Life expectancy by socioeconomic status: which model fits best?

Damien Bricard*, Carlo G. Camarda and Emmanuelle Cambois

Institut National d'Études Démographiques

Extended abstract submitted to the
European Population Conference 2016, Mainz, Germany
December 2015

Abstract

There is a need for routine life tables by socioeconomic status (SES) to monitor social inequalities in life and health expectancies. However, estimating mortality risks by SES requires large population datasets, with variables of social status, linked to vital statistics. Accurate datasets are scarce and samples are usually relatively small. Routine production of LE by SES therefore requires modeling mortality risks with a great variety in the methods and assumptions that can potentially be used. In this study, we use the census sample mortality follow-up to compare the accuracy of four models for estimating LE by SES. We used the French "Permanent demographic sample". EDP-Men (aged 30-100) are distributed according to 3 educational levels. We use deaths occurred in a given year between 2008 and 2013 for EDM-men who were surveyed once in the 5 preceding years. Four different Models are used to estimate LE at age 65 for each year between 2008 and 2013. Bayesian Information Criterion (BIC) indicates the "best" estimate for each year, in a statistical point of view. In our sample, LE at age 35 was around 45,5 years. It did not progress much over the 2008-2013 period, but with fluctuations. We confirmed the gap between the men in high-educated group and the men in low-educated group, reaching 6 years in 2013. The four Models provide different estimates of LE, differences being smaller than 1 year. Model 1 and 4 provides estimates which are closer to the raw data. Further analysis are needed to determine which of the Model is accurate to estimate and monitor LE differentials across educational levels. Replication with data for women and using other criteria for SES should bring new elements to formulate some recommendations.

KEYWORDS: Differential mortality; Educational level; Smoothing; Poisson; Socioeconomic status.

*Corresponding author: Institut National d'Études Démographiques. 133, Bd Davout, 75980 Paris Cédex 20, France; Tel. +33(0)1 5606 2155: damien.bricard@ined.fr

1 Introduction

The issue of reducing social inequalities in health, as well as improving the average population health level is now a well-established objective of health public policy in most European countries. These social inequalities in health and mortality are particularly important in France and in all developed countries and their costs are estimated to 20% of health spending in Europe (Mackenbach et al. 2010). A large literature documented the extent of inequalities in health and mortality, and showed significant inequalities in favor of the most educated, the most qualified workers and individuals with highest income and wealth (Cutler et al., 2006; Lantz et al., 2010; Mackenbach et al., 2011, 2008; Van Doorslaer and Koolman, 2004).

Estimating mortality risks by socioeconomic status (SES) requires a large population dataset comprising variables of social status and being linked to vital statistics. Previous studies in France have measured social inequalities in mortality using census samples with a mortality follow-up, which were initially set-up in the 1960's (Calot and Febvay, 1965; Desplanques, 1985). More recently, these datasets allowed producing various estimates of mortality by SES highlighting the excess mortality of the least educated or skilled groups of the population (Cambois, 2004a,b; Leclerc et al., 2006). Less frequent were the studies providing life tables and life expectancies (LE) by SES. In France, the National Institute for Economic and Statistical Studies produces every 5 years the life tables by occupational status to monitor trends in mortality differentials (Blanpain, 2011; Desplanques, 1993; Mesrine, 1999; Monteil and Robert-Bobée, 2005). The last results showed persistent and large differences in LE: for the period 2000-2008, manual workers' LE at 35 years reached 40.9 years while it is 47.2 for men in highly qualified occupations (Blanpain, 2011). Due to the relative scarcity of death events in age and occupation specific sub-groups, these LE estimates rely on large periods of mortality record (9 or ten years), limiting the analysis of thinner changes. In parallel, life tables by occupational classes were computed occasionally for more restraint periods of time to produce health expectancies and monitor simultaneously the health and mortality dynamics (Cambois et al., 2011; Cambois and Robine, 2001). There is a need for routine life tables by SES to monitor and analyze in a regular way the social differentials in life and health expectancies, for implementing public health and social policies (Jagger et al., 2013; Rechel et al., 2013). Recently the population census changed in France from an exhaustive population census every 9 or 10 years to a census with a yearly rotating sample over a 5 year-period. We took the opportunity of this change to test the accuracy of a routine production of life tables for educational groups based on the new design of the census sample mortality follow-up. This new design provides data more regularly but on smaller samples and requires modeling mortality risks with a great variety in the methods and assumptions that can potentially be used.

In this study, we compare the estimates produced by several methods to make recommendations on the choice of methods according to social variables and to provide a regular follow-up of social inequalities in mortality in France.

2 Methods

2.1 Data

We use the data from the Permanent Demographic Sample (EDP) elaborated with the 1968 French population census which records the information featuring in the census for all persons born on one of the designated EDP days. This sample was initially selected on 4 dates of birth and representative of 1% of the population and increased to 16 dates of birth (or 4% of the population) since 2008. EDP individuals are followed-up across censuses and civil statistics for every demographic events recorded (birth(s), marriage(s), divorce(s), death).

Since 2004, the French census is based on 5 annual census survey (ACS) being representative of the French population after the 5 year cycle is completed. Each ACS samples consist in 1/5 of the French municipalities bellow 10.000 inhabitants and a 8% rotating sample of the households in municipalities over 10.000 inhabitants. Then, the 5 ACS provides the population data for all small municipalities and 40% of the large ones. The EDP individuals are all the individuals born one of the 16 selected days of the year who participated to the ACSs. Compared to the prior census design, the new non-exhaustive design does not guarantee the follow-up of every single EDP individuals. However, it allows having yearly information. Given this feature, we organize our data to get the annual mortality records for the EDP individuals tracked back in five prior ACS samples (Fig. 1). For example, we collect data on deaths over the year 2008 among the surviving EDP individuals tracked in the ACSs 2004 to 2008.

In the aggregated datasets we used, age 27 was the first available age; we started our analysis from age 31 over which we had the same amount of information, i.e. the same number of ACS. We stop our analysis at age 100 where data were unreliable and extremely fluctuating.

We use the International Standard Classification of Education and distribute the population in three groups of education[1]. The "low education" group gathers people with no education or primary school education with the related minimum French diploma and certificates (ISCE 0-2 primary and lower secondary education). The "middle education"

[1] `uis.unesco.org/education/pages/international-standard-classification-of-education.aspx`

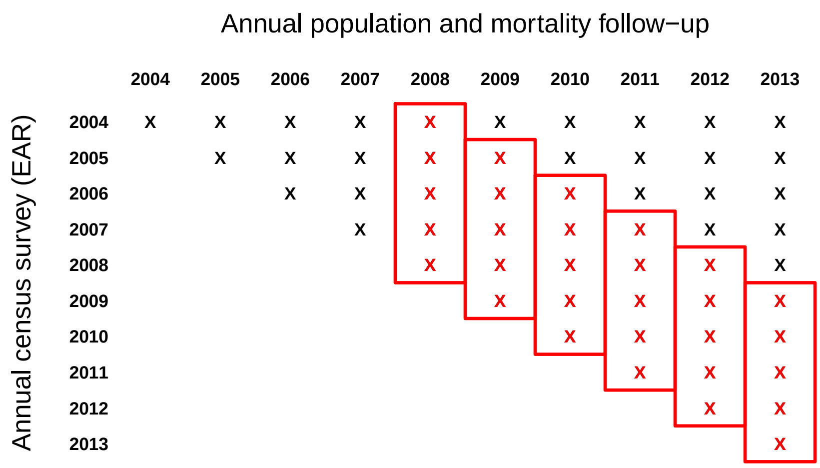

Figure 1: Annual mortality follow-up using annual census survey and mortality records

group gathers people who either went through vocational training or passed the French degree Baccalaureate (ISCE 3-4 upper secondary education). The "high education" group corresponds to people in tertiary education getting university or high school diploma (ISCE 5-6 tertiary education).

Mortality was measured by the date of death. In our datasets, for a given year, e.g. 2008, we can estimate mortality risks based on 1267312 individuals aged from 31 to 100 years old, and on 17123 deaths.

2.2 Models

Let define deaths and exposures for each year as the following matrices:

$$\boldsymbol{D} = (d_{x,k}) \quad \text{and} \quad \boldsymbol{E} = (e_{x,k})$$

where x and k are ages (from 31 to 100) and category (from educational level "primary" to "tertiary", respectively.

We decide to test 4 different models which are based on the same assumptions:

$$d_{x,k} \sim \text{Poi}(e_{x,k} \cdot \mu_{x,k})$$

i.e. deaths are realizations from a Poisson distribution where the expected values are the product between exposures and force of mortality. The aim of the paper is to test different assumptions and find a good representation of the force of mortality $\mu_{x,k}$. Table 1 presents briefly four models we will estimate.

Model 1. assumes independent estimation for each educational level and a smooth mortality trajectory over age. Moreover, dealing with adult mortality, we enforce a in-

Table 1: Lists of the four estimated models.

1.	Smooth class-specific:	$\ln(\mu_{x,k}) = s(x)_k$, $s(\cdot)$ smooth
2.	Piece-wise Proportional Hazard:	$\ln(\mu_{x,k}) = \beta_x + \theta_k$
3.	Smooth Proportional Hazard:	$\ln(\mu_{x,k}) = s(x) + \theta_k$
4.	Additive smooth:	$\ln(\mu_{x,k}) = s(x) + \theta(x)_k$, $\theta(\cdot)$ smooth

creasing monotonicity, too.

Model 2. is a simple piece-wise constant hazard model in which each age is estimated independently and the effect of a change in educational level is equal regardless the age and it is multiplicative with respect to the force of mortality.

Likewise the previous model, Model 3. assume a proportional hazard assumption over the covariate education, reference trajectory is here a smooth and monotonically increasing curve.

Finally, Model 4. attempt to gather the advantages of the previous models by assuming smoothness, monotonicity, relationship between educational levels, but without forcing a proportional hazard assumption. This additive model assumes a reference smooth curve over age and additional smooth functions which describe different between reference and other educational levels. Additionally we force monotonicity over educational attainment, i.e. a lower educational level is not allow to have lower mortality than a higher educational level. In formulas:

$$\text{if } s(x) = s(x)_{ter} \quad \Rightarrow \quad 0 = \theta(x)_{ter} \leq \theta(x)_{sec} \leq \theta(x)_{pri} \quad \forall x\,.$$

It is easy to view the additive smooth model as a generalization of the third model with the relaxation of the proportional hazard assumption. On the contrary, the additive smooth model can be also seen as a restricted version of the smooth class-specific model in which we have an additional constraint on the gradient over the educational levels.

The advantage of the Poisson assumption lays in the common estimation procedure for all four models. A year-specific likelihood approach has been employed, and an additional penalization has been added to enforce smoothness and monotonicity whenever it was needed. We have adapted and further developed approaches previously presented by Bollaerts et al. (2006); Camarda (2008, 2012); Currie et al. (2006); Eilers and Marx (1996).

Furthermore, this allows us to statistically compare models on the based on, for instance, the Bayesian Information Criterion (BIC) which measures the relative quality of statistical models for a given set of data (Schwarz, 1978). The BIC takes the achieved goodness of fit (measured in Poisson settings by the Deviance) and it penalizes with the complexity of the model. In our settings complexity refers to the number of parameters in case of Model 2. or to the effective dimensions in case of smooth models (Hastie and

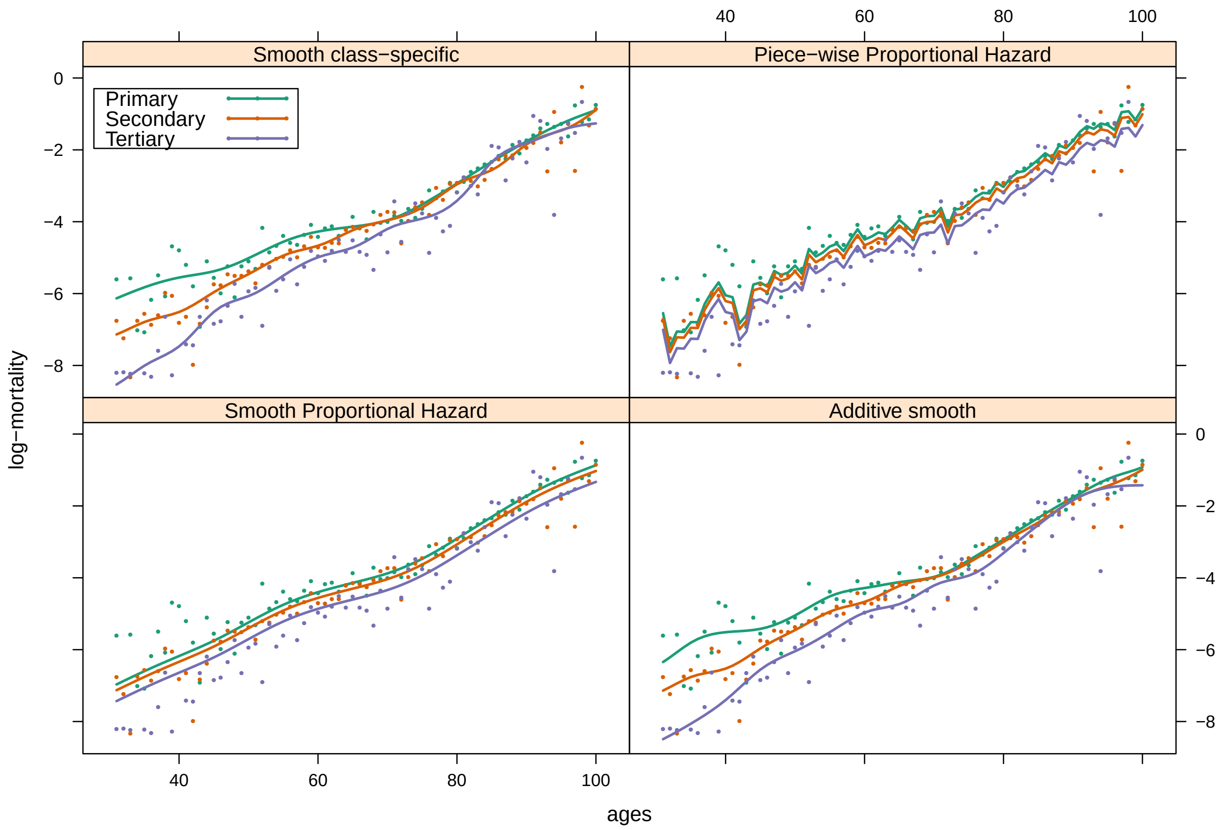

Figure 2: Actual and fitted death rates by 4 different models (in log scale). France, Males, 2010.

Tibshirani, 1990).

3 Preliminary results

Figure 2 presents the result for 2010 by educational levels and estimated models. Similar outcomes could be found for the remaining years.

Table 2 shows the estimated values for θ_k for both Piece-wise Proportional Hazard and Smooth Proportional Hazard models. There is not much difference between these tow models in terms of distance in mortality among educational levels.

Table 3 indicates that each year, Model 1 or Model 4 provide the lowest BIC.

In our sample, LE at age 35 was around 45.5 years and did not progress much over the 2008-2013 period. However, we found fluctuations over time in LE for the total sample and the three groups. We found no variation of LE across the 4 models for the middle-educated with a LE slightly above the total LE for each year. We confirmed the gap between the men in high-educated group and the men in low-educated group; it was of more than 7 years of LE in 2008, year at which it was the highest, and less than 6 years in

Table 2: Coefficients for the educational level for two models

		Piece-Wise PH		Smooth PH	
year	Primary (ref.)	Secondary	Tertiary	Secondary	Tertiary
2008	0.000	-0.280	-0.754	-0.276	-0.751
2009	0.000	-0.194	-0.498	-0.193	-0.496
2010	0.000	-0.160	-0.464	-0.160	-0.464
2011	0.000	-0.236	-0.508	-0.236	-0.508
2012	0.000	-0.228	-0.647	-0.227	-0.646
2013	0.000	-0.188	-0.473	-0.189	-0.472

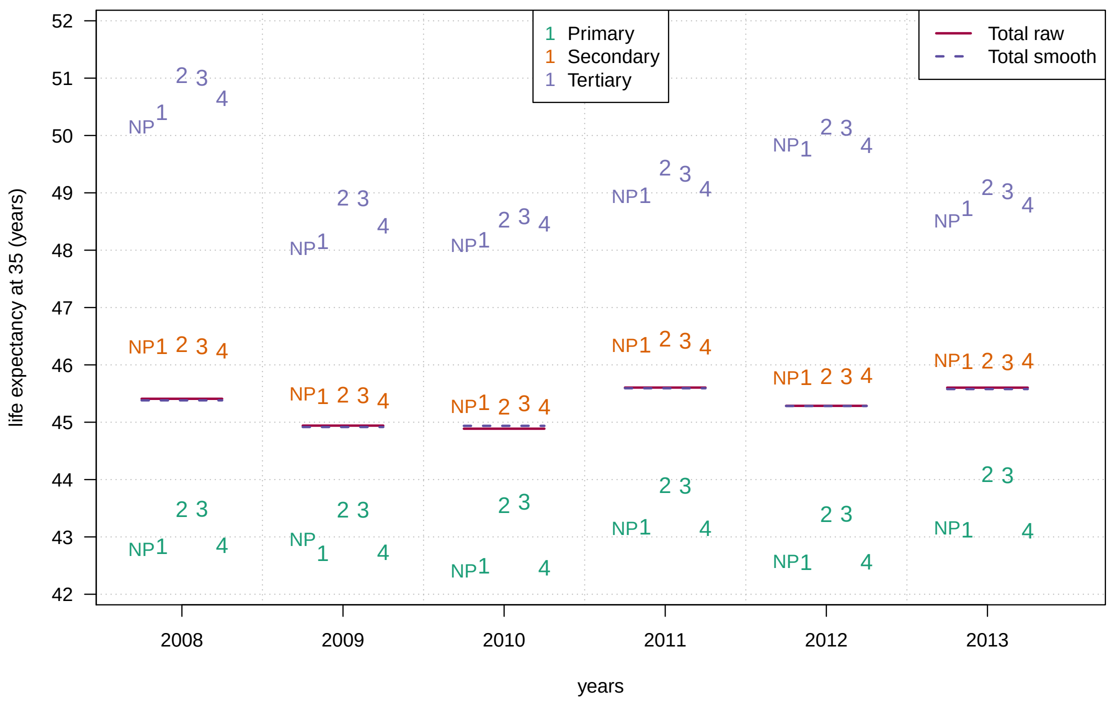

Figure 3: Life expectancy at 35 over year by educational level and estimated models. The NP depicts life expectancy at 35 computed directly on the raw data. Horizontal lines show life expectancy at 35 regardless the educational level, and for raw as well as for smooth data.

Table 3: Bayesian Information Criterion for the suggested four models. In red the lowest BIC of the year

	2008	2009	2010	2011	2012	2013
Smooth	1872.1	1609.7	2053.8	1576.8	1803.1	1898.7
PW PH	1988.2	1864.8	2308.7	1778.6	2112.3	2258.3
Smooth PH	2104.9	1883.7	2574.6	1825.7	2258.3	2460.3
Smooth Additive	1894.2	1603.8	2062.8	1573.2	1823.0	1896.0

2013. Models 2 and 3 provide higher LE than Models 1 and 4 for the both low-educated and the high-educated groups. But the differences are smaller than 1 year. Model 1 and 4 provides estimates which are close to the raw data.

4 Some conclusions

Whatever the model used, we confirmed the gap between the men in high-educated group and the men in low-educated group. We need to determine which method should be the more accurate to produce routinely life tables for educational groups, balancing the precision of the model, the likelihood of the estimated mortality curves and the potential for replicating estimations year after year. The recommendation should indeed take into account the need for monitoring trends over time in SES differentials in LE with the same method. Replicating the estimations for women or using other SES criteria than education (occupation, activity status) or eventually with data from other countries should also help figuring out which method provide the more stable results.

References

Blanpain, N. (2011). L'espérance de vie s'accroît, les inégalités sociales face à la mort demeurent. *Insee Première 1372*, 1–4.

Bollaerts, K., P. H. C. Eilers, and I. van Mechelen (2006). Simple and multiple P-splines regression with shape constraints. *British Journal of Mathematical and Statistical Psychology 59*, 451–469.

Calot, G. and M. Febvay (1965). La mortalité diffŕentielle suivant le milieu social: présentation d'une méthode expérimentée en france sur la période 1955-1960. *Etudes et Conjoncture 20*(11), 75–159.

Camarda, C. G. (2008). *Smoothing Methods for the Analysis of Mortality Development.* Ph. D. thesis, Programa de Doctorado en Ingeniería Matemática. Universidad Carlos III, Departamento de Estadística, Madrid.

Camarda, C. G. (2012). MortalitySmooth: An R Package for Smoothing Poisson Counts with *P*-Splines. *Journal of Statistical Software 50*, 1–24. Available on www.jstatsoft.org/v50/i01.

Cambois, E. (2004a). Careers and mortality: Evidences on how far occupational mobility predicts differentiated risks. *Social Science and Medicine 58*, 2545–2558.

Cambois, E. (2004b). Occupational and educational mortality differentials: magnitude and trends over last decades. *Demographic Research 2*, 278–304.

Cambois, E., C. Laborde, I. Romieu, and J.-M. Robine (2011). Occupational inequalities in health expectancies in france in the early 2000s: Unequal chances of reaching and living retirement in good health. *Demographic Research 25*, 407–436.

Cambois, E. and J.-M. Robine (2001). Social inequalities in disability-free life expectancy in the french male population. *Demographic Research 38*(4), 513–524.

Currie, I. D., M. Durbán, and P. H. C. Eilers (2006). Generalized Linear Array Models with Applications to Multidimentional Smoothing. *Journal of the Royal Statistical Society B 68*, 259–280.

Cutler, D. M., A. Deaton, and L.-M. A. (2006). The determinants of mortality. *Journal of Economic Perspectives 20*(3), 97–120.

Desplanques, G. (1985). *La mortalité des adultes: résultats de deux études longitudinales (période 1955-1980).* Les Collections de l'INSEE. D102.

Desplanques, G. (1993). L'inégalité sociale devant la mort. In INSEE (Ed.), *La société francaise: données sociales 1993*. Paris: INSEE.

Eilers, P. H. C. and B. D. Marx (1996). Flexible Smoothing with B-splines and Penalties (with discussion). *Statistical Science 11*, 89–102.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized Additive Models*. Chapman & Hall.

Jagger, C., M. McKee, K. Christensen, K. Lagiewka, W. Nusselder, H. Van Oyen, E. Cambois, B. Jeune, and J. Robine (2013). Mind the gap–reaching the european target of a 2-year increase in healthy life years in the next decade. *European Journal of Public Health 23*(5), 829–833.

Lantz, P., E. Golberstein, J. S. House, and J. Morenoff (2010). Socioeconomic and behavioral risk factors for mortality in a national 19-year prospective study of us adults. *Social Science and Medicine 70*(10), 1558–1566.

Leclerc, A., J. Chastang, G. Menvielle, and D. Luce (2006). Socioeconomic inequalities in premature mortality in france: have they widened in recent decades? *Social science and medicine 68*(2), 2035–2045.

Mackenbach, J. P., W. J. Meerding, and A. E. Kunst (2011). Economic costs of health inequalities in the european union. *Journal of Epidemiology and Community Health 65*(5), 412–419.

Mackenbach, J. P., I. Stirbu, A. J. R. Roskam, M. M. Schaap, G. Menvielle, M. Leinsalu, A. E. Kunst, and The European Union Working Group on Socioeconomic Inequalities in Health (2008). Socioeconomic inequalities in health in 22 european countries. *The New England Journal of Medicine 358*(23), 2468–2481.

Mesrine, A. (1999). Les différences de mortalité par milieu social restent fortes. In *La société francaise: données sociales 1999*. INSEE.

Monteil, A. and I. Robert-Bobée (2005). Les différences sociales de mortalité : en augmentation chez les hommes, stables chez les femmes. *INSEE Premiére 1025*, 1–4.

Rechel, B., E. Grundy, J. Robine, J. Cylus, J. P. Mackenbach, C. Knai, and M. McKee (2013). Ageing in the european union. *Lancet 381*(9874), 1312–1322.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics 6*, 461–464.

Van Doorslaer, E. and X. Koolman (2004). Explaining the differences in income-related health inequalities across european countries. *Health Economics 13*(7), 609–628.