Cohort Effects in US Cause-of-Death Mortality Data: The Link to Controllable Risk Factors

Andrew J.G. Cairns
joint work with Cristian Redondo Lourés

Heriot-Watt University, Edinburgh

HMD Conference, 4 & 6 October 2022

HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries

SOCIETY OF ACTUARIES

Outline

- Introduction
- Data
- The CBDX model
- CBDX-I: Cohort effects for *individual* causes of death
- The Common Cohort Effects (CCE-n) model
- Results and discussion

Introduction

- Death rates have been improving gradually over the last century (at least until 2020!)
- What will mortality rates be in 2025, 2030, 2050, 2100?
- Understanding past improvements will potentially help us to forecast the future
- Cause of death data might help

Motivation: How to understand all-cause mortality data better?

- What are the drivers of historical mortality improvements?
  - Medical and public health advances
  - Individual, controllable risk factors: smoking; diet; exercise; ...
  - Access to healthcare
- Cause of death (CoD) data can give us some insights:
  Good quality data by cause of death is now available (e.g. HMD, CDC, ONS)
  CoD $\times$ single or 5-year age group $\times$ single year
- Analysis:
  - Often use 5 to 7 CoD groups (e.g. all cancers in one group)
  - But a more granular approach can give more insight into what has happened in the past
- How do we make best use of this data?
- E.g. how can we exploit this data to get better insights into historical trends in all-cause mortality?

Data

- Details: Redondo Lourés and Cairns (2019, 2021)
- US data
- Sources: CDC (deaths); HMD, Current Population Survey (exposures)
- By sex
- By education level: low ($\leq$ high school graduation); high ($>$ high school)
- Single ages (40-84) (*)
- Single years (1989-2017) (*)
- 51 causes of death (contraction of HMD intermediate list)
- (*) Excluded: the oldest and youngest cohorts (too few observations)
  (*) Included: cohorts born in 1915-1970
- US analysis here $\Rightarrow$ what is potentially feasible for other countries e.g. using the HMD causes of death database (`causesofdeath.org`)

Cause of Death Groupings

1	Infectious diseases				
2	Cancer: mouth, gullet	3	Cancer: oesophageal		
4	Cancer: stomach	5	Cancer: colon	6	Cancer: rectum, anus
7	Cancer: liver	8	Cancer: pancreas	9	Cancer: other digestive system
10	Cancer: larynx	11	Cancer: lung, bronchus, trachea	12	Cancer: skin
13	Cancer: breast	14	Cancer: cervix	15	Cancer: uterus
16	Cancer: ovary	17	Cancer: other female genital	18	Cancer: prostate
19	Cancer: other male genital	20	Cancer: bladder	21	Cancer: urinary organs
22	Cancer: lymphatic etc.	23	Benign tumours	24	Cancer: other locations
25	Blood diseases	26	Diabetes		
27	Vascular dementia	28	Other mental illness	29	Parkinson's disease
30	Alzheimer's	31	Other diseases of nervous system		
32	Blood pressure + rheumatic fever	33	Ischaemic heart diseases	34	Non-rheumatic valve disorders
35	Other heart diseases	36	Cerebrovascular diseases	37	Circulatory diseases
38	Influenza	39	Pneumonia	40	Other acute respiratory infections
41	Chronic Obstructive Pulmonary Disease	42	Other respiratory diseases		
43	Liver cirrhosis	44	Other liver diseases	45	Other digestive diseases
46	Diseases: skin, bone, tissue	47	Diseases: urine, kidney,...		
48	Suicide	49	Road/other accidents	50	Accidental Poisonings
51	Other causes				

Detail $\Rightarrow$ able to separate causes with and without significant risk factors or inequality

E.g. cancers: some with strong single risk factors; some with multiple risk factors; some with no risk factors

Significant cohort effects and controllable risk factors

- (Mortality) cohort effect: variation in (sub-) population mortality by time and age has a significant element linked to cohort year of birth
- Potential reasons for a cohort effect include:
  - Variation in prevalence by cohort of controllable risk factors

    e.g. smoking, poor diet, exercise, alcohol etc.
  - Cohort-related preventative medical interventions: e.g. vaccination against Human Papilloma Virus (HPV) $\Rightarrow$ impact on cervical cancer
  - Other population-level early-life experiences

Significant cohort effects and controllable risk factors (cont.)

- Hypothesis:
  For a specific cause of death:
  - a significant cohort effect $\Rightarrow$??? one or more significant controllable risk factors
    (or a significant cohort-specific health intervention)
  - bigger cohort effect $\Rightarrow$??? bigger relative risk associated with specific risk factors
  - a significant gap between high and low-educated also $\Rightarrow$ one or more significant controllable or preventable risk factors
    even if there is no significant cohort effect
- 51 causes of death $\Rightarrow$ greater insight into individual controllable risk factors
- Cause-of-death cohort effects + controllable risk factors
  $\Rightarrow$ insight into all-cause mortality cohort effects

$$\log m(t,x) \quad = \quad \alpha(x) + \underbrace{\sum_{k=1}^{3} \beta_k(x)\kappa_k(t) + \gamma(t-x)}_{\text{CBD-M7 model}}$$

where

$$\beta_1(x) = 1, \qquad \beta_2(x) = x - \bar{x}, \qquad \beta_3(x) = (x-\bar{x})^2 - \sigma_x^2 \qquad \text{(fixed age effects)}$$

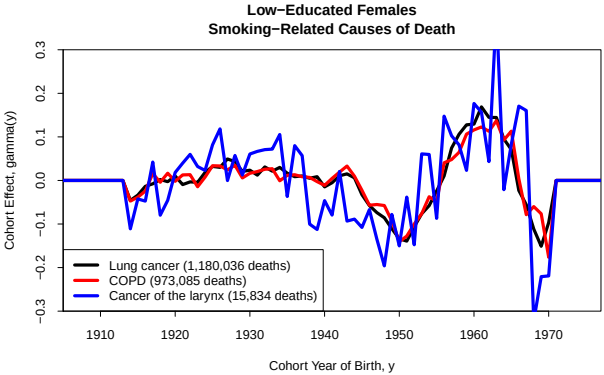
plus seven identifiability constraints.

Model each of the N_{cod} causes of death, c, individually:

$$\log m(c,t,x) \quad = \quad \alpha(c,x) + \sum_{k=1}^{3} \beta_k(x)\kappa_k(c,t) + \gamma(c,t-x)$$

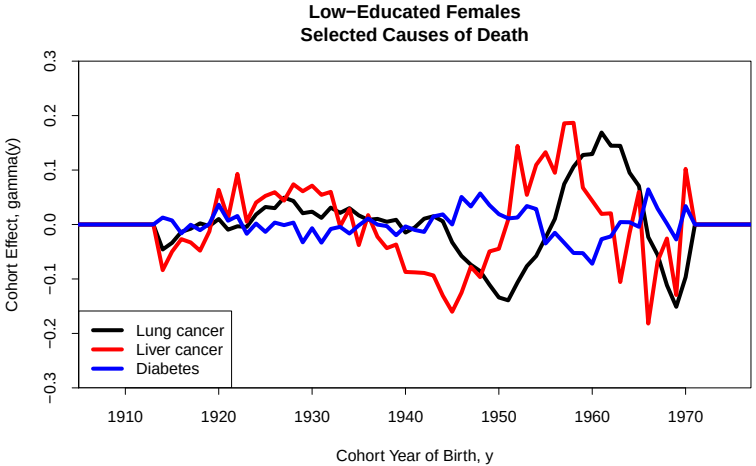
- Same model for each cause:
  - but different parameter estimates
  - general structure including non-parametric $\alpha(c,x)$ is needed to get a good fit in all years and ages and for all causes of death
- Different models customised would make it harder to identify common features
- $7N_{cod}$ identifiability constraints

CBDX-I: Selected Cohort Effects, $\gamma(c, y)$, for low-educated females



- Reported death counts in the legend represent the total for 1989-2017, ages 40-84
- Smoking is the main controllable risk factor for lung and laryngeal cancers, and Chronic Obstructive Pulmonary Disease (COPD)
- Very similar cohort effects
- But low death counts $\Rightarrow$ significant sampling variation in cancer of the larynx

CBDX-I: Selected Cohort Effects, $\gamma(c, y)$



- Other causes of death with different controllable risk factors have distinctly different cohort effects

Do we need 51 distinct cohort effects?

- Lung cancer, COPD and cancer of larynx $\Rightarrow$??? a common cohort effect
- Other causes of death with links to other controllable risks (e.g. excessive alcohol consumption) also have similar cohort effects
  and these have a different shape from a smoking cohort effect
- So perhaps we just need a small number of cohort effects that reflect variation in a small number of controllable risk factors
  - $\chi(\text{smoking}, y)$
  - $\chi(\text{alcohol}, y)$
  - $\chi(\text{diet/exercise/obesity}, y)$
  - ....
- where y = cohort year of birth
- Be aware: cohort effects for risk factors might be correlated
  e.g. a tendency by cohort to lead a generally healthy or unhealthy lifestyle

Model the N_{cod} causes jointly with common cohort effects

$$\log m(c,t,x) = \alpha(c,x) + \sum_{k=1}^{3} \beta_k(x)\kappa_k(c,t) + \underbrace{\sum_{j=1}^{n} \delta_j(c)\chi_j(t-x)}_{\gamma(c,t-x)}$$

where $\chi_1(y), \ldots, \chi_n(y)$ are n common cohort effects that apply to each cause of death

and the $\delta_j(c)$ control the contribution of each common cohort effect, $\chi_j(y)$, to the cause-specific cohort effect $\gamma(c,y)$

$$\log m(c,t,x) \quad = \quad \alpha(c,x) + \sum_{k=1}^{3} \beta_k(x)\kappa_k(c,t) + \sum_{j=1}^{n} \delta_j(c)\chi_j(t-x)$$

- Motivation: the n common cohort effects can be linked to n significant, underlying controllable risk factors
  e.g. smoking
- Each cause of death, c, has scaling factors $\delta_1(c), \ldots, \delta_n(c)$ attached to the common cohort effects
- Hypothesis
  Example: $\chi_1(y)$ links to cohort smoking prevalence & intensity
  Then, the size of $\delta_1(c)$ links to the *relative risk* of smoking for cause of death c.
  - If smoking is not a risk factor for CoD c then $\delta_1(c) = 0$
  - Bigger the relative risk $\Rightarrow$ bigger $\delta_1(c)$

Identifiability constraints

Similar identifiability constraints to the CBDX-I model, but fewer

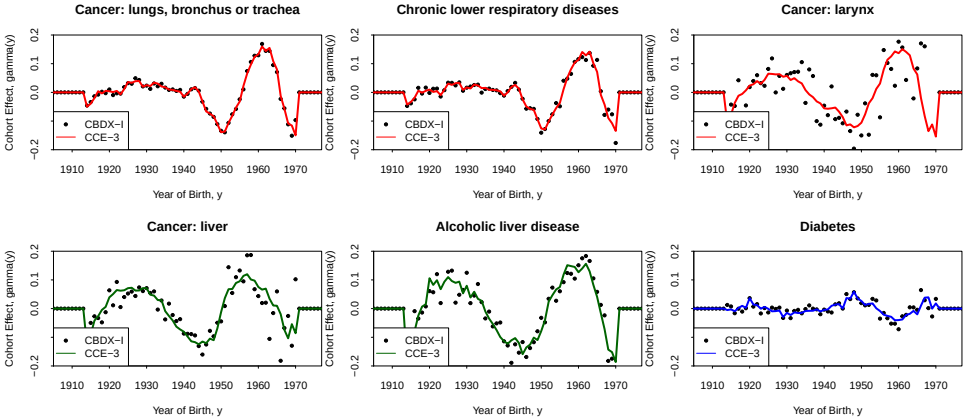
- $4n$ constraints on the n common cohort effects, $\chi_1(y), \ldots, \chi_n(y)$
- $3N_{cod}$ constraints on the cause-of-death specific period effects, $\kappa_k(c, t)$

Results: $n = 3$ common cohort effects

	Model:	CBDX-I			CCE-3		
Group	# obs, N_{obs}	maximum log-lik, $\hat{l}$	effective # params, ν	BIC	maximum log-lik, $\hat{l}$	effective # params, ν	**increase in BIC**
Males-Low	**55440**	-205252	8910	-253914	-206549	6003	+14580
Males-High	55440	-185460	8910	-234123	-186782	6003	+14555
Females-Low	**59136**	-209396	9504	-261609	-210775	6390	+15729
Females-High	59136	-185151	9504	-237365	-186435	6390	+15824

- CCE-3 Model fitted to each of the four sub-populations independently
- Bayes Information Criterion: $BIC = \log \hat{L} - 0.5\nu \log N_{obs}$ (so aim to maximise BIC)
- All four populations: BIC $\Rightarrow$ CCE-3 is the best model

Examples: Low educated females; CDBX-I versus CCE-3



- Good correspondence between CCE-3 and CBDX-I
- But with reduced noise in the estimated cohort effects
- Distinctly different cohort effects are now clearer

Discussion I

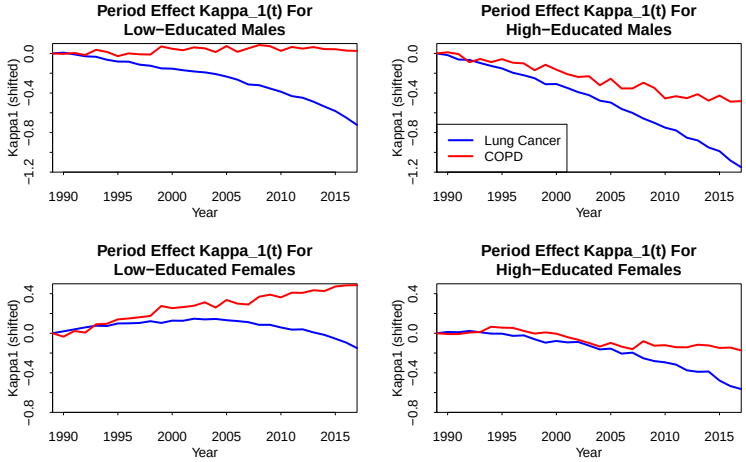
- The CCE-n model fitted to cause of death data avoids a significant amount of overfitting and leads to smoother and more robust estimates of the cohort effects
- The n fitted cohort effects give some insight into underlying risk drivers
- The n cohort effects can feed into all-cause mortality modelling

- Caution: e.g. smoking as a controllable risk factor
- $\gamma(c, y)$ is not the same as the underlying controllable risk factor
  - Impact depends on the prevalence and intensity of the controllable risk factor
    (e.g. 50% smoking prevalence: 40/10 heavy/light smokers different from 10/40)
  - Identifiability $\Rightarrow$ risk-factor$(y) \sim \gamma(y) +$ cubic(y)
  - Magnitude of $\gamma(c, y)$ depends on the relative risk linking the risk factor to CoD death rate

Discussion II

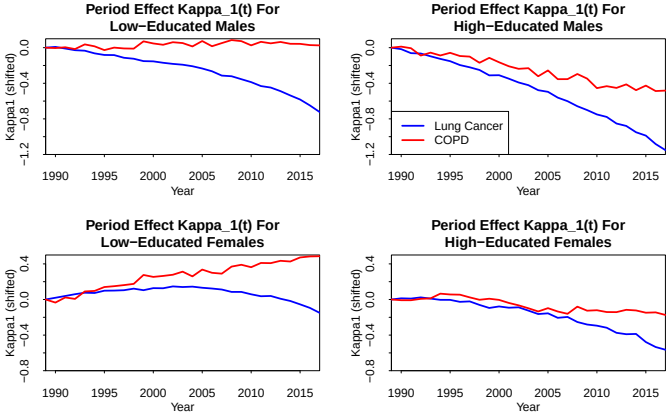
- If treatment of disease/illness is consistent between sub-populations
  then modelled period and cohort effects $\Rightarrow$ give insight into relative changes in the underlying behaviour
- The approach might also allow us to decompose cause-specific improvements by controllable risk factors *versus* medical and other advances
- We can also use the model to compare improvements in one CoD versus another one:
  e.g. lung cancer versus COPD

Comparison of the key period effect, $\kappa_1(t)$, for lung cancer and COPD



- Lung versus COPD gap: Consistent picture by male/female and low/high-educated
- Potential inference: Lung cancer treatment has been improving faster than COPD 1.5% to 2.5% per annum faster since 1989

Comparison of the key period effect, $\kappa_1(t)$, for lung cancer and COPD



- E.g. 4× COPD: Different trajectories also inform us about the relative changes in the prevalence/intensity of the underlying risk factor
- E.g. Low-educated females +40% versus High-educated males −40%

Discussion III

- Here: three common cohort effects.
  But is that the right number?
- Interpreting the common cohort effects, $\chi_i(y)$, is potentially challenging
  e.g. prevalence of controllable risk factors by year of birth might be correlated
- Common cohort effects can help explain cohort effects estimated at the all-cause level

Discussion IV

- Methodology can be adapted to HMD cause of death data
  - **Need to adapt to 5-year age groups**
    Problem: 5-year age groups $\Rightarrow$ we only observe blocks of 5 cohorts
    - 1989 + ages 70-74 $\Rightarrow$ 1915—1919 cohorts
    - 1990 + ages 70-74 $\Rightarrow$ 1916—1920 cohorts
    - 1991 + ages 70-74 $\Rightarrow$ 1917—1921 cohorts ...

    Needs some reprogramming of 1×1 APC models. Either:
    - work with single ages but likelihood function aggregates into 5×1;
    - e.g. 1989 + ages 70-74 $\Rightarrow$ cohort effect$= \gamma(1917)$
    - or e.g. 1989 + ages 70-74 $\Rightarrow$ cohort effect $\frac{1}{5}\sum_{y=1915}^{1919}\gamma(y)$

    Robustness experiment: group our US data into 5-year age bands and compare with single-age results
  - **How to handle smaller populations?** Potentially merge some smaller causes of death into coherent groups by controllable risk factors
    Robustness experiment: group US causes of death in the same way and compare results with 51-CoD results

Summary

- Work in progress
- We propose the Common Cohort Effect model as a way to link cohort effects for different causes of death to underlying controllable risk factors
- US data: Three common cohort effects were found to be very effective
- Potential to add insight into the contribution of different causes to all-cause mortality improvements
- Potential to provide insight into the effect of specific controllable risk factors at the all-cause level

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

HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries