Educational attainment and cause of death mortality in the United States

Cristian Redondo Loures, Andrew J. G. Cairns

Heriot-Watt University, Edinburgh

ELSA/BVZL Annual Conference, Munich, 24th September 2018

HERIOT WATT UNIVERSITY

Actuarial Research Centre

Institute and Faculty of Actuaries

Outline

- Introduction
- Objectives
- Data sources
- Dealing with problematic exposures
- Results

Actuarial Research Centre
Institute and Faculty of Actuaries

Why education?

- Education $\Rightarrow$ Socioeconomic status.
- "Fixed" through adulthood.
- Data readily available (quality to be assesed though!).

Case and Deaton (2015) found interesting trends in cause of death mortality in the US by education level.

Objectives

Gain insight on the impact of socioeconomic status on total mortality, which can help future mortality projections. We also want to analyse which causes of death affect different education groups.

We need to develop a method to deal with unreliable exposures if we want to obtain reliable death rates.

To this end we analyse US mortality by single years of age, sex, and single calendar years, for different CoD's and education levels.

Actuarial Research Centre
Institute and Faculty of Actuaries

Education recording

We will use three education groups in our analysis:

- Low educated: 12 years of schooling or less, High School degree or less.
- Medium educated: 1-3 years of college education, some college but no Bachelor's.
- High educated: 4+ years of college education, Bachelor's degree or higher.

These take into account two different recording standards used in our datasets in the period 1989-2015.

Actuarial Research Centre
Institute and Faculty of Actuaries

Deaths data

The Centers for Disease Control and Prevention (CDC) makes anonymised data from death certificates publicly available.

- Calendar year in which the death took place.
- Sex and age of the decedent.
- Educational attainment data.
- Multiple causes of death.

Deaths data

Multiple causes of death are recorded in death certificates, but we are only interested in underlying cause.

CoD's recorded using the International Classification of Diseases (ICD). ICD-9 until 1998, ICD-10 from 1999 onwards.

Change in ICD $\Rightarrow$ Change in the rules for determining underlying CoD.

Notation

$E^C(x,b,e) \rightarrow$ Exposures by cohort for people born in year b, at age x and with education e.

$E^C_T(x,b) = \sum_e E^C(x,b,e) \rightarrow$ Total exposures for cohort born in b at age x.

$R^C(x,b,e) = E^C(x,b,e)/E^C_T(x,b) \rightarrow$ Fraction of people in the cohort born in year b that, at age x, had education e

Actuarial Research Centre
Institute and Faculty of Actuaries

Exposures data

Sources of data for $E_T^C(x,b)$ and $R^C(x,b,e)$:

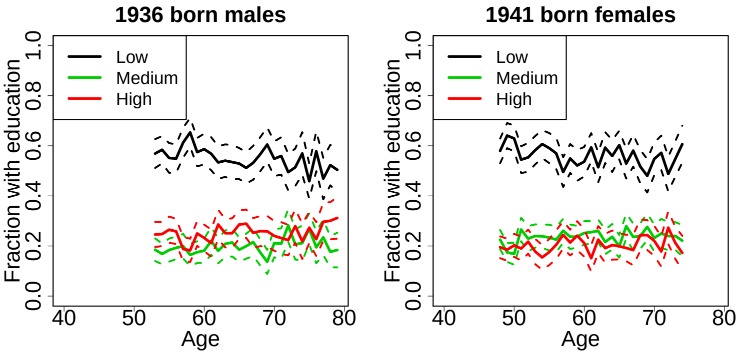
Human Mortality Database (HMD). Accurate source of total population estimates, $E_T^C(x,b)$.

Current Population Survey (CPS),

- estimates of educational attainment by sex and age ($\hat{R}^C(x,b,e)$).
- small sample ($\approx$ 60000 records per year).
- only up to age 79.

Actuarial Research Centre
Institute and Faculty of Actuaries

Ratios of educated people in the CPS are very noisy due to sampling variation.



Actuarial Research Centre
Institute and Faculty of Actuaries

Education in exposures

We will use the CPS ratios of educated people, $\hat{R}^C$, alongside HMD total exposures, E_T^C, to obtain exposures by education level, E^C.

We need to reduce the sampling noise $\Rightarrow$ Smoothing of the ratios by cohort.

We can make use of the deaths data to estimate the shape of the "real" population ratios and then fit this curve to the unsmoothed ratios.

Recurrence for the ratios:

$$R^C(x+1,b,e) = \frac{E^C(x+1,b,e)}{E_T^C(x+1,b)} = \frac{E^C(x,b,e) - \Delta(x,b,e)}{E_T^C(x,b) - \Delta_T(x,b)} =$$

$$= \frac{R^C(x,b,e)E_T^C(x,b) - \Delta(x,b,e)}{E_T^C(x,b) - \Delta_T(x,b)}.$$

$\Delta(x,b,e) \rightarrow$ Members lost by cohort born in b at age x with education e.

$\Delta_T(x,b) \rightarrow$ Total members lost by the cohort.

Actuarial Research Centre
Institute and Faculty of Actuaries

Education in exposures

We only need to estimate the initial ratio for each cohort, $R^C(x_0, b, e)$. We can do it by least squares:

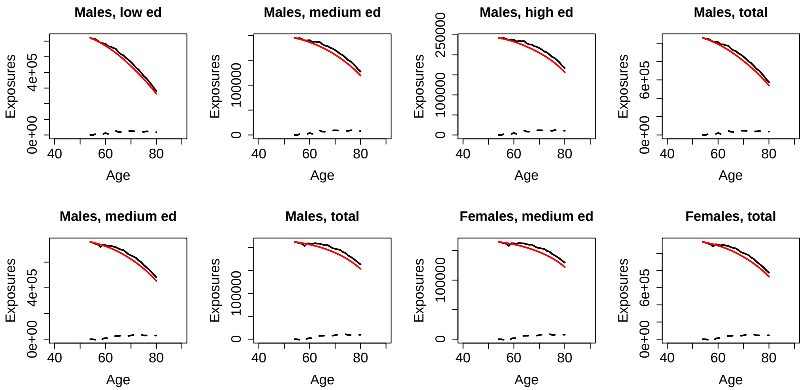
$$\mathcal{O}_b = \sum_x \left(R^C(x, b, e) - \hat{R}^C(x, b, e) \right)^2$$

$R^C(x, b, e)$ can be written in terms of $R^C(x_0, b, e)$.
$\hat{R}^C(x, b, e)$ are given by the CPS.

We need to assume the only change in cohort membership Δ is the number of deaths.

Actuarial Research Centre
Institute and Faculty of Actuaries

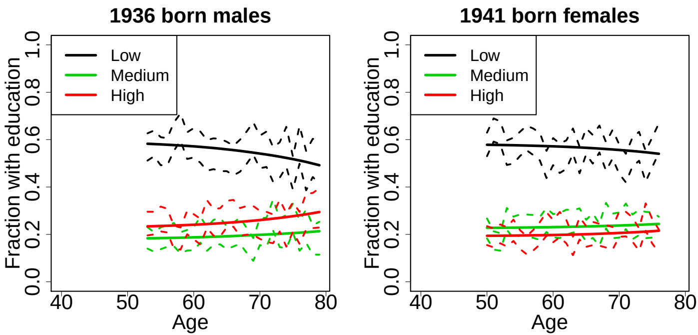
1935 cohort, $\Delta = D$ approximation



Actuarial Research Centre
Institute and Faculty of Actuaries

Education in exposures

Back to the previous cohorts, smoothed ratios:



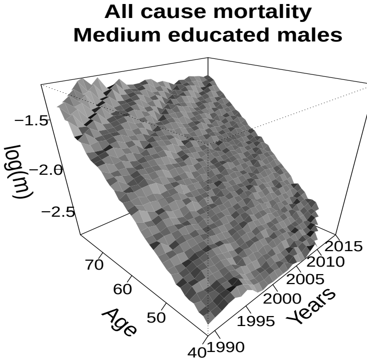
Really powerful method that allows for straightforward extrapolation to ages beyond 79!

Actuarial Research Centre
Institute and Faculty of Actuaries

Education in exposures

Cohort by cohort smoothing → "Wavy" death rates

Link cohorts by penalising concavity



Concavity penalisation: we want the death rates to be relatively linear in the log scale within calendar years.

$$\mathcal{O}_N = \sum_b \mathcal{O}_b + \sum_x \left[\alpha(C(x, b+x, e))^{2\beta}\right]$$

$$C(x,t,e) = \log(m(x,t,e)) - \frac{1}{2}\left(\log(m(x+1,t,e)) + \log(m(x-1,t,e))\right)$$

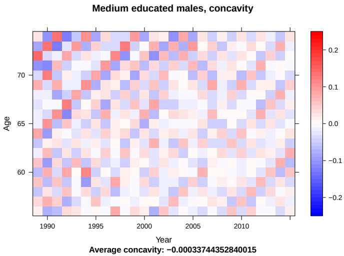
α, β parameters control the relative importance between least squares and concavity minimisation. In our analysis we use $\alpha = \beta = 1$.

Actuarial Research Centre
Institute and Faculty of Actuaries

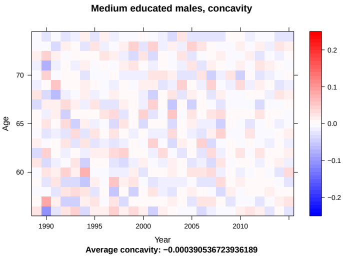
Education in exposures

Concavity heat maps:

No penalty ⇒ With penalty



⇒

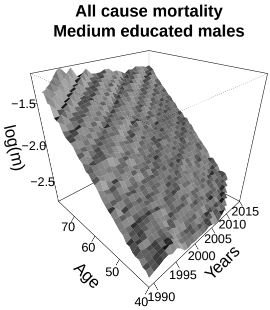


Actuarial Research Centre
Institute and Faculty of Actuaries

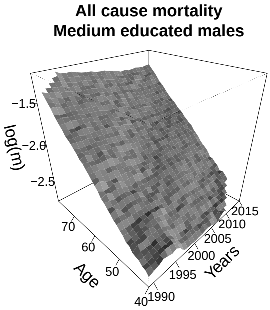
Effect of concavity correction on death rates

Link cohorts by penalising concavity:

No penalty $\Rightarrow$ With penalty



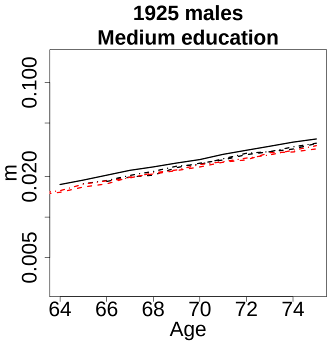
$\Rightarrow$



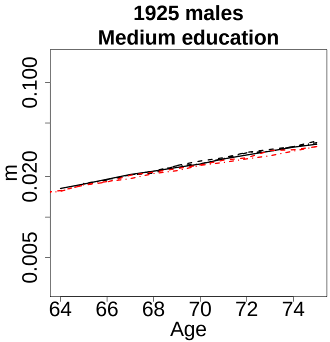
Education in exposures

Neighbouring cohorts:

No penalty ⇒ With penalty

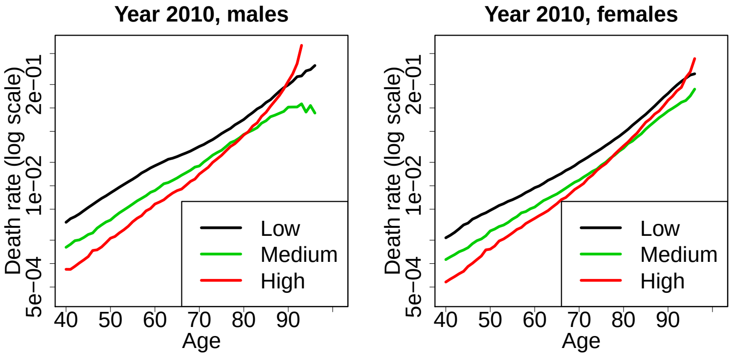


⇒



Education in deaths

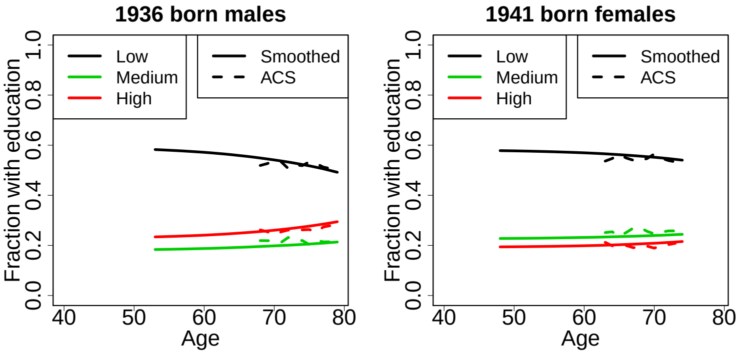
All cause mortality



Overestimated death rates for high educated people!

Actuarial Research Centre
Institute and Faculty of Actuaries

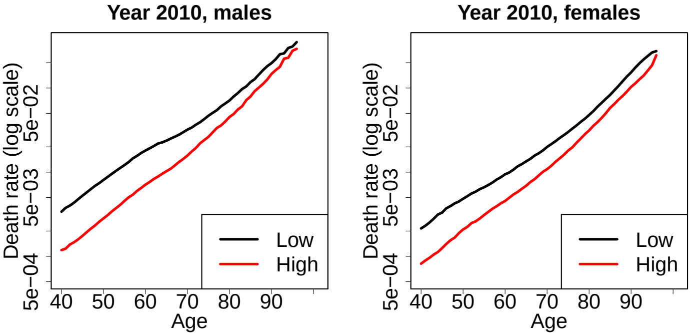
CPS and American Community Survey (ACS) comparison. No bias in the ratios → Problem must be biased reporting in the number of deaths!



Actuarial Research Centre
Institute and Faculty of Actuaries

Education in deaths

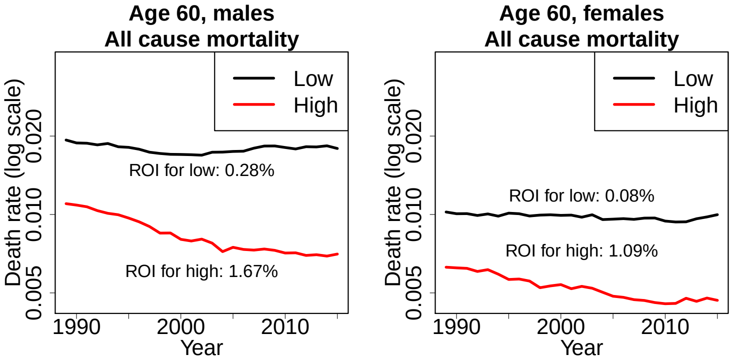
Merging the medium and high educated groups in a single group solves the issues at high ages.



Initial/final gap males: 2.86 $\rightarrow$ 1.20
Initial/final gap females: 2.62 $\rightarrow$ 1.12

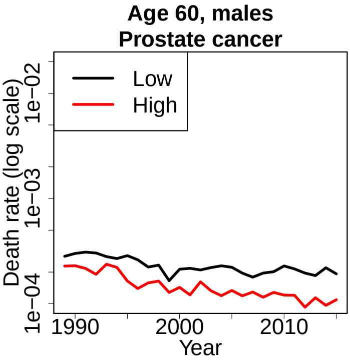
Actuarial Research Centre
Institute and Faculty of Actuaries

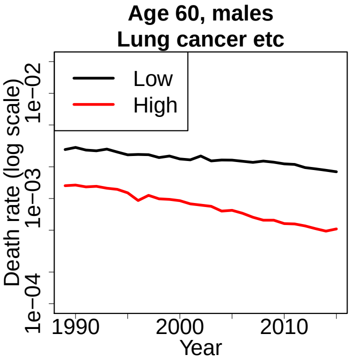
Death rates by education, age 60 by year



Initial/final gap males: 1.75 $\rightarrow$ 2.54
Initial/final gap females: 1.63 $\rightarrow$ 2.13

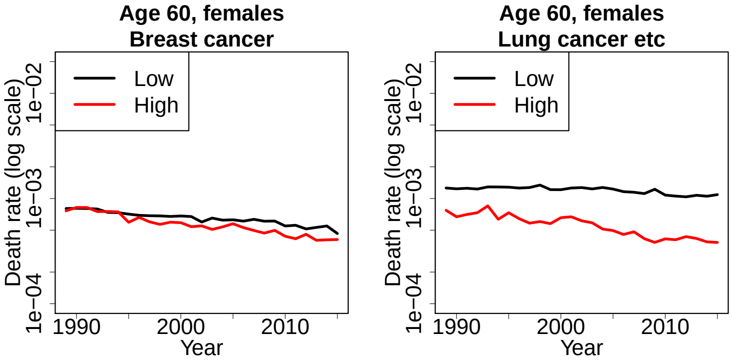
Death rates by education, age 60 by year





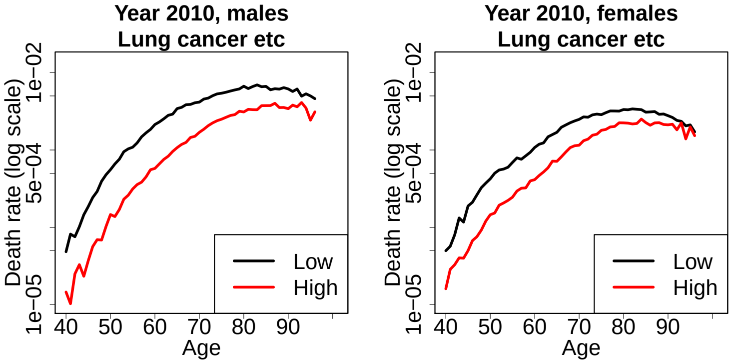
Initial/final gap prostate: $1.24 \to 1.76$
Initial/final gap lung: $2.21 \to 3.49$

Death rates by education, age 60 by year



Initial/final gap breast: 1.05 $\rightarrow$ 1.14
Initial/final gap lung: 1.63 $\rightarrow$ 2.84

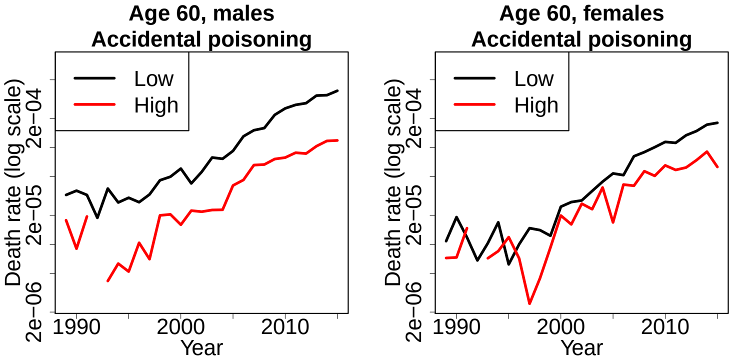
Death rates by education, 2010 by age



Initial/final gap males: 3.32 $\rightarrow$ 1.48
Initial/final gap females: 3.10 $\rightarrow$ 1.12

Actuarial Research Centre
Institute and Faculty of Actuaries

Death rates by education, age 60 by year



Huge increase for all groups!

And many more plots...

Broad causes of death, age 60 males

CoD	Gap		
	Initial	Final	Increase
Other cancer	1.30	1.98	1.52x
Ischaemic heart disease	1.75	2.39	1.36x
Other circulatory	1.96	2.62	1.34x
Respiratory cancer	2.21	3.49	1.58x
Cirrhosis, suicide, accidental poisoning	1.66	2.60	1.57x
Other	1.89	2.61	1.38x
Respiratory disease	2.38	4.33	1.82x
Accidents	1.89	2.65	1.40x
Infections	1.10	2.78	2.52x
Diabetes	2.34	2.38	1.02x
Mental and neurological	1.61	2.22	1.38x

Actuarial Research Centre
Institute and Faculty of Actuaries

Statistical testing

$\mathbf{H_0}$: Mortality for both subpopulations, L and H, is equal, only random variations cause the apparent difference. We test the hypothesis separately for:

- 5-year age groups (41-45 to 86-90)
- 30 causes of death (including all cause mortality).
- 2 genders.

Actuarial Research Centre
Institute and Faculty of Actuaries

Statistical testing

Given $\bar{D}(x,t,e) = m_T(x,t)E(x,t,e)$, the expected number of deaths in each education group if their mortality was equal to the whole population mortality:

- χ^2 test: Build $\chi^2 = \sum_x \sum_t \sum_e \frac{\left(D(x,t,e)-\bar{D}(x,t,e)\right)^2}{\bar{D}(x,t,e)}$
- Signs test: Given each residual $D(x,t,e) - \bar{D}(x,t,e)$ is as likely to be positive as it is to be negative, how likely is it we observe as many positive residuals in $D(x,t,L)$ as we do?

Actuarial Research Centre
Institute and Faculty of Actuaries

Statistical testing

We accept H_0 ($p>0.01$) in only very few cases, and usually in the extrapolated area (age 79+)

- χ^2: Accidental poisonings for females (76-90) and for males (81-90), breast cancer for males (86-90).
- Signs: Suicide for females (66-75 and 81-90), neurological illness for males (76-90) and breast cancer for males (86-90).

Actuarial Research Centre
Institute and Faculty of Actuaries

Conclusions

- Educational attainment has a very significant impact on mortality in the US.
- Educational differences play a bigger role in premature mortality than they do in death at high ages.
- Different causes of death have different gaps (although almost all of them do at all ages!).
- Differences in mortality between the two groups have been growing for the last 27 years.

Thank You!

Questions?