The Relationship between Socio-Economic Circumstances and Causal Mortality

Daniel Alai   Séverine Arnold (-Gaille)   Madhavi Bajekal
Andrés Villegas

Centre of Excellence in Population Ageing Research (CEPAR)

School of Mathematics, Statistics & Actuarial Science, University of Kent
Faculty of Business and Economics, University of Lausanne
Department of Applied Health Research, University College London
Cass Business School, City University London

Actuarial Teachers and Researchers Conference
1–2 December 2014

Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions

Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions

Introduction

- The motivation is to gain an improved understanding of mortality.
  ↳ Information is lost in aggregate mortality data . . .
  . . . Potentially found in causal mortality data!
- Reliable data is not readily available.
  ↳ Office for National Statistics data with socio-economic variables!
- What if circulatory-related deaths are considerably reduced?
  ↳ This scenario cannot be tested using aggregate models . . .
  . . . But it can be tested using causal models!
  ↳ Which socio-economic group stands to benefit most?
- Be careful! Causes are intrinsically dependent!
  ↳ Instantaneous probabilities vs. annual probabilities.
- Aim: quantify the impact on {residual} life expectancy.
  ↳ Study effects of scenarios on socio-economic gaps.

⇝ {This is a work-in-progress; feedback is most welcome!}

Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions

University of Kent

Data

- The UK Office for National Statistics.
  - ↳ Data, by gender, for 1981–2007.
  - ↳ Five-year age groups, from 25–29, . . . , 80-84, and 85+.
  - ↳ Socio-economic circumstances in quintiles.
- Deaths categorized by the International Classification of Diseases.
  - ↳ When classifications change, comparability ratios are applied.
  - ↳ This is to maintain some consistency under classification shifts.
- Adjusted death rates are produced and analyzed.
  - ↳ Relevant exposure adjustments are also made.

Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions

Multinomial Logistic Model

- Let $D_i(x,t)$ denote random deaths from cause i for age x at time t.
- Let $L(x,t)$ denote the subsequent survivors.
- Consider n causes.

$$Y(x,t) = (D_1(x,t), D_2(x,t), \ldots, D_n(x,t), L(x,t))'.$$

Assume $Y(x,t) \sim$ multinomial distribution $\{\pi(x,t),\ E(x,t)\}$, with

$$\pi(x,t) = \{q_1(x,t), q_2(x,t), \ldots, q_n(x,t), p(x,t)\}',$$

where,

$$\sum_{k=1}^{n} q_k(x,t) + p(x,t) = 1,$$

and

$$E(x,t) = L(x,t) + \sum_{k=1}^{n} D_k(x,t).$$

$\rightsquigarrow$ Annual probabilities and initial exposure.

Multinomial **Logistic** Model

- Adopt survival as the baseline category in the logistic framework.

$$\log \frac{q_i(x,t)}{p(x,t)} = X(x,t)\beta_i, \quad \text{for } i \in \{1,\ldots,n\}.$$

- $X(x,t)$ is the design matrix, and
- β_i the regression parameters suited to cause i.

The probabilities are given as follows:

$$\begin{aligned} q_i(x,t) &= \frac{\exp\{X(x,t)\beta_i\}}{1+\sum_k \exp\{X(x,t)\beta_k\}}, \quad \text{for } i \in \{1,\ldots,n\}, \\ p(x,t) &= \frac{1}{1+\sum_k \exp\{X(x,t)\beta_k\}}. \end{aligned}$$

University of Kent

Multinomial Logistic **Model**

Models typically include some combination of age, period, and cohort.

⇝ Consider a gender specific model with main and interaction effects:

- Age is given by age-groups, which we treat categorically.
  - ↳ 13 groups: 25-29, 30-34, … 80-84, 85+.
- Period is treated continuously.
  - ↳ Continuous time avoids time-series consideration when forecasting.
- Cohort is excluded as a *main* effect:
  - ↳ We have a limited number of periods.
  - ↳ Causal mortality is more intuitively linked to period effects.
- Socio-economic circumstance quintiles are treated categorically.

- Age–period interaction is included!
  - ↳ "Lee-Carter" observation: age-groups have different time trends.
- Include socio-economic–age and –period interactions!

University of Kent

The Regression Formula

$$
\begin{aligned}
\eta_i(g,x,s,t) \quad = \quad & \beta_{0,i} + \beta_{1,g,i} + \beta_{2,x,i} + \beta_{3,s,i} + \beta_{4,i}t \\
& + \beta_{5,g,x,i} + \beta_{6,g,s,i} + \beta_{7,g,i}t \\
& + \beta_{8,x,s,i} + \beta_{9,x,i}t + \beta_{10,s,i}t \\
& + \beta_{11,g,x,s,i} + \beta_{12,g,x,i}t + \beta_{13,g,s,i}t.
\end{aligned}
$$

where,

$$\eta_i(g,x,s,t) = \ln \frac{q_i(g,x,s,t)}{p(g,x,s,t)}.$$

Highlighted terms are gender-specific.

$\leadsto$ What remains is an intercept, three main and interaction effects.

Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions

Cause Elimination

- Consider the elimination of cause j.
- The probabilities in our model are adjusted as follows:

$$\begin{aligned} q_j(x,t) &= 0, \\ q_i(x,t) &= \frac{\exp\{X(x,t)\beta_i\}}{1+\sum_{k\neq j}\exp\{X(x,t)\beta_k\}}, \quad i \neq j \\ p(x,t) &= \frac{1}{1+\sum_{k\neq j}\exp\{X(x,t)\beta_k\}}. \end{aligned}$$

- Representative of a proportional re-weighting of the probabilities.

⇝ This essentially ignores extrinsic dependence amongst the causes.
↳ {We are considering "multi-cause data" to address this!}

Shocking Causal Mortality

- In general, suppose we introduce a shock $\rho_i \geq 0$ to cause i.
  - Values of $\rho_i > 1$ signify a marginal increase in mortality.
  - The value $\rho_i = 0$ corresponds to cause elimination.
- The probabilities are adjusted as follows:

$$q_i(x,t) = \frac{\rho_i \exp\{X(x,t)\beta_i\}}{1 + \sum_k \rho_k \exp\{X(x,t)\beta_k\}},$$

$$p(x,t) = \frac{1}{1 + \sum_k \rho_k \exp\{X(x,t)\beta_k\}}.$$

- Based solely on $\rho_i > 0$, will $q_i(x,t)$ increase or decrease?
- Previous work has considered shocks on an instantaneous basis.
  - The annual approach re-distributes less probability to survival.

    ↳ It is more conservative in a mortality-sense.

University of Kent

Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- **Preliminary Results**
- Conclusions

University of Kent

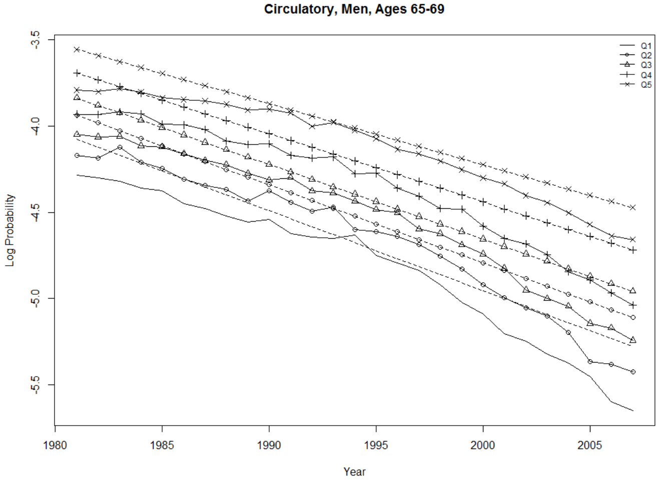
Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
gender	6	34	< 0.0001
age	72	22012	< 0.0001
sec	24	3407	< 0.0001
year	6	486	< 0.0001
gender*age	72	5016	< 0.0001
gender*sec	24	556	< 0.0001
gender*year	6	36	< 0.0001
age*sec	288	70431	< 0.0001
age*year	72	23398	< 0.0001
sec*year	24	3668	< 0.0001
gender*age*sec	288	2818	< 0.0001
gender*age*year	72	4937	< 0.0001
gender*sec*year	24	557	< 0.0001

Fitting the data results in high significance for all effects!

↳ {sec = socio-economic circumstances}

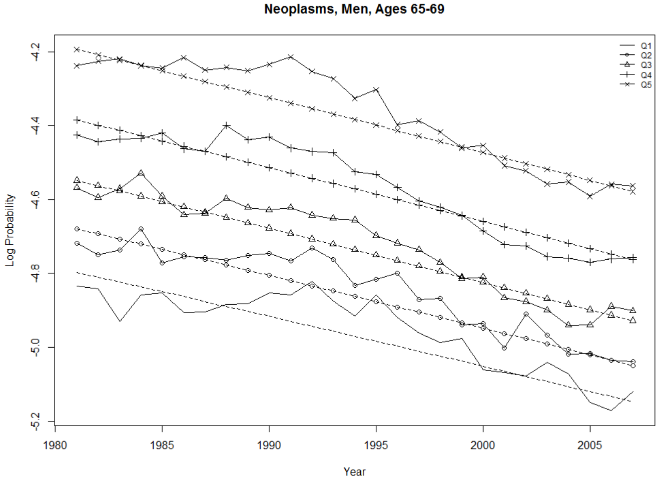
Observed and Fitted {log} Mortality



⇝ Overestimating mortality for this {and older} age-group!

↳ {Might have to consider quadratic time}

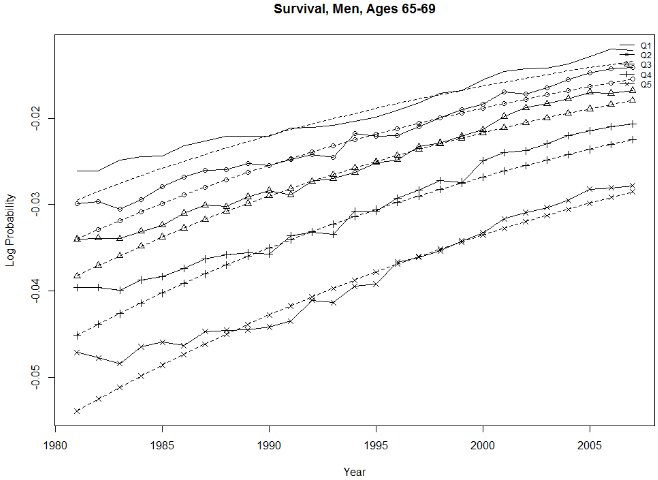
Observed and Fitted {log} Mortality



⇝ A pretty good fit, especially in the final year!
↳ {Linear time appears to suffice for this cause}

University of Kent

Observed and Fitted {log} Survival

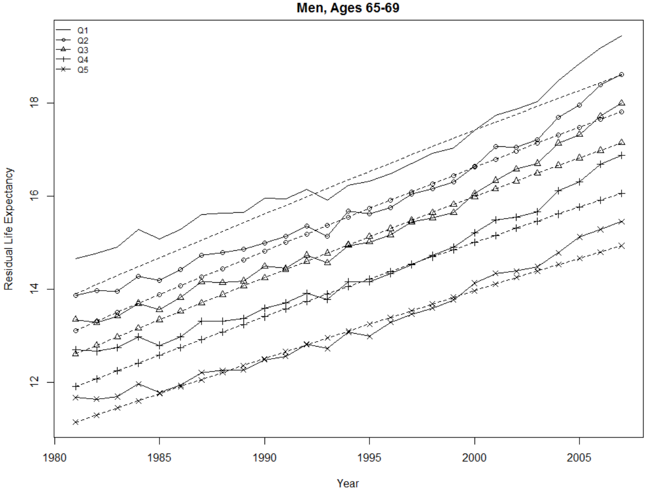


⇝ Evidence of compression; especially at higher quintiles!

↳ {Again, notice underestimation of survival}

University of Kent

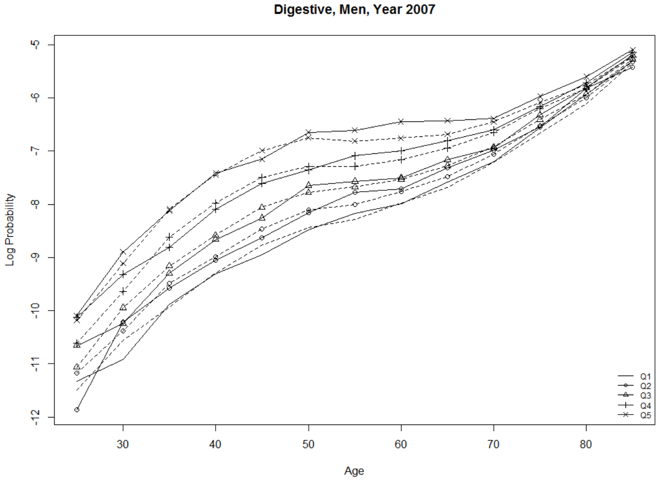
Observed and Fitted {Residual} Life Expectancy



⇝ Expansion in life expectancy?! {an aggregate measure!}

↳ {It captures the mortality expansion in later age-groups}

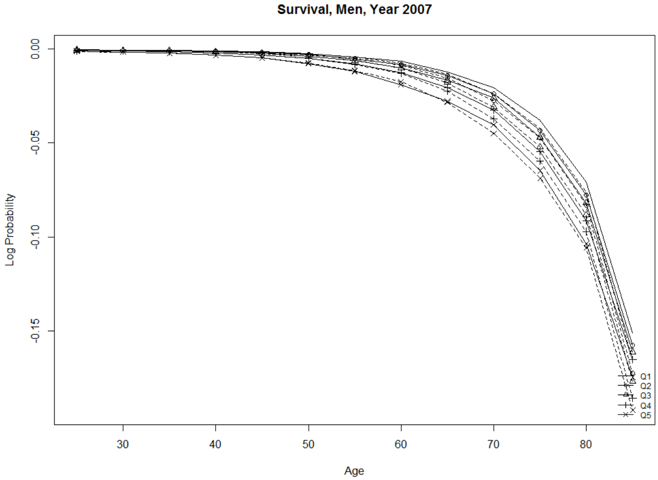
Observed and Fitted {log} Mortality



⇝ It looks like the role of 'sec' diminishes with age!

↳ {A difficult picture to digest, age-effect is dominating!}

Observed and Fitted {log} Survival



⇝ Survival plots tell a different tale. {Imperfect fit notwg!}
  ↳ {Of course, age-effect is, again, dominating.}

University of Kent

Where we aim to go from here ...

What happens when a cause is shocked {eliminated}!
↳ What happens to life expectancy?

Again, consider 65 year-old male residual life expectancy:

Life Expectancy	Q1	Q2	Q3	Q4	Q5
Fitted	18.59	17.81	17.14	16.06	14.93
Fitted {– circulatory}	24.04	23.21	22.48	21.23	19.95
Gain	5.46	5.40	5.34	5.18	5.02

⇒ The most affluent benefit most from a 'positive' shock to circulatory.

Given an ability to shock causes, what criteria should be optimized?

- Reduce the socio-economic gap?
- Provide the biggest life expectancy gains for the population?
- Etc.

University of Kent

Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions

Conclusions

- The motivation is to gain an improved understanding of mortality.
  ↳ Information is lost in aggregate mortality data …
  … Potentially found in causal mortality data!
- Reliable data is not readily available.
  ↳ Office for National Statistics data with socio-economic variables!
- What if circulatory-related deaths are considerably reduced?
  ↳ This scenario cannot be tested using aggregate models …
  … But it can be tested using causal models!
  ↳ Which socio-economic group stands to benefit most?
- Be careful! Causes are intrinsically dependent!
  ↳ Instantaneous probabilities vs. annual probabilities.
- Aim: Quantify the impact on {residual} life expectancy.
  ↳ Study effects of scenarios on socio-economic gaps.

⇝ {This is a work-in-progress; feedback is most welcome!}

Thank you!