Cause of death specific cohort effects in U.S. mortality

December 26, 2019

Cristian Redondo Lourés[1] and Andrew J.G. Cairns[2]

Maxwell Institute for Mathematical Sciences and Department of Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK.

Abstract

The socioeconomic mortality gap in the United States has attracted much attention from the research community in recent years. In this paper we use stochastic mortality modelling for U.S. data for years 1989-2015 and ages 50-75, separated by gender, educational attainment, and cause of death. Through the application of a stochastic mortality model we identify distinct cohort effects for each cause of death and aim to use these to better understand variations in lifestyle between cohorts (e.g. smoking prevalence). We find that cohort effects are much stronger for causes of death with risk factors that are both well known and easily preventable. We also find that cohort effects are very similar for causes of death with similar risk factors, clearly showcasing the link between cohort effects and health behaviours of the analysed population.

Keywords

Stochastic modelling, US mortality, mortality inequalities, Bayesian methods.

Acknowledgements

This work forms part of the research programme "Modelling, Measurement and Management of Longevity and Morbidity Risk" funded by the Actuarial Research Centre of the Institute and Faculty of Actuaries, the Society of Actuaries and the Canadian Institute of Actuaries. The authors gratefully acknowledge feedback from the respective steering and oversight groups.

[1]E: C.Redondo_Loures@hw.ac.uk

[2]E: A.J.G.Cairns@hw.ac.uk W: www.macs.hw.ac.uk/∼andrewc/ARCresources/

1 Introduction

Since the publication of Lee & Carter (1992), stochastic mortality models have been extremely useful tools for the modelling and forecasting of death rates. In the years since, many other models have been proposed, for example those by Cairns et al. (2006), Renshaw & Haberman (2006), or Plat (2009). Cairns et al. (2009) analysed the performance of several different models using data from England and Wales and the United States.

In this last publication there is some discussion about the importance of cohort effects in these models. In particular, for England and Wales it is well known that people born around 1930 experienced faster mortality improvements than people born before 1925 or after 1945, see Richards et al. (2006) and Willets (2004). Cohort effects typically appear due to changes in the exposure to risk factors for certain illnesses that are linked to when a person was born, as explained in Holford (1991). For example, people who smoke tipically start doing so in early adulthood and keep this habit for most of their lives. However, the percentage of people who smoke is very different for different birth cohorts. Therefore we would expect to see a strong relationship between year of birth and deaths due to lung cancer, which will appear as a cohort effect when we fit our models.

In this paper we will fit a mortality model with a cohort effect term to the data described in Redondo Lourés & Cairns (2019). The data available consists of exposures and number of deaths separated by gender, education level, and cause of death, for the period 1989-2015 and cohorts born between 1914 and 1975. We will use groupings of causes of death that have similar underlying risk factors, and fit the model separately for each of the four different groups (low/high educated males/females). The resulting cohort effects will give us insights on the lifestyle factors driving the widening mortality gap between different education groups, which has been widely reported in the literature, see for example Case & Deaton (2015), Olshansky et al. (2012), or Jemal et al. (2008). We will fit our model to the age range 50 to 75 only, as we know the gap is wider at these relatively young ages.

2 Data modelling

The theory behind our modelling uses the standard assumption that the number of deaths in a single year and at a certain age is a Poisson random variable with expected value given by the death rate times the central exposed-to-risk:

$$D(x,t) \sim \text{Poisson}(m(x,t)E(x,t)). \tag{1}$$

Now we need to specify the linear predictor for $m(x,t)$. Due to the wide range of causes of death we are going to analyse we use a fairly flexible model, in which there are one non parametric age effect, one non parametric period effect, and a non parametric period effect interacting linearly with age, in addition to the cohort effect. This model was

proposed in Plat (2009) as a reduction of a more complex model when we are only interested in mortality at high ages. In particular, we have:

$$\log[m_{ap}(x,t)] = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_c. \tag{2}$$

This allows us to write a likelihood function in terms of the parameters we want to estimate (the α_x, $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, and γ_c) and the known exposures and number of deaths. As is the case with most stochastic mortality models we need to impose some constraints on the parameters to ensure the solution is unique. The following transformation leave the death rates (and therefore the likelihood) unchanged:

$$\begin{pmatrix} \alpha_x \\ \kappa_t^{(1)} \\ \kappa_t^{(2)} \\ \gamma_c \end{pmatrix} \rightarrow \begin{pmatrix} \alpha_x + c_1 + c_2(\bar{x} - x) \\ \kappa_t^{(1)} - c_1 \\ \kappa_t^{(2)} - c_2 \\ \gamma_c \end{pmatrix} \tag{3}$$

$$\begin{pmatrix} \alpha_x \\ \kappa_t^{(1)} \\ \kappa_t^{(2)} \\ \gamma_c \end{pmatrix} \rightarrow \begin{pmatrix} \alpha_x + \phi_1 - \phi_2 x + \phi_2 x^2 \\ \kappa_t^{(1)} + \phi_2 t + \phi_2(t^2 - 2\bar{x}t) \\ \kappa_t^{(2)} + 2\phi_3 t \\ \gamma_c - \phi_1 - \phi_2(t - x) - \phi_3(t - x)^2 \end{pmatrix} \tag{4}$$

In particular, we see that constant shifts in the values of the $\kappa_t^{(1,2)}$ and γ_c can be absorbed in the α_x, and that linear and quadratic trends in the cohort effect can be absorbed by similar linear and quadratic transformations of the other parameters. This identifiability problem can be solved by imposing the following constraints:

$$\sum_t \kappa_t^{(1)} = 0 \ , \ \sum_t \kappa_t^{(2)} = 0, \tag{5}$$

$$\sum_c \gamma_c = 0 \ , \ \sum_c c\gamma_c = 0 \ , \ \sum_c c^2\gamma_c = 0. \tag{6}$$

Equations (5) mean that the period effects will average to zero, and equations (6) that the cohort effect will average to zero and also not have a linear or quadratic trend. These constraints will be imposed *a posteriori*, that is, we will first estimate the parameters using one of the methods described later in this section, and then we apply a transformation that ensures our estimates fulfil equations (5) and (6).

We will use two approaches to the estimation problem. First, we will use simple maximum likelihood estimation, which is completely straightforward to implement in R thanks to the StMoMo package (see Villegas et al. (2018) for details). The log-likelihood function is, up to a constant:

$$\ell_p = \sum_{x,t} \left[-E \exp(\alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_c) + D(\alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_c) \right], \tag{7}$$

where we have omitted the (x,t) dependence of the deaths, D, and exposures E, to simplify the notation. Confidence intervals are produced by bootstrapping using the function provided in that same R library.

Typically, when modelling cause of death data, the maximum likelihood estimates of non-parametric functions are very noisy, and that noise makes it difficult to identify true patterns in the cohort effect. Therefore we will use a second approach in which Bayesian techniques will help us fit a time series to $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, and γ_c. Forcing the period and cohort effects to follow a certain time series model will result in smoother estimates in which trends will be more readily visible. We will need to add new terms to the likelihood coming from the time series structure of $\kappa_t^{(1,2)}$ and γ_c, and priors for the unknown parameters in those time series, which will also need to be estimated. This will produce a log-posterior for all parameters that we can sample from to obtain the desired credible intervals. Due to the complexity of this log-posterior we need to use Markov Chain Monte Carlo (MCMC) methods to produce our sample. Note that the only purpose of the time series in this model is obtaining a smoother in-sample fit than the one obtained through maximum likelihood estimation, and this procedure is not intended for death rate forecasting.

Using the ideas of Cairns et al. (2011) we assume the vector of period effects $\boldsymbol{\kappa_t} = (\kappa_t^{(1)}, \kappa_t^{(2)})^T$ to follow a bivariate random walk with drift $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, and γ_c to be an AR(2) process of the form $\gamma_c = (\rho + \tau)\gamma_{c-1} - \rho\tau\gamma_{c-2} + \varepsilon_c$, where $\varepsilon_c \sim \mathcal{N}(0,1)$ for all c, and $|\rho|$, $|\tau| \leq 1$ to ensure stationarity. To simplify the Markov chain algorithm we fix $\boldsymbol{\kappa_1} = 0$ and $\gamma_1 = 0$, and later convert our results to fulfil the constraints given in equations (5) and (6). In more detail, the full log-likelihood will be (ignoring constants):

$$\ell = \ell_p + \ell_{rw,\kappa} + \ell_{ar,\gamma} \tag{8}$$

where ℓ_p is given in equation (7), and the other two terms are the time series contribution to the likelihood:

$$\ell_{rw,\kappa} = -\frac{1}{2}\sum_{t=2}^{27}\left[(\Delta_{\boldsymbol{\mu}}\boldsymbol{\kappa_t})^T\Sigma^{-1}\Delta_{\boldsymbol{\mu}}\boldsymbol{\kappa_t}\right] - \frac{26}{2}\log\left(|\Sigma|\right), \tag{9}$$

$$\ell_{ar,\gamma} = -\frac{1}{2\sigma_c^2}\sum_{c=3}^{52}(\gamma_c - (\rho+\tau)\gamma_{c-1} + \rho\tau\gamma_{c-2})^2 - \frac{50}{2}\log\left(\sigma_c^2\right) - \frac{1}{2}\log(\sigma_{ac}) - \frac{1}{2\sigma_{ac}}\gamma_2^2, \tag{10}$$

where Σ is the covariance matrix associated with the random walk for the vector $\boldsymbol{\kappa_t}$; $\Delta_{\boldsymbol{\mu}}\boldsymbol{\kappa_t} = \boldsymbol{\kappa_t} - (\boldsymbol{\kappa_{t-1}} + \boldsymbol{\mu})$; ρ and τ are the two parameters of the $AR(2)$ model for γ_c, and σ_c^2 its variance; and σ_{ac}^2 is given by:

$$\sigma_{ac} = \frac{1+\rho\tau}{1-\rho\tau}\frac{\sigma_c^2}{(1+\rho\tau)^2 - (\rho+\tau)^2}. \tag{11}$$

All of these new parameters, namely Σ, $\boldsymbol{\mu}$, ρ, τ, and σ_c, are also to be estimated, and

appropriate priors need to be provided for them. A standard choice for them would be:

$$\Sigma \sim Inv\text{-}Wishart(\nu, \mathbf{S}), \tag{12}$$

$$\mu_1 \sim \mathcal{N}(0, \sigma^2_{\mu_1}), \tag{13}$$

$$\mu_2 \sim \mathcal{N}(0, \sigma^2_{\mu_2}), \tag{14}$$

$$\sigma^2_c \sim Inv\text{-}Gamma(a, b), \tag{15}$$

$$\text{logit}(\rho) \sim \mathcal{N}(0, \sigma^2_\rho), \tag{16}$$

$$\text{logit}(\tau) \sim \mathcal{N}(0, \sigma^2_\tau), \tag{17}$$

where ν, $\mathbf{S}$, $\sigma^2_{\mu_1}$, $\sigma^2_{\mu_2}$, a, b, σ^2_ρ, and σ^2_τ are hyperparameters that we need to fix in order to run our MCMC algorithm. Our results have been found to be fairly robust against changes in the values of these hyperparameters. However they need to be chosen carefully: for example, σ^2_c controls the level of smoothing we will achieve in the cohort effects. If its value is too large the result will be as noisy as the maximum likelihood estimate, but if it is too small we will see oversmoothing that destroys the real underlying trends we want to observe. For the age, period, and cohort effects we choose improper uniform priors. Since the results we obtain for γ_c with this choice are very similar to the MLEs (as we will see in the next section) we do not find it necessary to use a different prior for these parameters.

Alternative ways of producing a smooth cohort effect, such as spline-fitting or two-step time series modelling have also been considered, but none of them were found to be better than the Bayesian approach. The former also have a somewhat stronger subjective component, since in the Bayesian method the role of the smoothing parameter is played by σ_c, which is not fixed to a specific value. Instead, it is estimated from the data with a certain level of flexibility given by its prior distribution.

These two models will be fitted separately to each of our 4 subpopulations (low educated males, low educated females, high educated males, and high educated females), and for each of the following groups of causes of death:

- Lung cancer.
- Lifestyle related cancers (mouth and gullet, stomach, gut or rectum, larynx, trachea, liver, and bladder cancers).
- Prostate and/or breast cancer.
- All other cancers (genitalia, pancreas, skin, urinary organs, lymphatic, benign or unespecified tumours, and all other locations of cancer).
- Chronic lower respiratory diseases (CLRD; mostly chronic obstructive pulmonary disease, COPD, above age 50).
- Diabetes.
- All heart diseases.

- Cerebrovascular diseases.
- Other circulatory diseases.
- Alzheimer's and mental illnesses.
- Accidental deaths (excluding accidental poisoning).
- Deaths of despair (chirrosis, accidental poisonings, and suicide).
- All other causes of death.
- All cause mortality.

There were two reasons to choose this grouping of causes of death. On the one hand, we tried to model groups that have a relatively large number of deaths and exposed to common underlying risk factors. For example, lung cancer and CLRD are closely related to smoking, while individual causes within the heart disease group exposed to a combination of smoking and poor diet/sedentary lifestyle. On the other hand, we grouped causes of death so that the ICD standard change described in Redondo Lourés & Cairns (2019) would not result in a big discontinuity in the death rates between the years 1998 and 1999, which would complicate the modelling exercise.

3 Results: cohort effects for different causes of death

We will now present the fitted cohort effects for different population groups and causes of death. We will mainly focus on causes of death with well known risk factors, which we will try to link to the observed patterns in γ_c.
Figure 1 shows the cohort effects for lung cancer and CLRD for all four population groups (low educated males, high educated males, low educated females, and high educated females). The fans show the 60%, 75%, and 90% confidence intervals. The orange fan in the backgrounds comes from maximum likelihood estimation and bootstrapping, whereas the grey fan in the foreground is the result obtained from the Bayesian approach. It can be seen that the Bayesian approach captures the main features of the MLE estimates, while avoiding most of the volatility resulting from sampling variation in the number of deaths.
It is interesting to note that the shape of γ_c is very similar for both causes of death within each of the four groups. This similarity is to be expected, since both lung cancer and CLRD are very heavily linked to cigarette smoking, see US Department of Health and Human Services (2004). The clear differences between males and females are therefore caused by the different patterns of smoking prevalence for different cohorts in each gender. It is also clear that, although smaller than the differences between genders, there is also a difference between the cohort effects for education groups, which reflects the different smoking behaviour within each gender. Note that, due to the constraints introduced by equations (6), the cohort effect cannot have linear or quadratic trends,

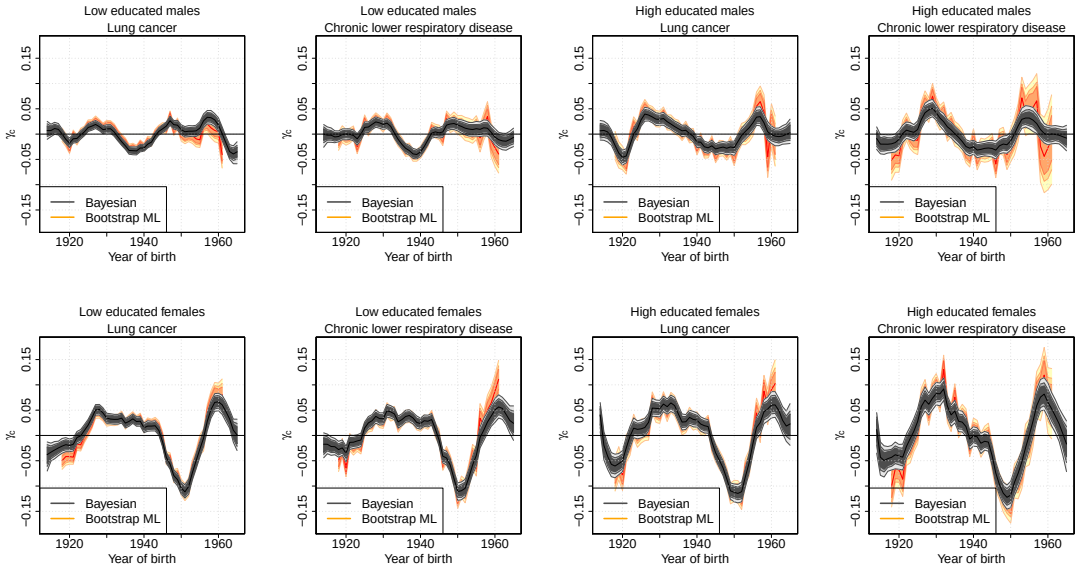

Figure 1: Cohort effects for CLRD and lung cancer for all four population groups, with 60%, 75%, and 90% confidence intervals. Orange and background is the result of the MLE-bootstrapping procedure. The grey fan is the result of the Bayesian method.

therefore these γ_c do not directly represent the proportion of smokers in each cohort. The overall decrease in smoking for younger cohorts will be absorbed by the period trends, and γ_c will tell for which cohorts the decrease has been faster or slower than the general trend. We discuss the link with smoking prevalence further in section 4.

In Figures 2-5 we plot the cohort effects for a selection of causes of death in each of the four groups analysed. The causes of death chosen have different links to lifestyle factors, so analysing the different trends observed for each of them in each population group can help us understand the different underlying health behaviours.

The general picture is that for causes of death with preventable risk factors (such as lifestyle cancers and deaths due to suicide, accidental poisoning, and cirrhosis) the cohort effects have clearer trends than they do for causes of death for which controllable risk factors do not exist or are not well established (prostate/breast cancer or other cancer, for example). In particular, we see that the trends for diabetes and cerebrovascular disease are very similar in all groups, probably due to the influence that poor diet and lack of exercise has on both of these causes of death (see, for example, figures 2 and 3). Heart disease, one of the most common causes of death in the US, shows similar patterns as well, although there seems to be a small bump for cohorts born in the 1920s. This is probably caused by the big influence of smoking on heart disease mortality (note that the cohort effect for lung cancer was also relatively high for these cohorts in all four groups), much higher than its impact in diabetes or cerebrovascular disease.

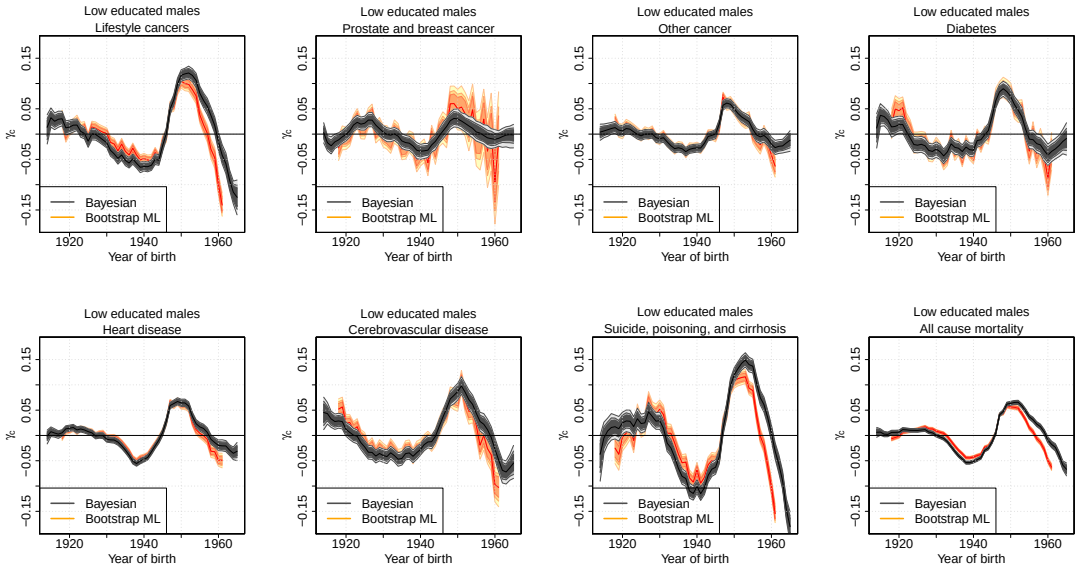

Figure 2: Cohort effects for several causes of death in low educated males, with 60%, 75%, and 90% confidence intervals. Orange and background is the result of the MLE-bootstrapping procedure. The grey fan is the result of the Bayesian method.

For causes of death for which the onset of the illness is not easily controllable, cohort effects tend to have small or no trends at all. For example, the "other cancer" category does not have any obvious trends except for low educated males. Lifestyle related cancers, on the other hand, have clear patterns in all groups, clearly showcasing the close relationship between cohort effects and health behaviours. Small trends in causes of death for which there are not any known effective prevention strategies can still arise through mechanisms other than changing health behaviours. In particular, changing access to healthcare, treatments, or screening programs can affect mortality due to most cancers (see Freeman (2004) for more detail), which may play a role in the trends observed for prostate and breast cancer. The existence of unknown lifestyle related risk factors is another reason why cohort effects trends can appear for these causes of death. For example, there might be a relation between smoking and cancer of the urinary organs, but evidence is not conclusive. These "hidden risk factors" can play a role in sharp changes in the trend of γ_c such as the one observed for other cancers in low educated males (figure 2) born around 1950. These were cohorts which suffered particularly bad health outcomes, and this peak seems to confirm the correlation between poor health behaviours and cancers for which lifestyle links have not been yet confirmed in the medical literature.

Looking at each group separately, we can see that for low educated males (Figure 2) γ_c sharply increases for cohorts born in the late 1940s for all the causes of death shown.

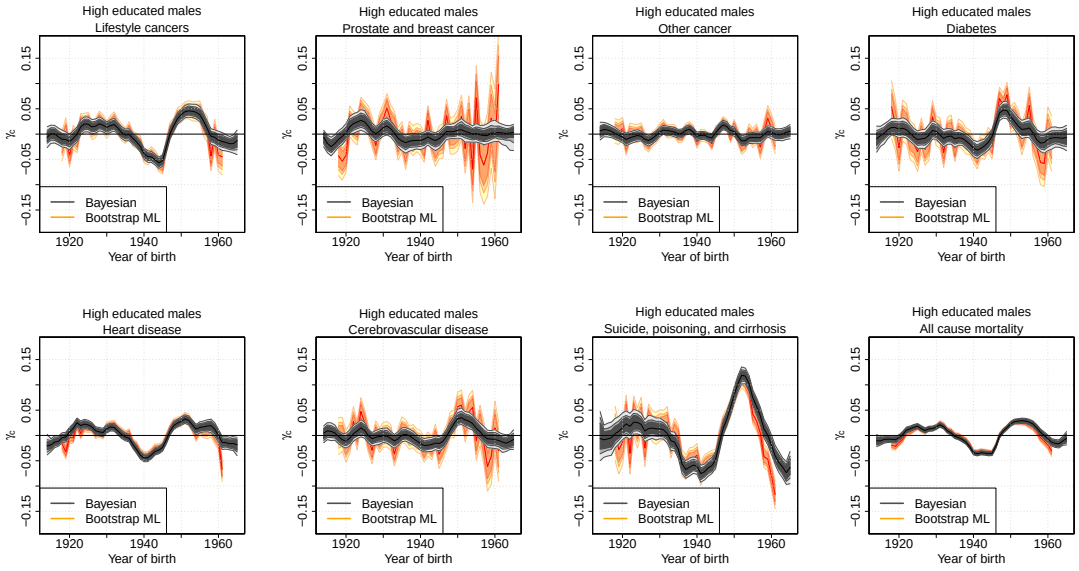

Figure 3: Cohort effects for several causes of death in high educated males, with 60%, 75%, and 90% confidence intervals. Orange and background is the result of the MLE-bootstrapping procedure. The grey fan is the result of the Bayesian method.

This might be related to the higher mortality experienced in later life by males who were drafted into the Vietnam war, see Schlenger et al. (2015). That is followed by a rapid decline for cohorts born in the 1960s. The strenght of this effect, even for causes of death with few or none known controllable risk factors, points out to a high prevalence of poor health behaviours in this group.

For high educated males (Figure 3) the picture is not so extreme. Although γ_c also increases for the Vietnam cohorts for most diseases highly correlated with lifestyle factors (diabetes, lifestyle cancers, heart disease, or sucide, poisoning, and cirrhosis), the increase is much smaller than for their lower educated peers. Trends in causes of death like prostate and breast cancer or other cancers have also disappeared, signalling the higher prevalence of generally healthy lifestyles in this group.

In the case of females, generally speaking (Figures 4 and 5), cohort effect trends are much smaller than they are for males. The patterns for both education groups are also very similar in most causes of death, probably pointing out that education plays a much smaller role in the differentials in health behaviour for females than they did for males. The most remarkable difference is the appearance of a (relatively speaking) small cohort effect for breast cancer in low educated females, which does not appear so clearly in the high educated group. Since breast cancer is not easily preventable this difference might be caused by differences in diagnosis rates or access to treatment, as speculated earlier.

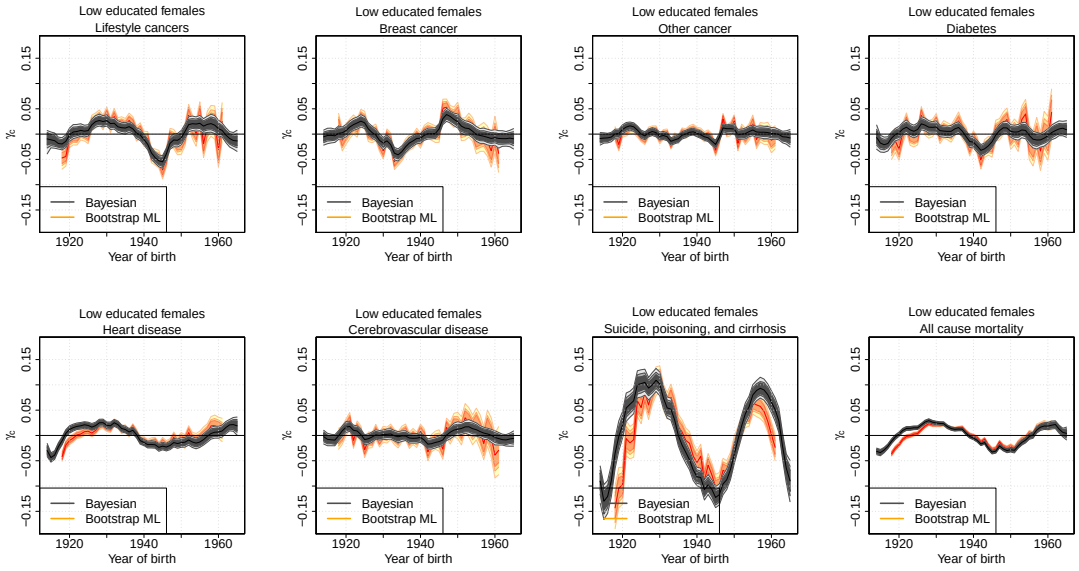

Figure 4: Cohort effects for several causes of death in low educated females, with 60%, 75%, and 90% confidence intervals. Orange and background is the result of the MLE-bootstrapping procedure. The grey fan is the result of the Bayesian method.

4 Smoking: comparison with survey data

In the previous section we singled out the results obtained for lung cancer and CLRD since a very large proportion of deaths due to these causes are linked to smoking. We claimed that the cohort effects seen in figure 1 were linked to the smoking trends in the underlying population, but that due to the identifiability problem the values of γ_c were not trivially related to the ratio of smokers. In this section we will compare the two and see that, even though the numerical values are not directly related, the trends of both do have clear similarities. Because of the relatively high quality data on smoking prevalence of the United States population, and the very strong link between smoking and lung cancer, this exercise can be used to test our main hypothesis that cohort effects reflect the changes in health behaviours of the population analysed.

We will use data on smoking habits of the US population form the National Health Interview Survey, run by the Centers for Disease Control and Prevention, CDC (2019). We use data for years 1990-2015, excluding 1996 (for which, along with 1989, data on smoking habits was not recorded.), and ages 30-79. Even though data is available for single ages and calendar years the relatively small sample size forces us to group the data in order to obtain less noisy estimates of the ratio of smokers in the population.

Instead of focusing on single years, we look at cohorts. For each cohort in the data we calculate the average ratio of ever smokers, weighting each (age, year) cell by the number of respondents. Note that, as smokers tend to die earlier in life, the actual ratio

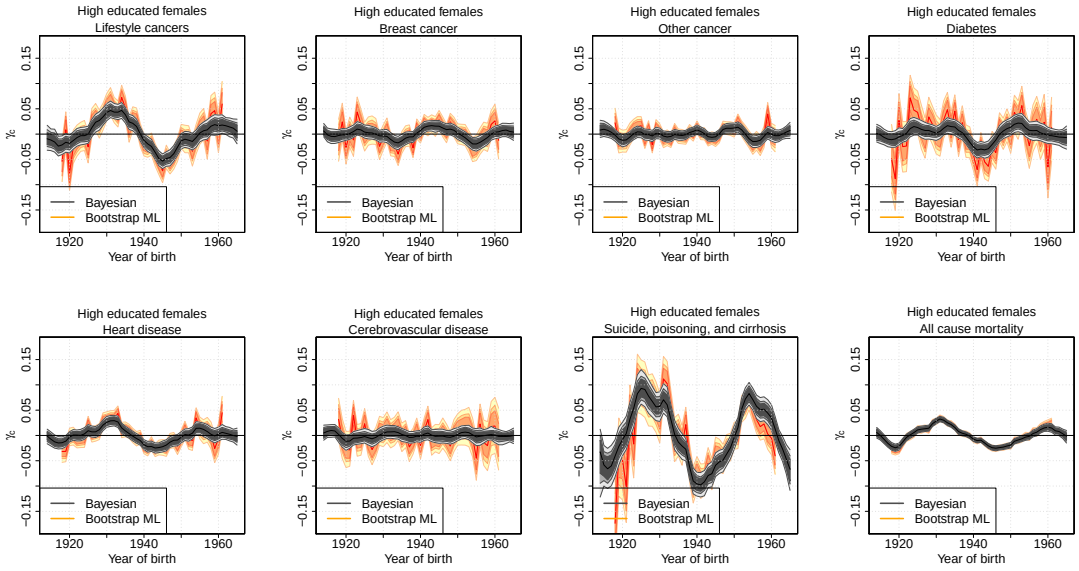

Figure 5: Cohort effects for several causes of death in high educated females, with 60%, 75%, and 90% confidence intervals. Orange and background is the result of the MLE-bootstrapping procedure. The grey fan is the result of the Bayesian method.

of smokers *decreases* as a cohort ages. Since we are only observing early cohorts at very high ages we will be slightly underestimating the true ratio of ever smokers in these groups. However, since our mortality data covers similar ages and calendar years, we expect the resulting cohort effects to be similarly biased.

Figure 6 shows, for each of the four subpopulations analysed, the average ratio of smokers (black line, left axis) and the median of the bayesian estimate for the lung cancer cohort effect (red line, right axis). As expected the cohort effect does not capture the general upwards or downwards trends due to the constraints imposed by (6), as explained earlier. However, big fluctuations of the cohort effects tend to appear whenever the general trend of smoking prevalence changes.

For example, for low educated males the big fall in the value of γ_c in the mid to late 1930s (A) comes with a corresponding fall in the smoking prevalence. Moreover, the peak of γ_c in the mid 1950s (B) is accompanied by a strange feature in the smoking prevalence curve, which shows a small "bump" that breaks an otherwise almost linear fall (which would be absorbed in $\kappa_t^{(1)}$, since γ_c is not allowed to have linear trends). The situation is very similar in low educated females: the overall increase in smoking prevalence is being transferred in the $\kappa_t^{(1,2)}$ parameters, but the cohort effect still reflects changes in the underlying trend. For example, the rapid increase in smoking for females born in the mid 1920s (C) is accompanied by a rise in the cohort effect, and a slight fall on the ratio of ever smokers around 1950 (D) sees γ_c fall very rapidly. For high educated females the

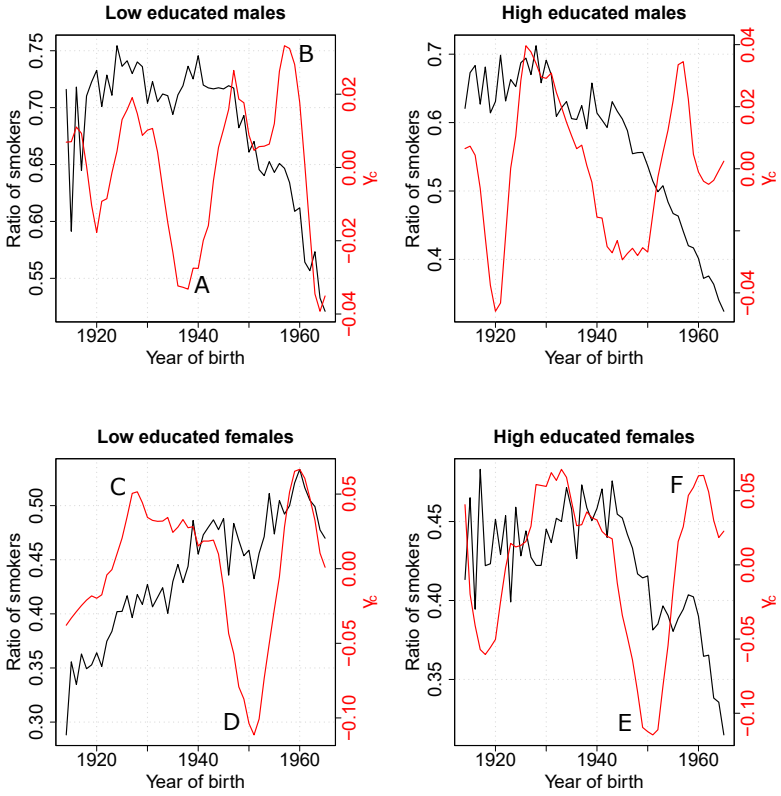

Figure 6: Average ratio of smokers (black line, left axis) and median of the bayesian estimate for the lung cancer cohort effect (red line, right axis) as a function of year of birth for all four subpopulations.

two most prominent features of γ_c are its rapid fall after 1940 (E), which coincides with a fast decrease in the ratio of smokers after years of slow but steady rise, and a peak around 1960 (F), which, similarly to low educated males, happens at the same time as a sharp increase in smoking prevalence, breaking the linear declining trend.
A remarkable exception to this ratio of smokers-cohort effect similarities is the big peak that γ_c shows for high educated males in the mid 1950s. For this group survey data shows a constant decline in smoking prevalence, whereas lung cancer mortality data would suggest that a small departure from the trend should have happened for these cohorts. This could be caused by poor data quality or changes in the underlying habits within the ever-smoker population (for example, a smaller ratio of heavier smokers could result in higher overall rates of lung cancer). A full analysis of this anomaly is however beyond the scope of this paper.
In summary, this section shows that the cohort effects plotted in figure 1 are indeed related to the trends of smoking prevalence in the US. This validates our approach of

using cause of death cohort effects as a proxy for lifestyle behaviours.

5 Conclusions

In this paper we have used statistical modelling in order to understand the lifestyle factors that drove the mortality inequalities described in Redondo Lourés & Cairns (2019). We have seen that cause of death specific cohort effects are an important tool for the understanding of the underlying changes in health behaviours of the population being analysed. In particular, the cohort effects for causes of death with very similar risk factors show very similar trends, which can help us identify an underlying "lifestyle cohort effect".

We have also seen that cohort effects are much stronger for causes of death with lifestyle-related risk factors, further reinforcing the idea that cohort effects arise as a consequence of changes in the health behaviours of the underlying population. With this in mind, we can clearly see that the evolution of these health behaviours has been very different for males and females and, particularly in the case of males, for people of different socioeconomic status (for which education is used as a proxy in this work). This points to diverging lifestyles as the cause underlying the increase in the mortality gap in recent years. For example, the faster increase in γ_c for low educated males born in the late 1940s (who would be now in their early 70s) with respect to their higher educated peers, and for almost all preventable causes of death, points at lifestyle choices as the main driver of excess early mortality in low educated American males.

One downside of this method is that the existence of identifiability problems in the model somehow complicates the direct interpretation of the cohort effects, and comparison with survey data is not completely straightforward. How to transform the known trends of a lifestyle risk factor (for example, smoking prevalence) into a cohort effect that can be used in the model, or use the fitted cohort effect as a way to estimate the prevalence of these factors where survey data is not available, is an interesting problem that will be the subject of future research.

Bibliography

Cairns, A. J. G., Blake, D. & Dowd, K. (2006), 'A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration', *Journal of Risk and Insurance* **73**(4), 687–718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. & Balevich, I. (2009), 'A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States', *North American Actuarial Journal* **13**(1), 1–35.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D. & Khalaf-Allah, M. (2011), 'Bayesian Stochastic Mortality Modelling for Two Populations', *ASTIN Bulletin* **41**(1), 29–59.

Case, A. & Deaton, A. (2015), 'Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century', *Proceedings of the National Academy of Sciences* **112**, 15078–15083.

CDC (2019), 'National Health Interview Survey (Accessed November 2019)'.
URL: *https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm*

Freeman, H. P. (2004), 'Poverty, Culture, and Social Injustice: Determinants of Cancer Disparities', *CA: A Cancer Journal for Clinicians* **54**(2), 72–77.

Holford, T. R. (1991), 'Understanding the Effects of Age, Period, and Cohort on Incidence and Mortality Rates', *Annual Review of Public Health* **12**(1), 425–457.

Jemal, A., Ward, E., Anderson, R. N., Murray, T. & Thun, M. J. (2008), 'Widening of Socioeconomic Inequalities in U.S. Death Rates, 1993–2001', *PLOS ONE* **3**(5), e2181.

Lee, R. D. & Carter, L. R. (1992), 'Modeling and Forecasting U.S. Mortality', *Journal of the American Statistical Association* **87**(419), 659–671.

Olshansky, S. J., Antonucci, T., Berkman, L., Binstock, R. H., Boersch-Supan, A., Cacioppo, J. T., Carnes, B. A., Carstensen, L. L., Fried, L. P., Goldman, D. P., Jackson, J., Kohli, M., Rother, J., Zheng, Y. & Rowe, J. (2012), 'Differences In Life Expectancy Due To Race And Educational Differences Are Widening, And Many May Not Catch Up', *Health Affairs* **31**(8), 1803–1813.

Plat, R. (2009), 'On stochastic mortality modeling', *Insurance: Mathematics and Economics* **45**(3), 393–404.

Redondo Lourés, C. & Cairns, A. J. G. (2019), 'Mortality In The US By Education Level', *to appear in Annals of Actuarial Science* .

Renshaw, A. & Haberman, S. (2006), 'A cohort-based extension to the Lee-Carter model for mortality reduction factors', *Insurance: Mathematics and Economics* **38**(3), 556–570.

Richards, S. J., Kirkby, J. G. & Currie, I. D. (2006), 'The Importance of Year of Birth in Two-Dimensional Mortality Data', *British Actuarial Journal* **12**, 5–38.

Schlenger, W. E., Corry, N. H., Williams, C. S., Kulka, R. A., Mulvaney-Day, N., DeBakey, S., Murphy, C. M. & Marmar, C. R. (2015), 'A Prospective Study of Mortality and Trauma-Related Risk Factors Among a Nationally Representative Sample of Vietnam Veterans', *American Journal of Epidemiology* **182**(12), 980–990.

US Department of Health and Human Services (2004), 'The Health Consequences of Smoking: A Report of the Surgeon General'.

Villegas, A. M., Kaishev, V. K. & Millossovich, P. (2018), 'StMoMo: An R Package for Stochastic Mortality Modeling', *Journal of Statistical Software* **84**(3), 1–38.

Willets, R. C. (2004), 'The Cohort Effect: Insights and Explanations', *British Actuarial Journal* **10**, 833–877.