly complex models required the use of random subsampling.[41–43] We repeated the estimation 250 times on a 1% sample of the dataset, randomly drawn without replacement and stratified by the major variables in analysis. The aim was to approximate the parameters' estimates based on the empirical distribution of the repeated estimates.

In the model without frailty, it was possible to include a finer calendar period division, 12-period variables of

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

Figure 1 Death rates on logarithmic scale for the birth cohort aged 50–59 at the beginning of the follow-up (1971) by three education levels: high, medium and low.

3 years each (1971–1973, 1974–1976…), while in the model with frailty, for computational reasons, the number of variables was reduced to two broader periods: 1971–1990 and 1991–2007.

Computations were realised with the software R.[44] Formal details are in online supplementary appendix A.

RESULTS

Figure 1 shows that the log-death rates by education level and gender, for the cohort aged 50–59 in 1971, converge at old ages. Other cohorts showed very similar patterns.

A preliminary analysis found that the reduction of the gradient over age is statistically significant and more pronounced among women (results are reported in online supplementary appendix B, table B1).

Frailty modelling

Table 1 shows AIC of the survival models, fitted to the all population mortality, with Gompertz and Makeham baselines. It also shows the results of the fit when unobserved frailty was controlled for. The comparison reveals that the Gompertz baseline was a better fit for male data, while the Makeham baseline was better for female data. In both cases, the models controlling for unobserved heterogeneity performed a better fit (table 1).

We then estimated the mortality differentials, using a cohort and a period approach, to control for mortality improvement over time. We included in the analysis the variables for education level (high, medium and low) and region of birth (North-West, North-East, Center, South and Abroad).

Tables 2 and 3 report the results of the models estimated with and without the unobserved heterogeneity component: the parameters of the baseline hazard (a and b of the Gompertz function for men and a, b and c of the Makeham function for women), the variance of frailty in the population and the rate ratios of the mortality differentials by education level and region of birth. Figure 2 compares the results for the educational gradient obtained by the models with and without frailty.

Educational gradient

In the model with the age-cohort improvement approach, the introduction of the frailty term made the male differences widen significantly, consistent with the statistical literature. The rate ratios with respect to high education

Table 1 Model selection for the baseline hazard and comparison of the model with best baseline hazard and unobserved heterogeneity of frailty component

	Model with different baseline hazards		Model with best baseline hazard and frailty	
	Gompertz	Makeham	Gamma-Gompertz	Gamma-Makeham
	ae^{bx}	$ae^{bx}+c$	$\frac{ae^{bx}}{1+\sigma^2(a/b)(e^{bx}-1)}$	$\frac{ae^{bx}+c}{1+\sigma^2(a/b)(e^{bx}-1)+cx}$
AIC women	1 327 474	1 326 878	–	1 326 695
AIC men	1 303 693	1 303 695	1 303 655	–

Comparison is based on AIC.

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

Table 2 Results of the regression models with cohort covariates

	Men				Women			
	Model without frailty		Model with frailty		Model without frailty		Model with frailty	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
a	5.241×10^{-5}	5.237×10^{-5} to 5.245×10^{-5}	4.495×10^{-5}	4.488×10^{-5} to 4.501×10^{-5}	3.767×10^{-6}	3.755×10^{-6} to 3.779×10^{-6}	1.605×10^{-6}	1.588×10^{-6} to 1.623×10^{-6}
b	0.081	0.080 to 0.082	0.083	0.082 to 0.084	0.106	0.105 to 0.107	0.117	0.115 to 0.119
c	–	–	–	–	0.001	0.001 to 0.001	0.001	0.001 to 0.001
σ^2	–	–	0.035	0.027 to 0.045	–	–	0.096	0.082 to 0.111
Cohort	0.016	0.015 to 0.016	0.016	0.015 to 0.016	0.016	0.015 to 0.016	0.017	0.016 to 0.017
Education level								
High	1	–	1	–	1	–	1	–
Medium	1.166	1.147 to 1.186	1.221	1.200 to 1.243	1.141	1.116 to 1.166	1.111	1.086 to 1.137
Low	1.239	1.221 to 1.257	1.302	1.283 to 1.322	1.246	1.222 to 1.270	1.213	1.188 to 1.238
Region of birth								
North-West	1	–	1	–	1	–	1	–
North-East	1.053	1.036 to 1.070	1.060	1.042 to 1.077	0.989	0.973 to 1.004	0.974	0.958 to 0.991
Centre	1.011	0.984 to 1.038	0.996	0.969 to 1.024	0.939	0.913 to 0.966	0.968	0.939 to 0.998
South	1.000	0.988 to 1.012	0.950	0.938 to 0.962	0.932	0.919 to 0.945	0.987	0.973 to 1.002
Abroad	1.031	1.006 to 1.057	0.998	0.974 to 1.024	1.071	1.047 to 1.096	0.993	0.968 to 1.018
logLk	−651 219		−651 082		−663 238		−663 098	
AIC	1 302 456		1 302 184		1 326 496		1 326 218	

Baseline parameters (Gompertz for men and Makeham for women) and rate ratios of the differentials by education and region of birth.

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

Table 3 Results of the regression models with period covariates

	Men				Women			
	Model without frailty		Model with frailty*		Model without frailty		Model with frailty*	
	Estimate	95% CI	Mean	0.025–0.0975	Estimate	95% CI	Mean	0.025–0.0975
a	4.159×10^{-5}	3.196×10^{-5} to 5.410×10^{-5}	0.004	0.000–0.010	8.031×10^{-6}	6.028×10^{-6} to 1.070×10^{-5}	0.008	0.000–0.016
b	0.096	0.095 to 0.096	0.069	0.061–0.163	0.121	0.120 to 0.122	0.084	0.073–0.106
c	–	–	–	–	0.001	0.001 to 0.002	2.852×10^{-6}	8.610×10^{-7}–2.997×10^{-5}
σ^2	–	–	0.269	0.026–0.367	–	–	0.292	0.174–0.367
Calendar period								
1971–1973	1	–	1	–	1	–	1	–
1974–1976	0.999	0.972 to 1.027			0.978	0.950 to 1.007		
1977–1979	0.947	0.921 to 0.973			0.919	0.893 to 0.946		
1980–1982	0.928	0.903 to 0.953			0.896	0.871 to 0.922		
1983–1985	0.943	0.918 to 0.969			0.967	0.941 to 0.994		
1986–1988	0.870	0.847 to 0.894			0.848	0.824 to 0.872		
1989–1991	0.820	0.798 to 0.843	0.728	0.613–0.985	0.796	0.774 to 0.818	0.888	0.671–1.035
1992–1994	0.796	0.774 to 0.817			0.757	0.736 to 0.778		
1995–1997	0.741	0.721 to 0.762			0.704	0.684 to 0.724		
1998–2000	0.701	0.682 to 0.721			0.682	0.663 to 0.701		
2001–2003	0.670	0.652 to 0.689			0.657	0.639 to 0.676		
2004–2007	0.631	0.615 to 0.648			0.625	0.608 to 0.642		
High	1	–	1	–	1	–	1	–
Medium	1.204	1.184 to 1.225	1.277	1.054–1.349	1.107	1.083 to 1.131	1.256	1.053–1.347
Low	1.301	1.282 to 1.320	1.268	1.074–1.591	1.209	1.186 to 1.232	1.475	1.103–1.641
North-West	1	–	1	–	1	–	1	–
North-East	1.040	1.024 to 1.057	1.075	0.855–1.220	0.963	0.948 to 0.978	1.122	0.888–1.217
Centre	0.943	0.917 to 0.969	1.081	0.854–1.212	0.964	0.938 to 0.992	1.102	0.864–1.218
South	0.900	0.889 to 0.911	1.037	0.854–1.216	0.962	0.949 to 0.975	1.130	0.904–1.220
Abroad	0.965	0.941 to 0.989	1.082	0.864–1.218	0.985	0.962 to 1.009	1.082	0.847–1.215
logLk	–650 997		Na		–663 081		Na	
AIC	1 302 034		Na		1 326 204		Na	

Baseline parameters (Gompertz for men and Makeham for women) and rate ratios of the mortality differentials by education and region of birth.
*The model with frailty does not report conventional point estimates and CI, but the mean value and the 0.025 to 0.975 quantiles of the empirical distribution of the parameters obtained from the repeated estimates via random subsampling.

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

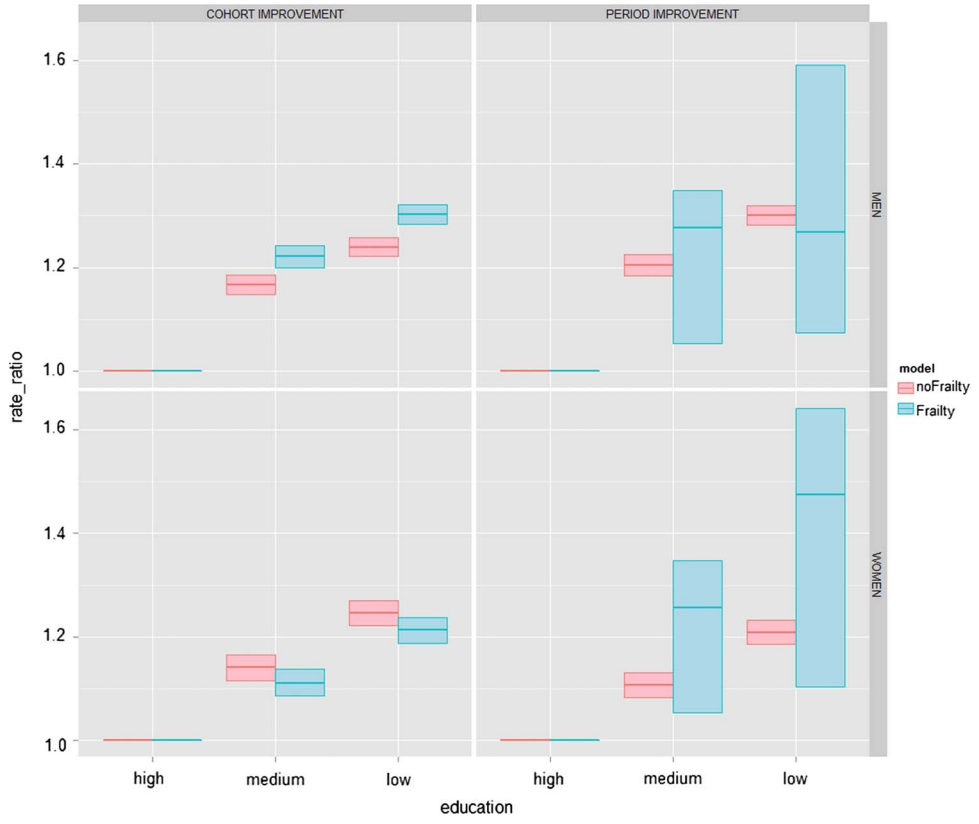

Figure 2 Mortality rate ratios by education level in the models with cohort and period improvement, without and with frailty.

changed from 1.16 (95% CI 1.15 to 1.19) to 1.22 (1.20 to 1.24) for medium education and from 1.24 (1.22 to 1.26) to 1.30 (1.28 to 1.32) for low education (left upper panel of figure 2). Among women, on the contrary, there was a slight reduction, but the confidence regions of the estimates in the two cases overlap: for medium education, the rate ratio went from 1.14 (1.12 to 1.17) to 1.11 (1.08 to 1.14), and for low education from 1.25 (1.22 to 1.27) to 1.22 (1.19 to 1.24; lower left panel of figure 2). AIC indicates that the models with frailty fit the data significantly better than the models without.

In the model adopting the age-period improvement approach, the AIC comparison of the models with and without frailty was not possible, because the utilisation of random subsampling for the estimation of the frailty model[41–43] did not allow for obtaining a likelihood value comparable with the values of the models without frailty. Moreover, it is necessary to consider that we are comparing conventional point estimates and CIs with values obtained via bootstrapping methods, whose confidence regions are usually wider than conventional CI. Nevertheless, a comparison is still possible.

The introduction of frailty affected the mortality gradient by education. Although the uncertainty around the estimates does not allow for assessing a precise effect, the rate ratios of medium and low education in respect to high education in the models with frailty lie in a higher confidence region than in the models without: among women with a medium education level, it lies between 1.05 and 1.34 compared with 1.08 and 1.13 of the model without frailty and for the low-education group, between 1.1 and 1.6, compared with 1.18 and 1.23. The same pattern can be observed among men.

The male difference between the medium-education and low-education group, on the contrary, was not as clear as that among women.

Other results and the impact of the macroregion of birth on mortality

As expected, the variance of frailty in the cohort models was smaller than in the period models, since periods are more heterogeneous than cohorts.

Women were more heterogeneous than men: 0.09 (0.08 to 0.11) vs 0.04 (0.03 to 0.05) in the age-cohort models and 0.29 (0.17 to 0.37) vs 0.27 (0.00 to 0.36) in the age-period models.

This is consistent with the more pronounced convergence of the hazards by education at old age found among women compared with men. According to the framework of the frailty models, converging hazards are the result of the effect of selection on the population hazards, due to how much variance of unobserved frailty is present in the population at the initial age of observation. The bigger the variance, the stronger the convergence is. For more information about frailty models, the process of selection and how they relate to narrowing mortality differentials at old ages, see online supplementary appendix A.

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

In the age-cohort models the introduction of unobserved frailty affected the coefficient for the macroregion of birth significantly. Among men, holding education equal, those born in the South show a significant survival advantage over the natives of the North-West, while in the model without frailty there was no such advantage. Among women, the model without frailty showed a significant survival advantage for those born in the South but when frailty was controlled for, this became not significant.

The pattern also resembles the regional mortality macrodynamics that have characterised Italy for most of the 20th century (although the two patterns refer to different phenomena, the first one refers to mortality by region of birth), when male mortality in the South was lower than in the North.[45–48] Cohort-based analyses have highlighted that, in recent cohorts (those born after WWII), there is a reversing trend.[48 49]

The models with an age-period perspective did not identify any significant geographical differences. This could be due to the utilisation of random subsampling of a 1% sample. Although 250 repetitions are considered by the literature to be a sufficient number for very complex models,[50–52] it is possible that the number was inadequate to identify a clear pattern from the small sample. For more detailed results, see tables 1 and 2.

DISCUSSION

Interest in the role of unobserved heterogeneity in a life-course approach to socioeconomic mortality differences has recently increased. Most of the studies focus on health outcomes,[53–58] while fewer studies also analyse mortality.[59–61] Their findings are not consistent and fuel a still controversial debate.

In this study, we investigated the role of unobserved individual heterogeneity on the estimation of mortality differentials at adult-old ages by education level in a longitudinal perspective. This study investigated if the estimates of the mortality differentials are affected by the introduction of the unobserved heterogeneity component into the models.

We fitted survival analysis models with and without controlling for the unobserved heterogeneity and we found that when this component was included, the models gave a significantly better fit.

We also found that in the majority of cases, the educational gradient estimated by the models with frailty was higher than the one estimated by the models without frailty. When big uncertainty around the estimates did not allow for assessing a precise value, the confidence regions in the models with frailty spanned over higher values than those in the models without frailty. It must be pointed out that, in the age-period approach, the peculiar statistical procedure used to estimate the frailty models did not allow for obtaining a likelihood value comparable with the one of the model without frailty. Thus, the statistical comparison of the models via AIC was not possible, making this evidence weaker. Nevertheless, the results seem to point to a direction that is consistent with the statistical literature about unobserved heterogeneity.[19–26]

Among men, such a pattern was found in both the age-cohort and age-period approaches. Among women, on the contrary, this pattern was less clear: in the age-cohort model, controlling for hidden frailty resulted in a slight reduction of the mortality gradient. Social determinants also act on mortality through risk factors that are known to affect more men than women. Moreover, owing to a lag in the smoking and fertility transitions, highly educated women in Turin are more exposed to risk factors like cigarette smoking and a smaller number of children. Therefore, controlling for hidden frailty in the case of women might reduce the educational gradient.

In the models with age-cohort perspective, controlling for the hidden frailty also affected the estimates of the differentials by macroregion of birth, showing a survival advantage of the men born in the South, but not of the women, for whom an advantage was instead detected by the model that did not control for frailty.

The healthy migrant effect[62–67] could cause this pattern. Among the cohorts involved in the migration, women were likely to be more passive actors than men in the migratory decision,[68–70] which might have been responsible for their being selected less than men. Frailty is a general concept embedding all the hidden factors that affect the individual survival chances: innate and acquired frailty, exposure to risk factors, lifestyle factors and so on. Therefore, when controlling for frailty, the survival advantage of women was reduced, as they might have been less health selected than men by the migration. On the contrary, the advantage of men was uncovered. However, another recent study on the impact of migration on all-cause mortality in Turin did not find particularly strong gender differences in the so-called healthy migrant effect,[66] and this point deserves to be investigated further in future.

The study spanned over a long observation window of 36 years. Therefore, it was important to control for the general mortality improvement that took place during this time. We did so by adopting both an age-period and an age-cohort approach.

The age-period models, as expected, estimated higher heterogeneity than the age-cohort models. Periods aggregate different generations and are expected to be more heterogeneous than the cohorts themselves. In both the period and cohort models, the variance of frailty was higher among women than among men, indicating that men are more homogeneous than women. This could be attributed to a stronger selection process due to mortality that is usually observed to be higher among men than among women.

On the other hand, it is also possible that the industrialisation process and the internal migration experienced by Italy after WWII[34] played a role. The vast majority of

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

less educated individuals in Turin came from the South, seeking a job in the car factories of the city. As less educated men were mainly employed in heavier and riskier jobs and were exposed to higher mortality, it is possible that during their life they were selected at a faster pace than other educational groups and women. This might have reduced the differences in susceptibility to death among men, contributing to determining a lower level of heterogeneity than among women.

CONCLUSION

This study found that neglecting selection effects due to unobserved heterogeneity in longitudinal analyses could lead to an underestimation of mortality differentials by social class. In the majority of cases, the models that controlled for unobserved heterogeneity estimated higher educational differences in mortality than the models that did not control for it.

Moreover, when compared with the AIC, the models that controlled for unobserved heterogeneity gave a statistically significantly better fit than the models that did not control for it. Although the best AIC just shows that the more complex model approximates the data better, it does not represent an unequivocal proof of the selection hypothesis; however, the results point to the possibility that the data could be better described by this hypothesis.

Author affiliations
[1]Max Planck Odense Center on the Biodemography of Aging and Institute of Public Health, University of Southern Denmark, Odense, Denmark
[2]Max Planck Institute for Demographic Research, Laboratory of Survival and Longevity, Rostock, Germany
[3]Department Epidemiology, Local Health Unit TO3, Piedmont Region & Italian Ministry of Health, Rome, Italy
[4]Department of Clinical and Biological Science, University of Turin, Turin, Italy
[5]Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy

Contributors VZ participated in the conception and design of the study, analysis and interpretation of the data and results, drafting of the article and in revising it; GrC participated in the interpretation of the results, drafting of the article and in revising it critically; CM and GiC participated in revising the article for important intellectual content. All authors have read and approved the final manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None..

Ethics approval Anonymous data from the city population registry and censuses were used.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Mackenbach JP, Kunst AE, Cavelaars AEJM, *et al*. Socioeconomic inequalities in morbidity and mortality in western Europe. *Lancet* 1997;349:1655–9.
2. Mackenbach JP, Kunst AE, Groenhof F, *et al*. Socioeconomic inequalities in mortality among women and among men: an international study. *Am J Public Health* 1999;89:1800–6.
3. Mackenbach JP, Stirbu I, Roskam AJR, *et al*. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med* 2008;358:2468–81.
4. Antonovsky A. Social class, life expectancy and overall mortality. *Milbank Memorial Fund Q* 1967;45:31–73.
5. Huisman M, Kunst AE, Mackenbach JP. Socioeconomic inequalities in morbidity among the elderly; a European overview. *Soc Sci Med* 2003;57:861–73.
6. Dalstra J, Kunst A, Mackenbach J, *et al*. A comparative appraisal of the relationship of education, income and housing tenure with less than good health among the elderly in Europe. *Soc Sci Med* 2006;62:2046–60.
7. Martelin T. Mortality by indicators of socioeconomic status among the Finnish elderly. *Soc Sci Med* 1994;38:1257–78.
8. Huisman M, Kunst AE, Andersen O, *et al*. Socioeconomic inequalities in mortality among elderly people in 11 European populations. *J Epidemiol Community Health* 2004;58:468–75.
9. House JS, Lepkowski JM, Kinney AM, *et al*. The social stratification of aging and health. *J Health Soc Behav* 1994:213–34.
10. Decker S, Rapaport C. Medicare and disparities in women's health. National Bureau of Economic Research, 2002.
11. Dor A, Sudano J, Baker DW. The effect of private insurance on the health of older, working age adults: evidence from the Health and Retirement Study. *Health Serv Res* 2006;41(3p1):759–87.
12. Marmot MG, Shipley MJ. Do socioeconomic differences in mortality persist after retirement? 25 year follow up of civil servants from the first Whitehall study. *BMJ* 1996;313:1177–80.
13. Elo IT, Preston SH. Educational differentials in mortality: United States, 1979–1985. *Soc Sci Med* 1996;42:47–57.
14. Liang J, Bennett J, Krause N, *et al*. Old age mortality in Japan. *J Gerontol Ser B* 2002;57:S294.
15. Caselli G, Vaupel JW, Yashin AI. Explanation of the decline in mortality among the oldest-old: a demographic point of view. Human Longevity, Individual Life Duration, and the Growth of the Oldest-Old Population, 2006:395–413.
16. Manton KG, Stallard E. Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: explaining the black/white mortality crossover by a model of mortality selection. *Hum Biol* 1981;53:47.
17. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979;16:439–54.
18. Vaupel JW, Yashin AI. Heterogeneity's ruses: some surprising effects of selection on population dynamics. *Am Statistician* 1985:176–85.
19. Aalen OO. Effects of frailty in survival analysis. *Stat Methods Med Res* 1994;3:227.
20. Aalen OO. Heterogeneity in survival analysis. *Stat Med* 1988;7:1121–37.
21. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 1984;71:431.
22. Trussell J, Rodriguez G. *Heterogeneity in demographic research. Convergent issues in genetics and demography*. Oxford University Press, 1990:111–32.
23. Chamberlain G. Heterogeneity, omitted variable bias, and duration dependence. In: Heckman JJ, Singer B.eds *Longitudinal Analysis of Labor Market Data* Cambridge: Cambridge University Press, 1985:3–38.
24. Schumacher M, Olschewski M, Schmoor C. The impact of heterogeneity on the comparison of survival times. *Stat Med* 1987;6:773–84.
25. Schmoor C, Schumacher M. Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Stat Med* 1997;16:225–37.
26. Bretagnolle J, Huber-Carol C. Effects of omitting covariates in Cox's model for survival data. *Scand J Stat* 1988;15:125–38.
27. Wienke A. *Frailty models in survival analysis*. Chapman & Hall/CRC, 2010.
28. Marinacci C, Spadea T, Biggeri A, *et al*. The role of individual and contextual socioeconomic circumstances on mortality: analysis of time variations in a city of north west Italy. *J Epidemiol Community Health* 2004;58:199–207.
29. Costa G, Cardano M, Demaria M, *et al*. *Storie di salute in una grande città. Città di Torino, Ufficio di statistica*. Osservatorio socioeconomico torinese, 1998.
30. Doblhammer G, Hoffmann R, Muth E, *et al*. A systematic literature review of studies analyzing the effect of sex, age, education, marital status, obesity, and smoking on health transitions. *Demogr Res* 2009;20:37–64.
31. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health* 1997;18:341–78.
32. Mirowsky J, Ross CE. *Education, social status, and health*. Aldine de Gruyter, 2003.

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

33. Galobardes B, Shaw M, Lawlor DA, *et al*. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health* 2006;60:7–12.
34. Bonifazi C, Heins F. Long-term trends of internal migration in Italy. *Int J Popul Geography* 2000;6:111–31.
35. Gompertz B. 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. *Philos Trans R Soc London* 1825;115:513–83.
36. Akaike H. A new look at the statistical model identification. *Automatic Control, IEEE Trans* 1974;19:716–23.
37. Holford TR. Analysing the temporal effects of age, period and cohort. *Stat Methods Med Res* 1992;1:317–37.
38. Osmond C, Gardner M. Age, period, and cohort models. Non-overlapping cohorts don't resolve the identification problem. *Am J Epidemiol* 1989;129:31.
39. Glenn ND. Cohort analysts' futile quest: statistical attempts to separate age, period and cohort effects. *Am Sociol Rev* 1976;41:900–4.
40. Yang Y, Land KC. *Age-period-cohort analysis: new models, methods, and empirical applications*. Chapman & Hall, 2013.
41. Hartigan JA. Using subsample values as typical values. *J Am Stat Assoc* 1969;64:1303–17.
42. Politis DN, Romano JP. Large sample confidence regions based on subsamples under minimal assumptions. *Ann Stat* 1994;22:2031–50.
43. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat* 1979;7:1–26.
44. R Development Core Team. *R: a language and environment for statistical computing*. Vienna, Austria, 2011.
45. Barbi E, Caselli G. Selection effects on regional differences in survivorship in Italy. *Genus* 2003:37–61.
46. Caselli G, Egidi V. Le differenze territoriali di mortalità in Italia. Tavole di mortalità provinciali (1971–72). 1980.
47. Caselli G, Egidi V. L'analyse des données multidimensionnelles dan l'étude des relations entre mortalité et variable socio-économiques d' envirnment et de comportement individuel [Multivariate methods in the analysis of the relations between mortality and socio-economic, environmental and behavioural variables]. *Genus* 1981;37:57–91.
48. Caselli G, Reale A. Does cohort analysis contribute to the study of the geography of mortality? *Genus* 1999;55:27–59.
49. Biggeri A, Accetta G, Egidi V. Evoluzione del profilo di mortalita 30–74 anni per le coorti di nascita dal 1889 al 1968 nelle regioni italiane [Mortality Time Trends 30–74 years by Birth Cohorts 1889–1968 in the Italian Regions]. *Epidemiol Prev* 2011;35:50–67.
50. Efron B, Tibshirani R. *An introduction to the bootstrap*: Chapman & Hall/CRC, 1993.
51. Manly BFJ. *Randomization, bootstrap and Monte Carlo methods in biology*: Chapman & Hall/CRC, 1997.
52. Pattengale ND, Alipour M, Bininda-Emonds ORP, *et al*. How many bootstrap replicates are necessary? *J Comput Biol* 2010;17:337–54.
53. Beckett M. Converging health inequalities in later life-an artifact of mortality selection? *J Health Soc Behav* 2000:106–19.
54. Ferraro KF, Farmer MM. Double jeopardy, aging as leveler, or persistent health inequality? A longitudinal analysis of white and black Americans. *J Gerontol Ser B* 1996;51:S319.
55. Herd P. Do functional health inequalities decrease in old age? *Res Aging* 2006;28:375–92.
56. Kim J, Durden E. Socioeconomic status and age trajectories of health. *Soc Sci Med* 2007;65:2489–502.
57. Lynch SM. Cohort and life-course patterns in the relationship between education and health: a hierarchical approach. *Demography* 2003;40:309–31.
58. McMunn A, Nazroo J, Breeze E. Inequalities in health at older ages: a longitudinal investigation of the onset of illness and survival effects in England. *Age Ageing* 2009;38:181.
59. Dupre ME. Educational differences in age-related patterns of disease: reconsidering the cumulative disadvantage and age-as-leveler hypotheses. *J Health Soc Behav* 2007;48:1–15.
60. Hoffmann R. Do socioeconomic mortality differences decrease with rising age? *Demogr Res* 2005;13:35–62.
61. Hoffmann R. Socioeconomic inequalities in old-age mortality: a comparison of Denmark and the USA. *Soc Sci Med* 2011;72:1986–92.
62. Anson J. The migrant mortality advantage: a 70 month follow-up of the Brussels population. *Eur J Popul* 2004;20:191–218.
63. Feinleib M, Lambert PM, Zeiner-Henriksen T, *et al*. The British-Norwegian migrant study—analysis of parameters of mortality differentials associated with angina. *Biometrics* 1982:55–71.
64. Kington R, Carlisle D, McCaffrey D, *et al*. Racial differences in functional status among elderly US migrants from the south. *Soc Sci Med* 1998;47:831–40.
65. Norman P, Boyle P, Rees P. Selective migration, health and deprivation: a longitudinal analysis. *Soc Sci Med* 2005;60:2755–71.
66. Rasulo D, Spadea T, Onorati R, *et al*. The impact of migration in all-cause mortality: the Turin Longitudinal Study, 1971–2005. *Soc Sci Med* 2012;74:897–906.
67. Singh GK, Siahpush M. All-cause and cause-specific mortality of immigrants and native born in the United States. *Am J Public Health* 2001;91:392.
68. Bielby WT, Bielby DD. I will follow him: family ties, gender-role beliefs, and reluctance to relocate for a better job. *Am J Sociol* 1992;97:1241–67.
69. Cooke TJ. Gender role beliefs and family migration. *Popul Space Place* 2008;14:163–75.
70. Mincer J. Family migration decisions. *J Political Economy* 1978;86:749–75.

BMJ Open: first published as 10.1136/bmjopen-2013-002841 on 3 July 2013. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

Appendix

A. Frailty models and Survival Analysis

Frailty models

Hidden differences in survival chances make individuals differ in their susceptibility to death. This complex set of characteristics, called unobserved frailty, does not distinguish between acquired weakness, life style factors, environmental risks and innate biological frailty, but it indicates a general susceptibility to death (1).

In cohort analyses, as the population ages, frailer individuals die faster and gradually select the survivors in terms of robustness, because the population undergoes a compositional change. This causes the population hazard to decelerate at very old ages because, at every age, the death rate is computed based on a population at risk whose composition is gradually converging towards the low frailty individuals, who have also lower mortality. The greater the variance of unobserved heterogeneity of frailty at the initial age of observation, the stronger the selection process and, therefore, the faster the deceleration of the hazard observed at the population level as age goes by.

Neglecting the presence of unobserved frailty and its selection processes can lead in survival analysis models to possible biases in the estimates of the regression coefficients. In the case of mortality by socioeconomic position, education level or income groups, higher mortality groups are selected at a faster rate than lower mortality groups (because the higher the mortality the stronger the force of selection). Therefore, the frailest individuals in these groups are selected out at a faster pace. Consequently, at the same age, what is left in the high mortality group is a more selected population in terms of robustness, compared to the low mortality group, which undergoes a slower pace of selection. The difference between the rates of selection causes the

mortality curves to converge and gives the impression that the effect of the covariate that defines the two groups (for example education level) declines with age. Also in this case, the greater the variance of unobserved heterogeneity in the population at the initial age of observation, the stronger the selection process and, therefore, the stronger the convergence between subgroups at old ages.

Main equations of the framework of the frailty models.

Frailty models assume that every individual has a specific level of unobserved frailty, *z*, that defines its hazard in a context of proportional hazard models. There is a standard individual, whose frailty *z*, is standardized to 1, and all the others have a frailty that is proportional to the frailty of the standard individual. If $\mu(x)$ is the hazard of the standard individual (or baseline hazard), defined as a function of age and frailty:

$$\mu(x,z) = z\mu(x)$$

at any age, what is observed at the population level is the mean mortality rate at that age, $\bar{\mu}(x)$, for the survivors of each frailty. That is, the standard individual hazard multiplied by the mean frailty among survivors at that age, which is a decreasing quantity:

$\bar{\mu}(x) = \mu(x)\bar{z}(x)$ Assuming that unobserved frailty follows a Gamma distribution, the population hazard $\bar{\mu}(x)$ at any age *x* is expressed as a mixture of individual hazards $\mu(x)$, by 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 the frailty distribution with mean 1 at the initial age and $\mu(x)$ is the hazard experienced by the standard individual with frailty 1. The optimization problem estimates the baseline hazard parameters and the variance of the frailty in the population.

Survival analysis without unobserved heterogeneity

The only variability controlled for is the one explained by the observed covariates, *u*, included in the model. Their effect on the baseline hazard $\mu_0(x)$ is estimated as follows:

$$\mu_i(x \mid u_i) = \mu_0(x) e^{\beta u_i} \tag{2}$$

The likelihood function in case of right censored and left truncated survival data is:

$$L(\beta, \theta) = \prod_{i=1}^{n} \frac{\left(\mu(x_i, \theta) e^{u_i \beta}\right)^{\delta_i} S(x_i, \theta)^{e^{u_i \beta}}}{S(y_i, \theta)^{e^{u_i \beta}}} \tag{3}$$

Where for each individual *i*, y_i is the entry time, x_i in the exit time, δ_i is the status (1=dead, 0=right censored), u_i is the covariate profile with effect β and $\mu(.)$ denotes the hazard, $S(.)$ the survival function and θ is the vector of parameters of the baseline hazard.

Univariate frailty models

An individual random effect for the frailty is introduced in the model as a multiplicative term on the baseline hazard:

$$\mu_i(x \mid u_i, z_i) = z_i \mu_0(x) e^{\beta u_i} \tag{4}$$

The likelihood function in case of right censored and left truncated survival data is:

$$L(\beta, \theta, \sigma^2) = \prod_{i=1}^{n} \frac{\left(\frac{\mu(x_i, \theta) e^{u_i \beta}}{1 + \sigma^2 M(x_i, \theta) e^{u_i \beta}}\right)^{\delta_i} \left(1 + \sigma^2 M(x_i, \theta) e^{u_i \beta}\right)^{-\frac{1}{\sigma^2}}}{\left(1 + \sigma^2 M(y_i, \theta) e^{u_i \beta}\right)^{-\frac{1}{\sigma^2}}} \tag{5}$$

Where for each individual *i*, y_i is the entry time, x_i in the exit time, δ_i is the status (1=dead, 0=right censored), u_i is the covariate profile with effect β and $\mu(.)$ denotes the hazard, $M(.)$ the cumulative hazard, θ is the vector of parameters of the baseline hazard and σ^2 is the variance of frailty.

Shared frailty models

In the case of repeated survival spells for the same individual i, the shared frailty models assume that those spells share the same hidden frailty, as showed by equation (6):

$$\mu_i(x \mid u_{i,j}, z_i) = z_i \mu_0(x) e^{\beta u_{i,j}} \tag{6}$$

Where the indexes *j* and *i* represent the survival spell *j* of the individual (cluster) *i*.

The cluster (individual) likelihood function in case of right censored and left truncated survival data is (2):

$$L_i = \left(\prod_{j=1}^{n_i} \left(\mu(x_{ij}, \theta) e^{u_{ij}\beta} \right)^{\delta_{ij}} \right) \frac{\Gamma\left(\frac{1}{\sigma^2} + D_i \right)}{\Gamma\left(\frac{1}{\sigma^2} \right)} \left(\sigma^2 \right)^{D_i} \left(1 - \sigma^2 \sum_{j=1}^{n_i} \ln\left(S_{ij}\left(y_{ij}, \theta \right)^{e^{u_{ij}\beta}} \right) \right)^{\frac{1}{\sigma^2}} \left(1 - \sigma^2 \sum_{j=1}^{n_i} \ln\left(S_{ij}\left(x_{ij}, \theta \right)^{e^{u_{ij}\beta}} \right) \right)^{-\frac{1}{\sigma^2} - D_i} \tag{7}$$

Where for each j-th individual in the i-th cluster, y_{ij} is the entry time, x_{ij} in the exit time, δ_{ij} is the status (1=dead, 0=right censored), u_{ij} is the covariate profile with effect β and $\mu(.)$ denotes the hazard, $S(.)$ the survival function, θ is the vector of parameters of the baseline hazard, σ^2 is the variance of frailty and $D_i = \sum \delta_{ij}$.

The overall likelihood function is simply:

$$L(\beta,\theta,\sigma^2)=\prod_{i=1}^{n}L_i \qquad (8)$$

B. Exponential model

Table B1 reports the results of the exponential model with age as covariate. The exponential baseline hazard, $\mu(x)=\lambda$, is constant and does not change with age. This allows us to include the age as a covariate and to have it interacted with the covariate for education level. The aim is to investigate whether there is a statistically detectable convergence of hazards at old ages by education group, by testing whether there is a significant interaction between the variables education and age.

The single parameter baseline hazard was modulated by the covariate for the age groups. Equations 9 and 10 describe the hazard and the survival functions of the exponential model with covariates.

$$\mu(x) = \lambda e^{\beta\, cov} \qquad (9)$$

$$S(x) = (e^{-\lambda x})^{e^{\beta\, cov}} \qquad (10)$$

The identity between an exponential hazard modulated by an age covariate and the Gompertz model makes such exponential models appropriate for human adult mortality data. The age was divided into two groups: 50-80 and 80+. Education was divided into three groups: low, medium and high. In addition to age and education, the model controlled also for period effects by introducing a variable for the calendar years.

For the sake of simplicity table B1 does not report the coefficients for the period variable and for the λ parameter of the exponential hazard. The results show that the risk of death is inversely proportional to the educational level. However, the relative difference between low education

and high education narrows at older ages and the reduction is more pronounced among women than among men.

A likelihood ratio test between the simple model without age-education interaction and the model which includes such interaction was performed. The test showed that the interaction term significantly improved the fit of the model.

Table B1. Mortality rate ratios between education groups and age groups estimated from an exponential survival hazard model with covariates education, age and their interaction. The table also reports the likelihood ratio test between this model and a model without an age-education interaction term.

	Men				Women			
	50-80 years		80+ years		50-80 years		80+ years	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
High	1	-	1	-	1	-	1	-
Medium	1.234	1.209-1.259	1.082	1.048-1.116	1.250	1.213-1.289	1.040	1.008-1.073
Low	1.571	1.544-1.598	1.172	1.143-1.202	1.594	1.552-1.637	1.170	1.136-1.202
Likelihood ratio test with reduced model (without age-education interaction)								
	D statistics: 395.193		Df: 2	p-value:0.000	D statistics:319.833		Df: 2	p-value:0.000

1. Manton KG, Stallard E, Vaupel JW. Methods for comparing the mortality experience of heterogeneous populations. Demography. 1981;18(3):389-410.
2. Van den Berg GJ, Drepper B. Inference for Shared-Frailty Survival Models with Left-Truncated Data. IZA Discussion Paper No 6031. 2011.