Cause of death specific cohort effects in US mortality

Cristian Redondo Loures, Andrew J. G. Cairns

Heriot-Watt University, Edinburgh

Longevity 15, Washington, September 2019

HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries
Outline

• Introduction
• Data availability
• A look at crude death rates
• Stochastic modelling
• Conclusions
Introduction

Gain insight into the impact of socioeconomic status on mortality to inform projections. Analyse which causes of death affect different groups within a population. Try to understand the drivers of the observed behaviour through the analysis of cohort effects.

US mortality by single ages and calendar years, independently for each gender, CoD, and education level.
Introduction

Education as a covariate in death rates

Education ⇒ Socioeconomic status.

“Fixed” through adulthood.

Data readily available (quality to be assessed though!).

Data availability

Data combines the Centers for Disease Control and Prevention public database (number of deaths) and Human Mortality Database/Current Population Survey mid-year population estimates.

Calendar years 1989-2015
Years of birth 1914-1970
2 genders
2 education levels
13 groups of causes of death (more possible too!)
Data availability

List of causes of death:

1. Lung and bronchus cancer
2. Lifestyle cancers
3. Prostate or breast cancer
4. Other cancers
5. Chronic lower respiratory disease
6. Diabetes
7. Heart disease
8. Cerebrovascular disease
9. Other circulatory disease
10. Dementia and other mental illness
11. Accidental
12. Suicide, poisoning, and cirrhosis
13. All other
ALL CAUSE MORTALITY
A look at crude death rates

Initial/final gap males: 2.86 → 1.20
Initial/final gap females: 2.62 → 1.12
Death rates by education, age 60 by year

**Age 60, males**
**All cause mortality**

- ROI for low: 0.28%
- ROI for high: 1.67%

**Age 60, females**
**All cause mortality**

- ROI for low: 0.08%
- ROI for high: 1.09%

Initial/final gap males: 1.75 → 2.54
Initial/final gap females: 1.63 → 2.13
CAUSE OF DEATH MORTALITY
Death rates by education, age 60 by year

Age 60, males
Prostate cancer

Death rate (log scale)

Year

Age 60, males
Lung cancer etc

Death rate (log scale)

Year

Initial → final ratio for prostate cancer: 1.24 → 1.76

Initial → final ratio for lung cancer: 2.21 → 3.49
Death rates by education, age 60 by year

Age 60, females
Breast cancer

Age 60, females
Lung cancer etc

Death rate (log scale)

Year

1.05 → 1.14
Initial→final ratio for breast cancer

1.63 → 2.84
Initial→final ratio for lung cancer
STOCHASTIC MODELLING
We will fit a stochastic model for the mortality rates independently to each gender, education group, and group of causes of death, for ages 50-75 and years 1989-2015.

Our goal is to identify distinct trends in the cohort effects for different causes of death that point at both the drivers behind mortality for each cause and the changing behaviours of the population analysed.
Model for the mortality rates $m(x, t)$:

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

where

$$\log(m(x, t)) = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_{t-x}$$

Plat model restricted to high ages.

$\gamma_{t-x} \equiv \gamma_c$ cohort effect. From a formal standpoint: mortality is a function of year of birth.
How do cohort effects arise in practice?

Non-homogeneities that are strongly linked to when a person was born, and not to the period they live in. Examples:

- Smoking
- Type of work (blue collar, white collar)
- Diet (home cooked/processed food)

Typically, health behaviours fall into this category.
Males, Lung Cancer Death Rates
Ratio of Low Education to High Education

Year
Age

Ratio
45
50
55
60
65
70

Cristian Redondo Loures, Andrew J. G. Cairns
Cause of death specific cohort effects in US mortality
Identifiability constraints for the Plat model:

\[ \sum_t \kappa_t^{(1)} = 0, \quad \sum_t \kappa_t^{(2)} = 0, \quad \sum_c \gamma_c = 0, \]

\[ \sum_c c \gamma_c = 0, \quad \sum_c c^2 \gamma_c = 0. \]

All time series parameters will average to zero, and the cohort effect \( c \) will have no linear or quadratic trends!
Two approaches:

Estimate the $\alpha$’s, $\kappa$’s, and $\gamma$’s from maximum likelihood, bootstrap to obtain CIs.

Assume time series structure for $\gamma_c$, $\kappa_t^{(1)}$, and $\kappa_t^{(2)}$, and use Bayesian techniques.

For the second approach we assume:

\[
\gamma_c \text{ is an AR}(2) \text{ process,}
\]

\[
\kappa_t = \begin{pmatrix} 
\kappa_t^{(1)} \\
\kappa_t^{(2)}
\end{pmatrix}
\text{ is a RW with drift } \mu.
\]
Bayesian approach

Priors for the AR(2) parameters:

\[
\begin{align*}
\sigma_c^2 & \sim \text{Inv-Gamma}(a, b) \\
\logit(\rho) & \sim \mathcal{N}(0, \sigma_\rho^2) \\
\logit(\tau) & \sim \mathcal{N}(0, \sigma_\tau^2)
\end{align*}
\]

Priors for the random walk:

\[
\begin{align*}
\Sigma & \sim \text{Inv-Wishart}(\nu, S) \\
\mu_1 & \sim \mathcal{N}(0, \sigma_{\mu_1}^2) \\
\mu_2 & \sim \mathcal{N}(0, \sigma_{\mu_2}^2)
\end{align*}
\]
Bayesian approach

The final contributions to the log posterior are:

\[
\log \left( P(\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)}, \gamma_c, \sigma_c, \rho, \tau, \Sigma, \mu_1, \mu_2 | D, E) \right) \propto \\
\ell_P + \ell_{ar\gamma} + \ell_{rw\kappa} + \\
\log(\pi(\sigma_c)) + \log(\pi(\tau)) + \log(\pi(\rho)) + \\
\log(\pi(\Sigma)) + \log(\pi(\mu_1)) + \log(\pi(\mu_2))
\]

Not known analytical solutions. MCMC methods required.
Cohort effects

Cristian Redondo Loures, Andrew J. G. Cairns

Cause of death specific cohort effects in US mortality
Cohort effects

Low educated females
Lung cancer

Year of birth

γ_c
Bayesian
Bootstrap ML

1920 1940 1960
−0.10 0.00 0.05 0.10
Low educated females
Chronic lower respiratory disease

Year of birth

γ_c
Bayesian
Bootstrap ML

1920 1940 1960
−0.10 0.00 0.05 0.10

Cristian Redondo Loures, Andrew J. G. Cairns

Cause of death specific cohort effects in US mortality
Cohort effects

Low educated males
Lung cancer

\[ \gamma_c \]

Bayesian
Bootstrap ML

Low educated males
Diabetes

\[ \gamma_c \]

Bayesian
Bootstrap ML

Cristian Redondo Loures, Andrew J. G. Cairns
Cause of death specific cohort effects in US mortality
Cohort effects

Cohort effects with 60%, 75%, and 90% CIs

Low educated males, heart disease

\( \gamma_c \)

Year of birth
Cohort effects for different causes of death
Low educated males, heart disease

\[ \gamma_c \]

Heart disease, Lung cancer, Diabetes, Linear comb

Year of birth
1920 1930 1940 1950 1960

Cristian Redondo Loures, Andrew J. G. Cairns

Cause of death specific cohort effects in US mortality
Cohort effects for different causes of death
High educated males, heart disease

\[ \gamma_c \]

Heart disease
Lung cancer
Diabetes
Linear comb

Year of birth
1920 1930 1940 1950 1960
−0.10 0.00 0.05 0.10
Cohort effects for different causes of death
Low educated females, heart disease

\[ \gamma_c \]

- Heart disease
- Lung cancer
- Diabetes
- Linear comb

Year of birth

1920 1930 1940 1950 1960

-0.10 0.00 0.05 0.10
Cohort effects for different causes of death
High educated females, heart disease

\[ \gamma_c \]

Year of birth

- Heart disease
- Lung cancer
- Diabetes
- Linear comb

1920 1930 1940 1950 1960

-0.10 0.00 0.05 0.10
Conclusions

→ Modelling cause of death data provides insight into the reasons behind the evolution of all cause mortality.

→ The shape of the cohort effects for different causes of death is linked to different underlying risk factors.

→ Different groups show different cohort effect patterns, reflecting different health behaviours.

→ Complex interactions between lifestyle factors are behind inequalities in death rates in the US.
Future work

There are several parts of this work that can be extended in the future:

Systematise the cohort effect analysis.

Mortality projections.

Applications to annuities and life insurance.

Extension of the analysis to other populations.
Thank You!

Questions?
Poisson log-likelihood (both approaches):

\[ \ell_P \propto \sum_{x,t} \left[ -E_{xt} \exp(\alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_c) + 
D_{xt}(\alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_c) \right] \]

Parameters easily estimated using R (StMoMo library). Bootstrapping techniques help obtain confidence intervals.
Random walk log-likelihood for $\gamma_c$:

$$\ell_{rw\gamma} \propto -\frac{1}{2\sigma_c^2} \sum_{c=3}^{n_c} (\gamma_c - (\rho + \tau)\gamma_{c-1} + \rho\tau\gamma_{c-2})^2$$

$$- \frac{n_c - 2}{2} \log(\sigma_c^2) - \frac{1}{2} \log(\sigma_{ac}) - \frac{1}{2\sigma_{ac}^2} \gamma_c^2,$$

where we have used the autocovariance

$$\sigma_{ac} = \frac{1 + \rho\tau}{1 - \rho\tau} \frac{\sigma_c^2}{(1 + \rho\tau)^2 - (\rho + \tau)^2}.$$
Random walk log-likelihood for $\kappa_t$:

$$l_{rw\kappa} \propto -\frac{1}{2} \sum_{t=2}^{n_t} \left[ (\kappa_d)^T \Sigma^{-1} \kappa_d \right] - \frac{n_t - 1}{2} \log (|\Sigma|),$$

with

$$\kappa_d = \begin{pmatrix} \kappa_t^{(1)} - (\kappa_{t-1}^{(1)} + \mu_1) \\ \kappa_t^{(2)} - (\kappa_{t-1}^{(2)} + \mu_2) \end{pmatrix}$$
Priors:

\[ \sigma_c \sim \text{Inv-Gamma}(10, 5 \cdot 10^{-4}), \]
\[ \Sigma \sim \text{Inv-Wishart} \begin{pmatrix} 10, \begin{pmatrix} 10^{-4} & 0 \\ 0 & 10^{-4} \end{pmatrix} \end{pmatrix}, \]
\[ \mu_1 \sim \mathcal{N}(0, 1^2), \]
\[ \mu_2 \sim \mathcal{N}(0, 1^2), \]
\[ \text{logit}(\rho) \sim \mathcal{N}(0, 0.5^2), \]
\[ \text{logit}(\tau) \sim \mathcal{N}(0, 0.5^2). \]
Parameters with 60%, 75%, and 90% CIs
Low educated males, heart disease

Year of birth
\(\alpha_x\)

Bayesian
Bootstrap ML

Parameters with 60%, 75%, and 90% CIs
Low educated males, lung cancer

Year of birth
\(\alpha_x\)

Bayesian
Bootstrap ML
Parameters with 60%, 75%, and 90% CIs
Low educated males, heart disease

\[ \kappa_t(1) \]

Bayesian
Bootstrap ML

Year of birth

1990 2000 2010

\[ \kappa_t(2) \]

Bayesian
Bootstrap ML

Year of birth

1990 2000 2010

Cristian Redondo Loures, Andrew J. G. Cairns
Cause of death specific cohort effects in US mortality
Residuals for lung cancer, low educated males
Residuals for diabetes, low educated males

- Year: 1990, 2000, 2010
- Age: 50, 55, 60, 65, 70, 75
- Birth: 1920, 1940, 1960

Standardised residuals

-3, -2, -1, 0, 1, 2, 3

Birth, Age, Year
Residuals for heart disease, low educated males

- Year range from 1990 to 2010
- Age range from 50 to 75
- Birth range from 1920 to 1960
- Standardised residuals range from -3 to 3
Residuals for all cause, low educated males
Low educated males
Lifestyle cancers

Year of birth

Prostate and breast cancer

Year of birth

Bayesian
Bootstrap ML

Cristian Redondo Loures, Andrew J. G. Cairns
Cause of death specific cohort effects in US mortality
Cause of death specific cohort effects in US mortality

Cristian Redondo Loures, Andrew J. G. Cairns
Cause of death specific cohort effects in US mortality

Cristian Redondo Loures, Andrew J. G. Cairns

Low educated males

Accidental death

Year of birth

\( \gamma_c \)

Bayesian
Bootstrap ML

Suicide, poisoning, and cirrhosis
Figure 1: Cause of death specific cohort effects in US mortality.

Low educated males
Other causes of death

Low educated males
All cause mortality