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
The UK Office for National Statistics.
   - 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 $Y(x, t) \sim \text{multinomial}$ 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

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

 quindi 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
- \(\beta_i\) the regression parameters suited to cause \(i\).

The probabilities are given as follows:

\[
q_i(x, t) = \frac{\exp\{X(x, t)\beta_i\}}{1 + \sum_k \exp\{X(x, t)\beta_k\}}, \quad \text{for } i \in \{1, \ldots, n\},
\]
\[
p(x, t) = \frac{1}{1 + \sum_k \exp\{X(x, t)\beta_k\}}.
\]
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!
The Regression Formula

\[ \eta_i(g, x, s, t) = \beta_{0,i} + \beta_{1,g,i} + \beta_{2,x,i} + \beta_{3,s,i} + \beta_{4,it} + \beta_{5,g,x,i} + \beta_{6,g,s,i} + \beta_{7,g,it} + \beta_{8,x,s,i} + \beta_{9,x,it} + \beta_{10,s,it} + \beta_{11,g,x,s,i} + \beta_{12,g,x,it} + \beta_{13,g,s,it}. \]

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.

\( \Rightarrow \) What remains is an intercept, three main and interaction effects.
Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions
Consider the elimination of cause \( j \).

The probabilities in our model are adjusted as follows:

\[
q_j(x, t) = 0,
\]

\[
q_i(x, t) = \frac{\exp\{X(x, t)\beta_i\}}{1 + \sum_{k \neq j} \exp\{X(x, t)\beta_k\}}, \quad i \neq j
\]

\[
p(x, t) = \frac{1}{1 + \sum_{k \neq j} \exp\{X(x, t)\beta_k\}}.
\]

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!}
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.
Plan

- Introduction
- Data
- Methodology
- Shocking Causal Mortality
- Preliminary Results
- Conclusions
## Analysis of Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>6</td>
<td>34</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>age</td>
<td>72</td>
<td>22012</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>sec</td>
<td>24</td>
<td>3407</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>year</td>
<td>6</td>
<td>486</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>gender*age</td>
<td>72</td>
<td>5016</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>gender*sec</td>
<td>24</td>
<td>556</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>gender*year</td>
<td>6</td>
<td>36</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>age*sec</td>
<td>288</td>
<td>70431</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>age*year</td>
<td>72</td>
<td>23398</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>sec*year</td>
<td>24</td>
<td>3668</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>gender<em>age</em>sec</td>
<td>288</td>
<td>2818</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>gender<em>age</em>year</td>
<td>72</td>
<td>4937</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>gender<em>sec</em>year</td>
<td>24</td>
<td>557</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Fitting the data results in high significance for all effects!

\[ \text{sec} = \text{socio-economic circumstances} \]
Overestimating mortality for this {and older} age-group!

\{ Might have to consider quadratic time \}
A pretty good fit, especially in the final year!

\[\text{Linear time appears to suffice for this cause}\]
Evidence of compression; especially at higher quintiles!

\{Again, notice underestimation of survival\}
Expansion in life expectancy?! \{an aggregate measure!\}

\{It captures the mortality expansion in later age-groups\}
It looks like the role of ‘sec’ diminishes with age!
\{A difficult picture to digest, age-effect is dominating!\}
Survival plots tell a different tale. {Imperfect fit not wg!}

{Of course, age-effect is, again, dominating.}
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:

<table>
<thead>
<tr>
<th>Life Expectancy</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitted</td>
<td>18.59</td>
<td>17.81</td>
<td>17.14</td>
<td>16.06</td>
<td>14.93</td>
</tr>
<tr>
<td>Fitted {- circulatory}</td>
<td>24.04</td>
<td>23.21</td>
<td>22.48</td>
<td>21.23</td>
<td>19.95</td>
</tr>
<tr>
<td>Gain</td>
<td>5.46</td>
<td>5.40</td>
<td>5.34</td>
<td>5.18</td>
<td>5.02</td>
</tr>
</tbody>
</table>

⇒ 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.
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!