Death rates by educational attainment level in the US

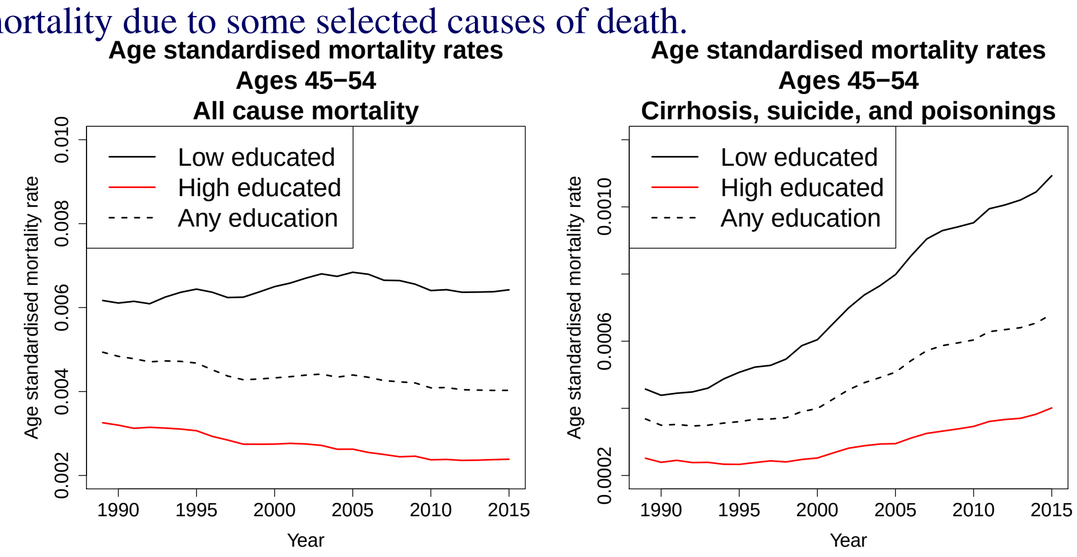
Cristian Redondo Loures, Andrew J. G. Cairns

Actuarial Research Centre of the Institute and Faculty of Actuaries, Heriot-Watt University, UK.

Email: c.redondo_loures@hw.ac.uk / a.j.g.cairns@hw.ac.uk

Introduction

Case and Deaton (2015) found a rise in mortality amongst middle aged non-hispanic whites in the US, concentrated in those with a high school diploma or less formal education. This increase was related to the upward trend in mortality due to some selected causes of death.



Can we go further?

- Individual years of age.
- Extension to higher ages.
- Analysis of many other causes of death.

Our data sources

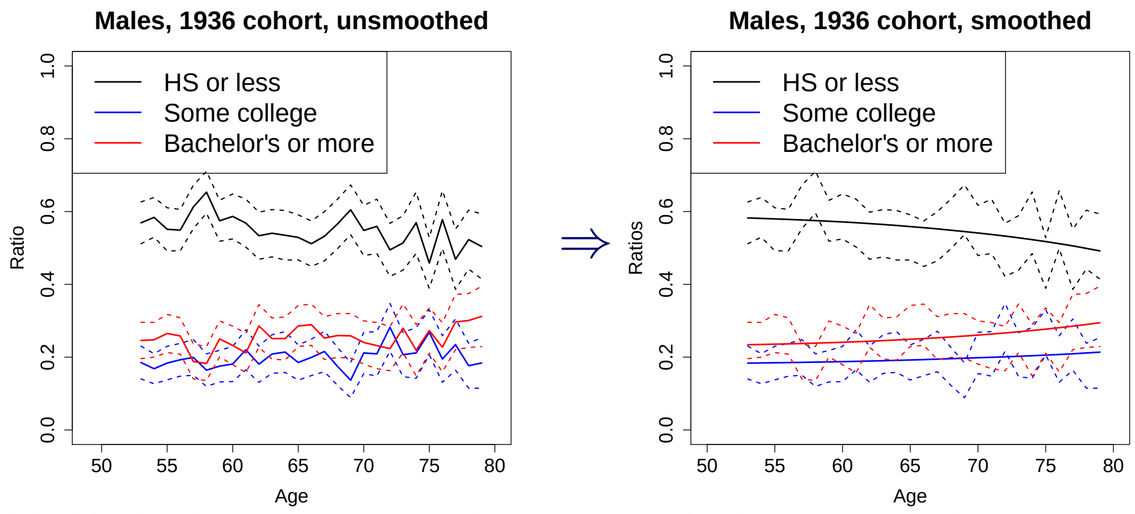
Source of death counts: CDC data. Complete database of individual death records containing all relevant information for years 1989-2015.

Sources of exposures:

- HMD: provides reliable population estimates. No information on educational attainment.
- ACS/Census: provide reliable population estimates AND information on educational attainment. Available only for selected years in the period analysed.
- CPS: provides information on educational attainment for the whole period under analysis. Big sampling error → noise in the education-dependent exposures. Restricted age range (only up to 79 by individual year of age).

Problem: unreliable population data

Combining HMD estimates for the total population and CPS data for the ratio of educated people we obtain exposures by educational attainment. These are very noise and need smoothing.



Reliable death counts → use information on deaths to smooth the noise in the exposures.

Recurrence for the ratios within a cohort:

$$R_{i+1}^{(e,c)} = \frac{E_{i+1}^{(e,c)}}{E_{i+1}^{(c)}} = \frac{E_i^{(e,c)} - \Delta_i^{(e,c)}}{E_i^{(c)} - \Delta_i^{(c)}} = \frac{R_i^{(e,c)} E_i^{(c)} - \Delta_i^{(e,c)}}{E_i^{(c)} - \Delta_i^{(c)}}, \tag{1}$$

$e \rightarrow$ Educational attainment.

$c \rightarrow$ Characteristics of the cohort.

$\Delta \rightarrow$ Decrease in membership (approximated by the number of deaths).

The smoothed ratios will minimise

$$\mathcal{O}_N = \sum_i \left[\left(R_i^{(e,c)} - \hat{R}_i^{(e,c)} \right)^2 + (C_i^{(e,c)})^2 \right] \tag{2}$$

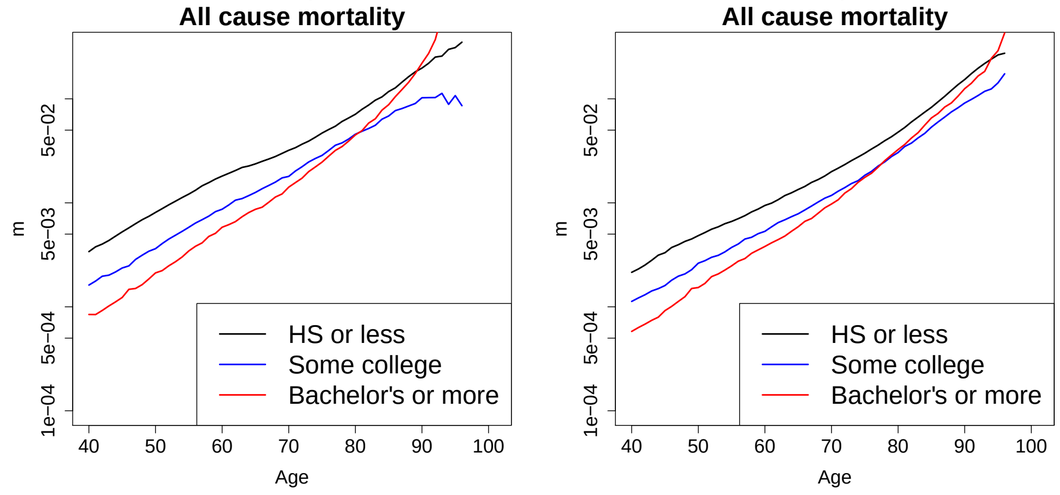
$$C(x,t) = \log(m(x,t)) - \frac{1}{2}\left(\log(m(x+1,t)) + \log(m(x-1,t))\right)$$

Eqs. (1) and (2) provide a systematic way of estimating (smooth) ratios of educated people, and extrapolating them up to any age desired.

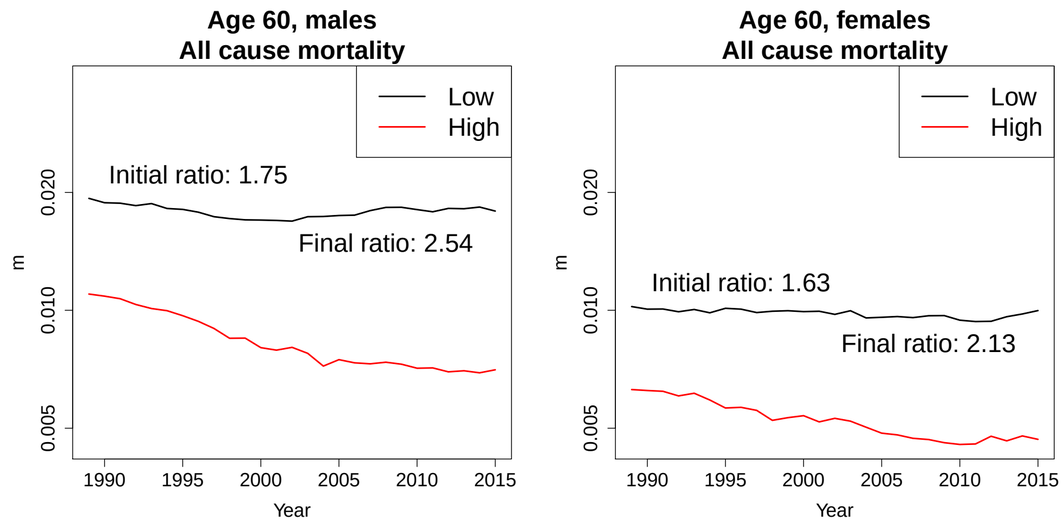
Now we have all the tools needed to calculate mortality rates!

Results

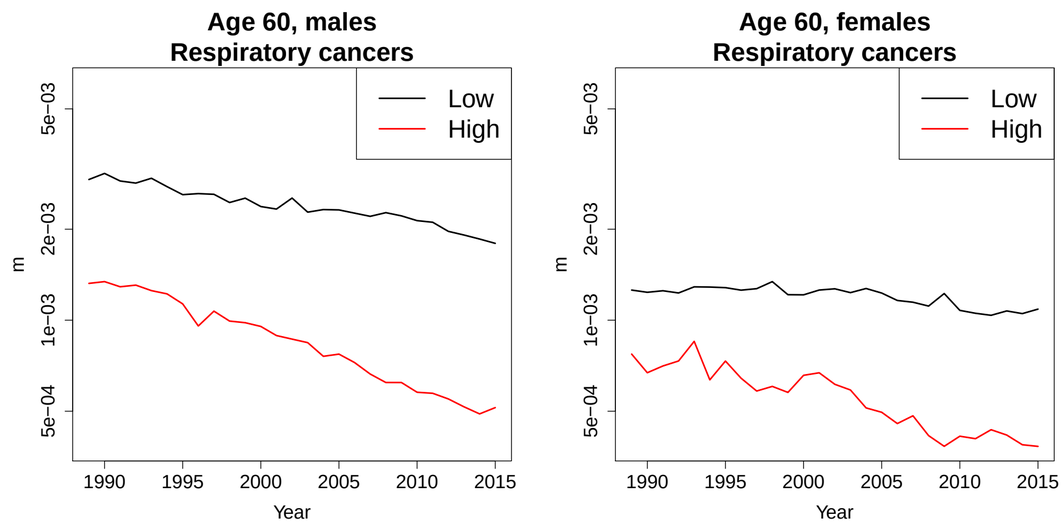
One last issue: The "Some college" and "Bachelor's or more" education groups show inconsistent results at very high ages (extrapolated area). Comparison with ACS/Census shows that the problem does not lie with the data and methods underpinning the in-sample exposure estimates → Likely bias is in the CDC data! (Overreporting of college graduates).



We need to merge the "Some college" and "Bachelor's or more" categories into a broad "Attended University" category. We relabel the two remaining categories as "Low" and "High" education.



Gap in all cause mortality has increased for both males and females in the period 1989-2015.



For some lifestyle-related causes of death (like cancer in the respiratory system, heavily linked to tobacco) the gap is growing very rapidly.

Conclusions

- We managed to obtain smooth death rates by education level for individual years of age despite the noise in the educational attainment data.
- We managed to study several causes of death and the evolution of the difference in mortality rates due to differences in educational attainment.

Further research

- Combining our findings with data on lifestyle factors (like smoking) could further our understanding about the effects of these habits on the mortality rates of different subpopulations.
- A natural next step is estimating the impact of these lifestyle factors on mortality (i.e., if lower educated people had copied the smoking habits of their higher educated counterparts, how might their mortality have evolved over the period considered in our analysis?)

Multi-population Mortality Modelling Quebec Pension Plan (QPP)

Jie Wen; Prof. Andrew Cairns; Dr. Torsten Kleinow

School of Mathematical and Computer Sciences, Heriot-Watt University

Actuarial Research Centre
Institute and Faculty of Actuaries

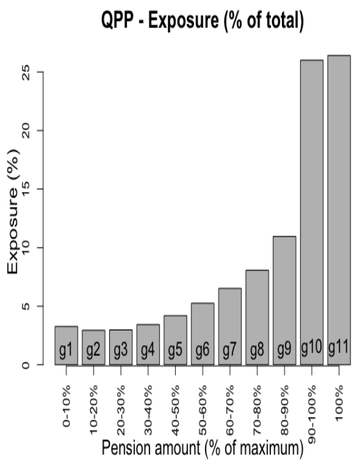
HERIOT WATT UNIVERSITY

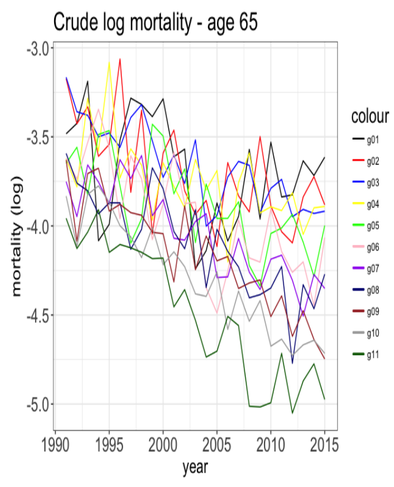
ABSTRACT

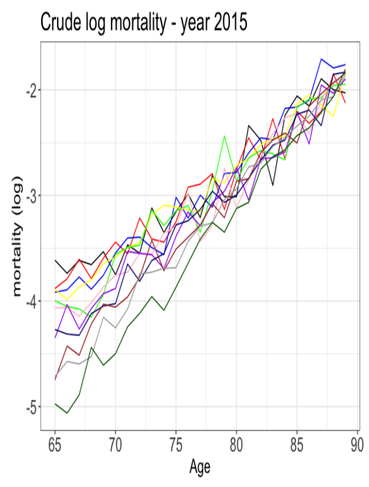
Modelling of mortality rates for specific population(s) is key for longevity risk management. Reliable multi-population stochastic models help capture different mortality patterns and project future longevity. We have carried out analysis into several multi-population datasets including English IMD, Canada Pension Plan (CPP) and QPP. We present modelling results of several mortality models from dataset provided by QPP. There are 15 models fitted to QPP data, and we present 4 of them.

1. QPP DATA OVERVIEW

- QPP data consists of 11 sub-population groups from group 1 to 11, sorted by increasing pension amount in 10% bands.
- The underlying data range for model fitting is age 65-89, and year 1991-2015.
- The 11 groups are not equally sized, groups with high-pension population tend to be in larger size in terms of exposures.







2. MODEL SPECIFICATION

There are 4 stochastic models' fitting and analysis presented today.

- m1 $\log m_{xti} = \alpha_{xi} + \kappa^1_{ti} + (x - \bar{x})\kappa^2_{ti}$ (Plat (2009))
- m2 $\log m_{xti} = \alpha_{xi} + \beta^1_x \kappa^1_{ti} + \beta^2_x \kappa^2_{ti}$ (CAE model by Kleinow (2014))
- m6 $\log m_{xti} = \alpha_x + \kappa^1_{ti} + (x - \bar{x})\kappa^2_{ti}$ (Plat model with common α_x)
- m14 $\log m_{xti} = \alpha_x + \beta^1_x \kappa^1_{ti} + \beta^2_x \kappa^2_{ti}$ (CAE model with common α_x)

3. NOTATION & MATHEMATICAL PRINCIPLE

D_{xti}, E_{xti}: number of deaths/exposures observed at age x in year t for group i, from underlying QPP data. m_{xti}, $\hat{m}_{xti}$: crude/estimated death rate at certain age, year and group.

The log-likelihood function is created from the Poisson assumption for the number of deaths, D_{xti}.

$$D_{xti} \sim Poi(E_{xti}\hat{m}_{xti})$$

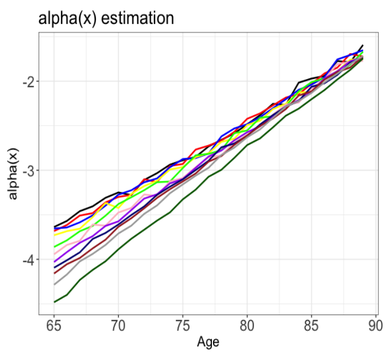
where m_{xti} is the underlying death rate observed. For example, for m1 the distribution of death is:

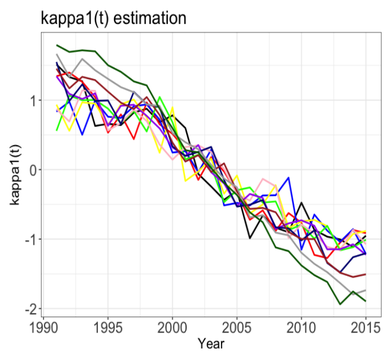
$$D_{xti} \sim Poi\Big(E_{xti}\exp(\alpha_{xi} + \kappa^1_{ti} + (x - \bar{x})\kappa^2_{ti})\Big)$$

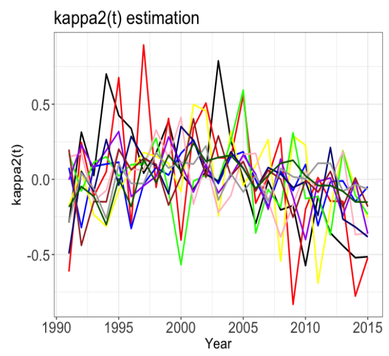
We use maximum log-likelihood (MLE) to estimate all parameters.

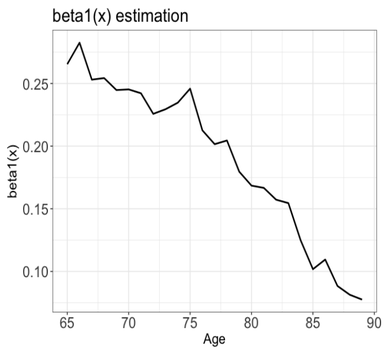
4. ESTIMATION RESULTS - M2 and M6

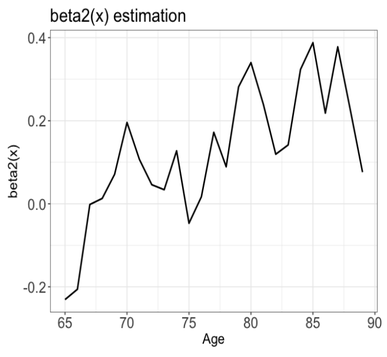
Model m2:







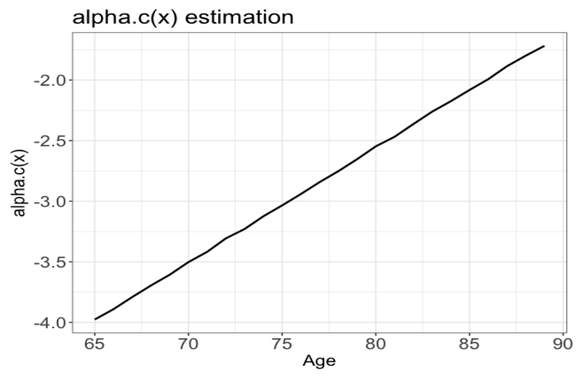


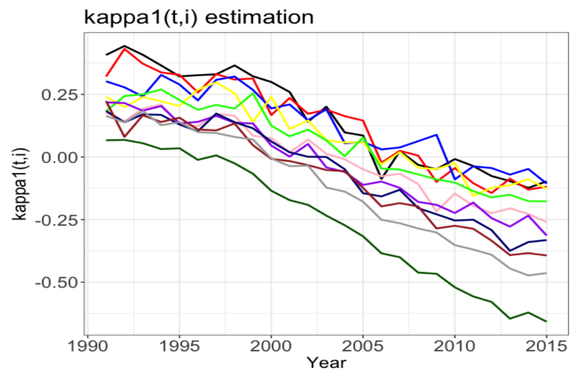


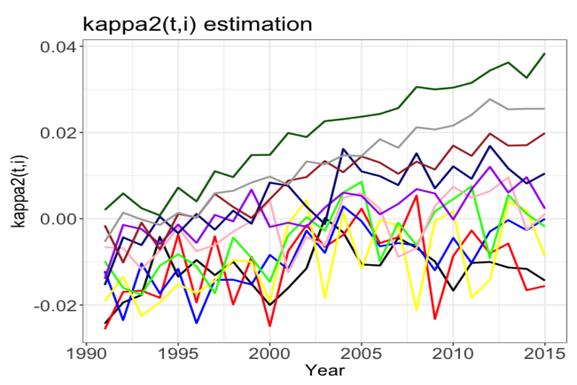
- Group specific α_{xi} gives clear group rankings.
- κ^1_{ti} are distributed very close for subgroups with full of crossover.

4. ESTIMATION RESULTS - M2 and M6 - cont.

Model m6:







- As α_x is common, variations between subgroups are captured by κ^1_{ti} and κ^2_{ti}.
- Gap between subgroups in κ^1_{ti} has widened.
- Group 11 stands well below all other groups.

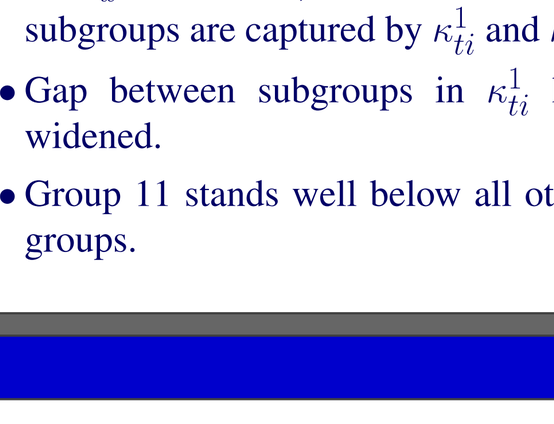
5. MODEL SELECTION

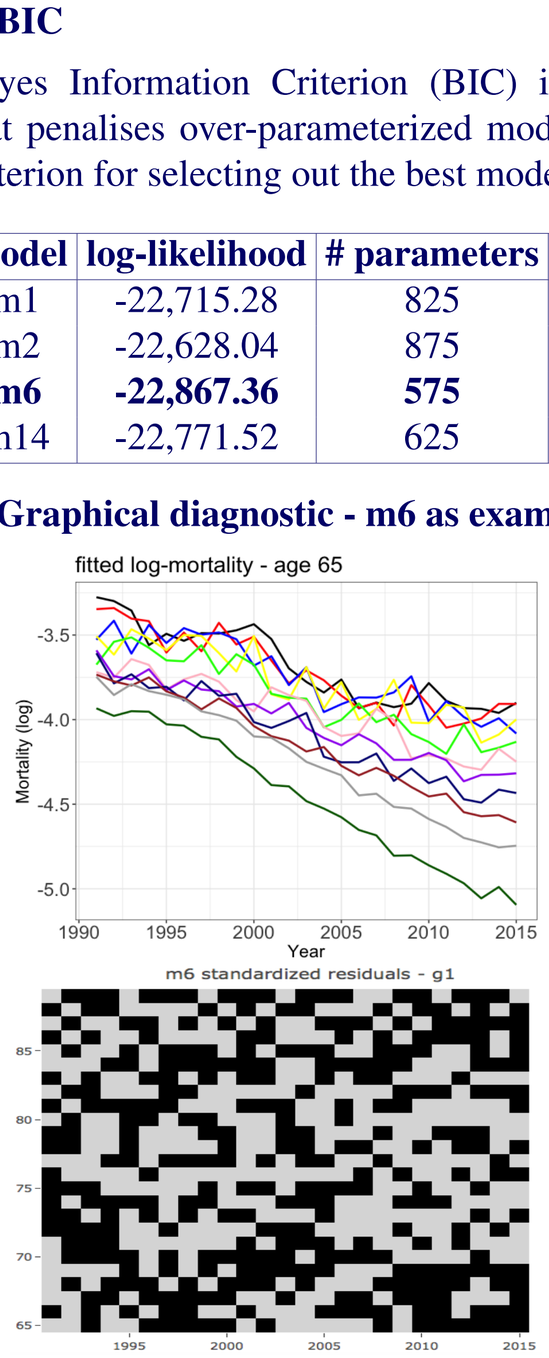
1. BIC

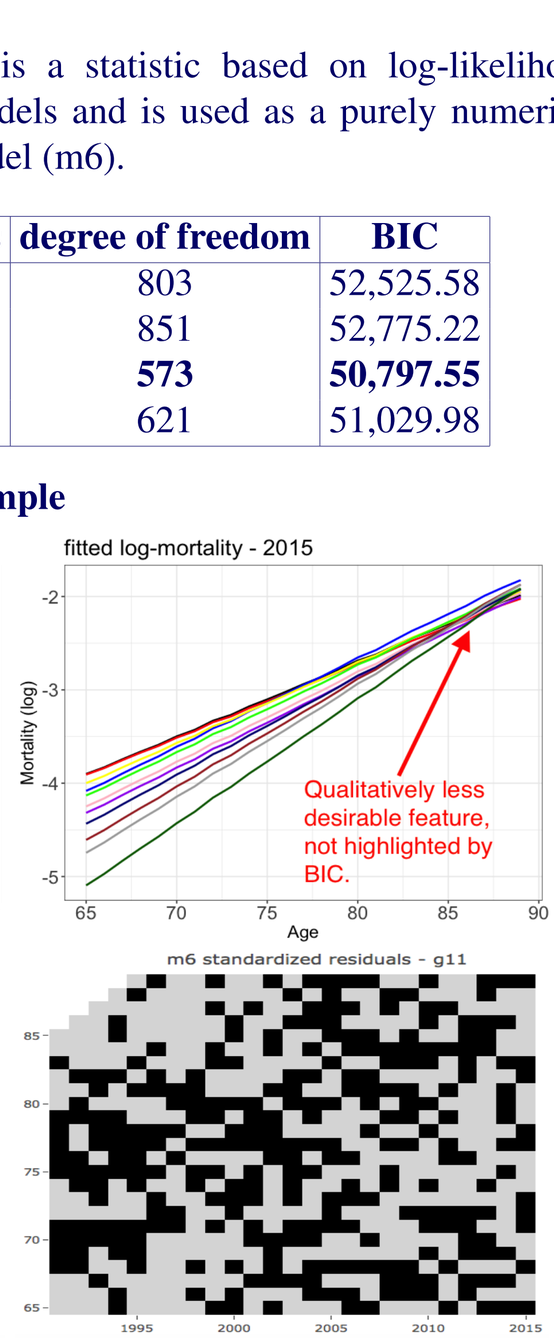
Bayes Information Criterion (BIC) is a statistic based on log-likelihood that penalises over-parameterized models and is used as a purely numerical criterion for selecting out the best model (m6).

Model	log-likelihood	# parameters	degree of freedom	BIC
m1	-22,715.28	825	803	52,525.58
m2	-22,628.04	875	851	52,775.22
m6	**-22,867.36**	**575**	**573**	**50,797.55**
m14	-22,771.52	625	621	51,029.98

2. Graphical diagnostic - m6 as example







* From left to right: Heatmaps of standardized (Pearson) residuals from m6, for group 1 and 11. (Black cell: positive figure; Grey cell: negative figure)
** White cell: zero exposures from QPP data.

CONCLUSION AND FUTURE WORKS

- For multi-population demongraphic dataset with quite high volatility across different groups, simpler models with linear terms tend to produce better BIC result, i.e. model m6 and m14. For the specific QPP dataset, common α_x performs better.
- Besides numerical criteria, we also assess model fitting results by graphical diagnostics, includes Pearson residuals distribution, comparison between fitted and crude mortalities, etc.
- Based on the best-fitting model we can project future death rates using a multivariate time series model for the κ's. For other populations with a cohort effect we add an extra cohort index.
- We would also carry out cluster analysis on multi-population datasets to work out the best clustering structure for model fitting.

Bayesian modelling of critical illness insurance claim rates

Chunxiao Xie, Dr. George Streftaris, Dr. Erengul Dodd, Prof. Angus Macdonald

Actuarial Research Centre of the Institute and Faculty of Actuaries, Heriot-Watt University, UK.

Email: cx3@hw.ac.uk / g.streftaris@hw.ac.uk / e.dodd@soton.ac.uk/ a.s.macdonald@hw.ac.uk

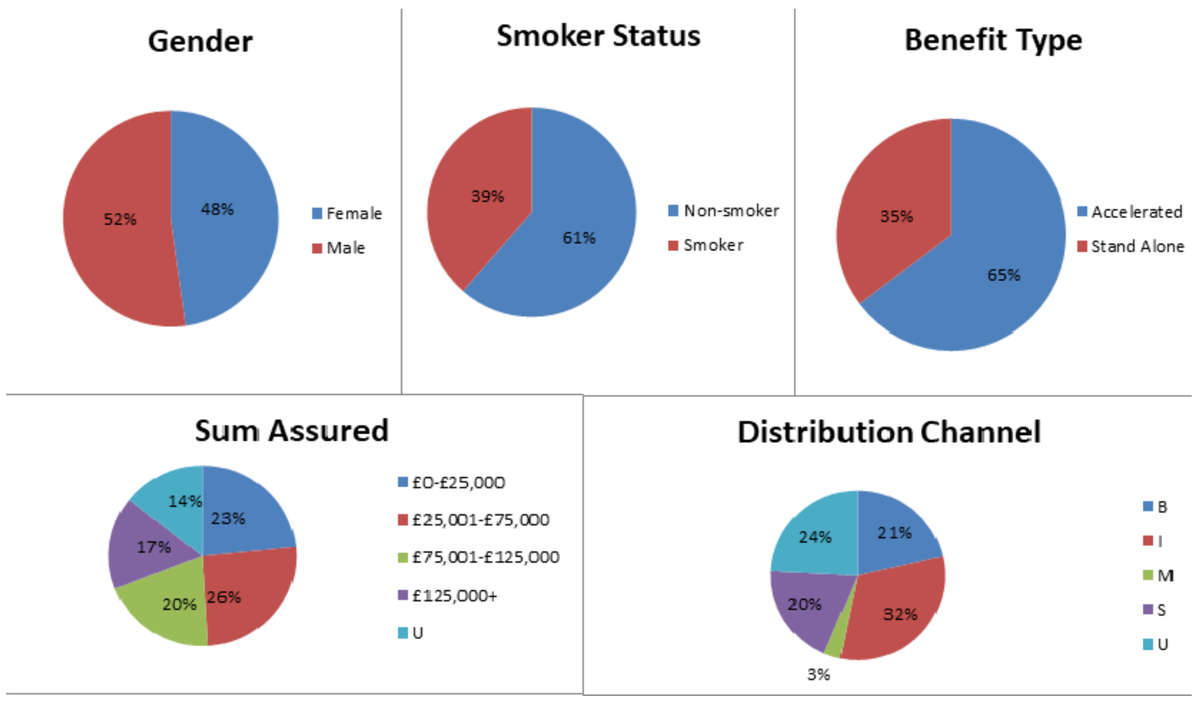
Introduction

This work focuses on the change of Critical illness insurance claim rate (CII) claim rates over time when Incurred-but-not-Settled adjustments are involved. Modelling the delay between dates of diagnosis and settlement of claims is also important for reserving purposes.

Data

Data Overview(CMI 2007-2010)

Data relate to CII claims enforceable and/or settled in 2007-2010. There are total 525,780 observations, grouped by combining the same characteristics such as gender, smoke status, benefit type, and benefit amount. Earlier CMI data also available from 1999 to 2005.



Methodology: Delay time modelling

Denote that y_i is the delay for claim i in days. $\boldsymbol{x_i}^{\boldsymbol{T}} = (x_{1,i}, x_{2,i}, \cdots, x_{p,i})$ are the risk factors: age, gender, benefit type, smoker status, settlement year, benefit amount, and policy duration. The delay time modelling is based on 1999 - 2005 Dataset.

Delay time generalized beta type 2 distribution (GB2) model

$$Y_i \sim GB2(\alpha, \tau, \gamma, s_i)$$

with

$$E(Y_i) = \exp(\eta_i) \qquad for\ i = 1, \cdots, n$$

where η_i is the linear predictor for claim i has: $\eta_i = \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta}$

Methodology: Claim rate modelling

Exposure adjustment

Denote $E(x, u; \theta)$as the original exposure with age last birthday of diagnosis being x and risk factors vector θ recorded at calender year u. Adjust exposures as:

$$E^*(x, u; \theta) = E(x, u; \theta)F(2010.5 - u; x, \theta)$$

$F(2010.5 - u; ; x, \theta)$: probability that a critical illness diagnosed at calender year u will be settled before the end of the last contribution year 2010.5 with the delay time distribution estimated from previous dataset.

Inception rate estimation

Denote that $C_{x,i,\theta}$ is the number of claim for policy i with risk profile θ at age x. $E_{x,i,\theta}$ and $\lambda_{x,i,\theta}$ are the exposure and inception rate for policy i, respectively. In order to smooth the inception rate, we add a age-smoker interaction to current covariates. The model shown as:

$$C_{x,i,\theta} \sim Poi(\lambda_{x,i,\theta} * E_{x,i,\theta})$$

with

$$\log(\lambda_{x,i,\theta}) = \delta x + \beta\boldsymbol{\theta} \qquad for\ i = 1, \cdots, n$$

where δ and β have normal vague priors.

Results

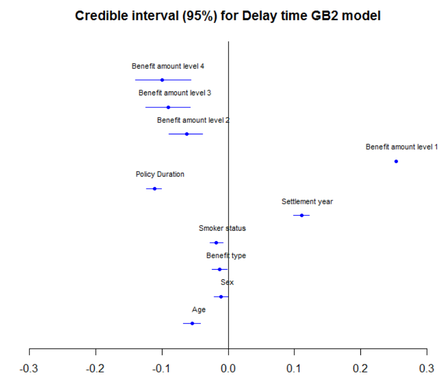


Figure 1: Risk factor estimates and 95% CIs for delay time

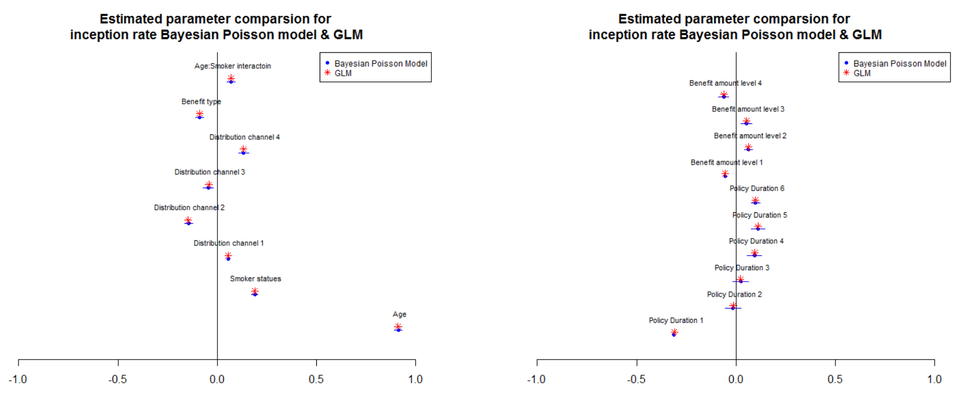


Figure 2: Risk factor estimates and 95% CIs for CII rates

Bayesian estimation of CII rates

GVS method is used in model selection for inception rates under a Poisson distribution. The best model involves covariates age, smoker status, distribution channel, policy duration, benefit amount, benefit type and age-smoker interaction.

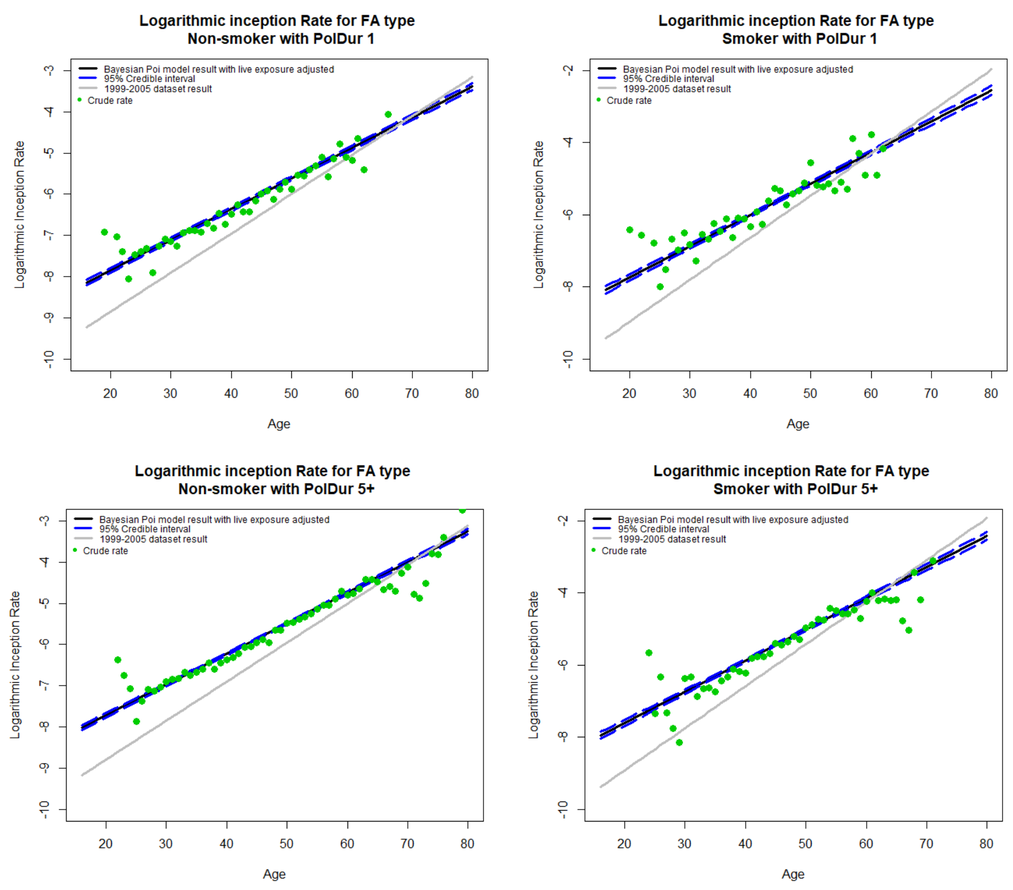


Figure 3: Estimated inception rates

Conclusions

- Our GB2 model for Delay distribution provides better fit compared with other models used before such as Lognormal model and Burr model.
- CII rates estimated using Bayes GLM methodology show a considerable increase for the 2007-2010 data at younger ages compared to 1999-2005 data.

Further Research

- Apply more types of different error structures (e.g.negative binomial distribution) Bayes model to the inception rate estimation and compute credible intervals for all risk factors.
- Compare the smooth claim rates obtained by our method to the ones obtained by the Continuous Mortality Investigation (CMI).

Age Heaping in Population Data of Emerging Countries

Andres Barajas, Prof. Andrew Cairns, Dr. Torsten Kleinow

Actuarial Research Centre of the Institute and Faculty of Actuaries, Heriot-Watt University, UK.

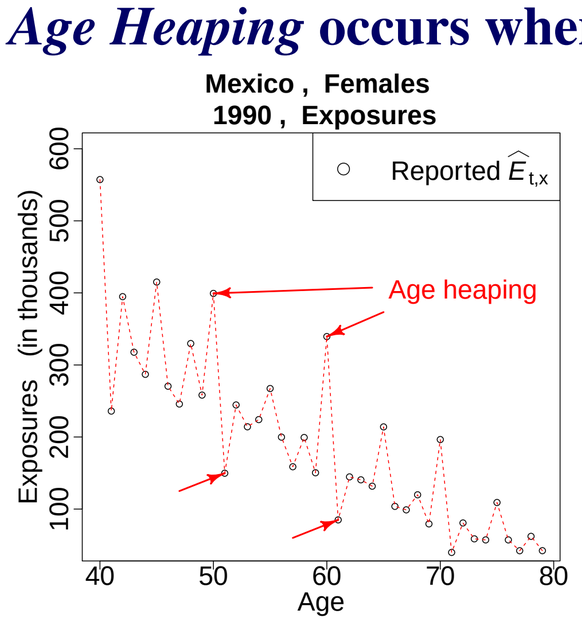
Email: ab108@hw.ac.uk / a.j.g.cairns@hw.ac.uk / t.kleinow@hw.ac.uk

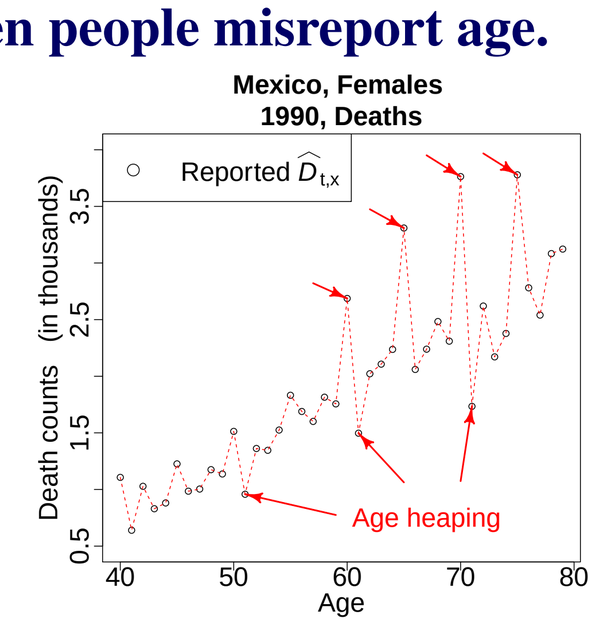
Introduction

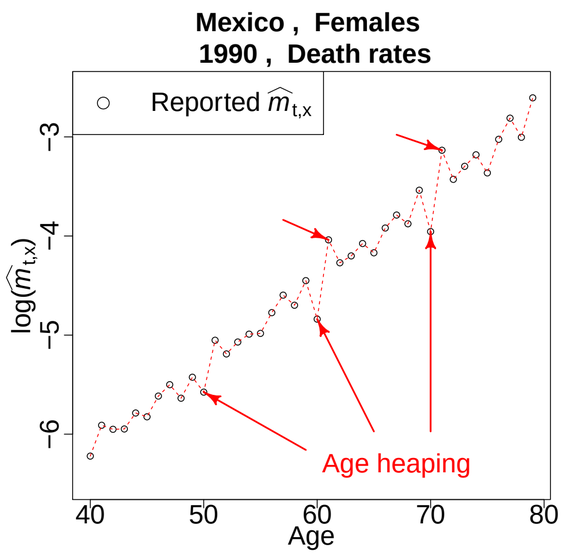
The purpose of the current research is to develop a model for emerging countries to improve the quality of the data of those countries which are not included in the Human Mortality Database such as Mexico.
Application: Reported data → Smoothed HMD → International Reinsurance.

Problem: Age heaping

Age Heaping occurs when people misreport age.







Data

Exposures: census years 1990, 1995, 2000, 2005, 2010 (reported / observed).
Death counts: annual.

Figure 1: Reported exposures $\widehat{E}_{t,x}$, deaths $\widehat{D}_{t,x}$, and death rates $\widehat{m}_{t,x}$, Mexico 1990.

Main Objective

To develop a model to improve the quality of the data related to death counts and exposures with *age heaping*.

Model and Notation

We design a penalised *log-likelihood* for the two dimensional data, such that for any cohort y we denote by $|y|$ the number of ages available for this cohort, that is, $n_y = |y|$ is the length of cohort y in our data set. The corresponding set of ages x is denoted by $\mathcal{X}_y$.

$\widehat{E}_{x,y}$ = reported exposures at age x, cohort y
$\widehat{D}_{x,y}$ = reported death counts at age x, cohort y
$E_{x,y}$ = true / actual exposures at age x, cohort y (not observed)
$D_{x,y}$ = true / actual death counts at age x, cohort y (not observed)

$$E_{x,y} \xrightarrow{\textit{Age heaping}} \widehat{E}_{x,y}$$
$$D_{x,y} \xrightarrow{\textit{Age heaping}} \widehat{D}_{x,y}$$

$$D_{x,y} \sim Poisson\Big(m_{x,y}E_{x,y}\Big),$$

where the true cohort mortality $m_{x,y}$ is given by:

$$m_{x,y} = \exp\Big[a_y + b_y\big(x-\overline{x}\big) + c_y\Big(\big(x-\overline{x}\big)^2 - \sigma_x^2\Big)\Big],$$

and the reported death rate is given by $\widehat{m}_{x,y} = \dfrac{\widehat{D}_{x,y}}{\widehat{E}_{x,y}}$

Approximate *log-likelihood*

$$l = \sum_{x,y} \widehat{D}_{x,y} \log\Big(m_{x,y}\widehat{E}_{x,y}\Big) - m_{x,y}\widehat{E}_{x,y} + C.$$

Smoothing

Penalised *log-likelihood*

$$llp = l - \lambda_1 p\big(a_y\big) - \lambda_2 p\big(b_y\big) - \lambda_3 p\big(c_y\big), \qquad p\big(\xi_y\big) = \sum_{\widetilde{y}=2}^{n_y-1}\Big(\triangle^2\xi_y\Big)^2,$$

where $\triangle^2\xi_y$ is the second order difference of ξ_y, and λ_1, λ_2 and λ_3 are the smoothing parameters.

Results

Fitted exposures $\widetilde{E}_{x,y} = \dfrac{\widehat{D}_{x,y}}{\widetilde{m}_{x,y}} = \dfrac{\text{Reported deaths}}{\text{Fitted Death Rate}}$

where $\widetilde{m}_{x,y} = \exp\Big[\widetilde{a}_y + \widetilde{b}_y\big(x-\overline{x}\big) + \widetilde{c}_y\Big(\big(x-\overline{x}\big)^2 - \sigma_x^2\Big)\Big]$.

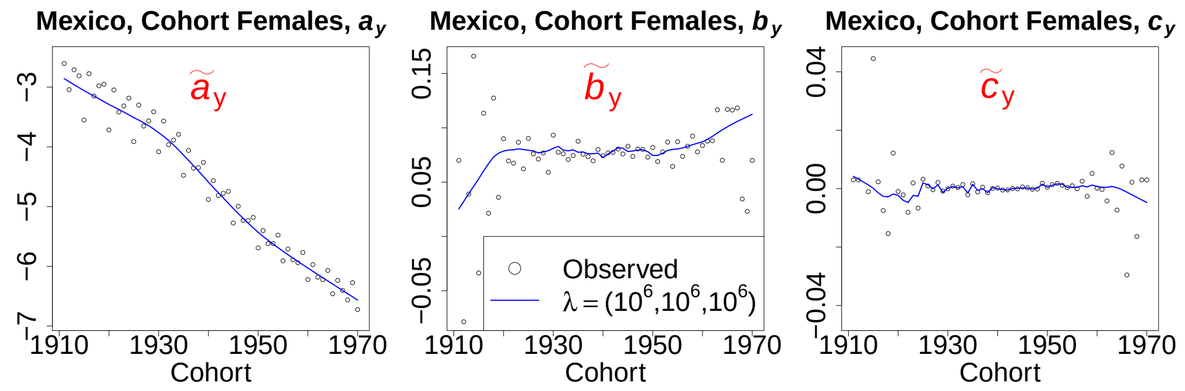


Figure 2: Fitted death rate parameters $\big(\widetilde{a}_y, \widetilde{b}_y, \widetilde{c}_y\big)$, Mexico.

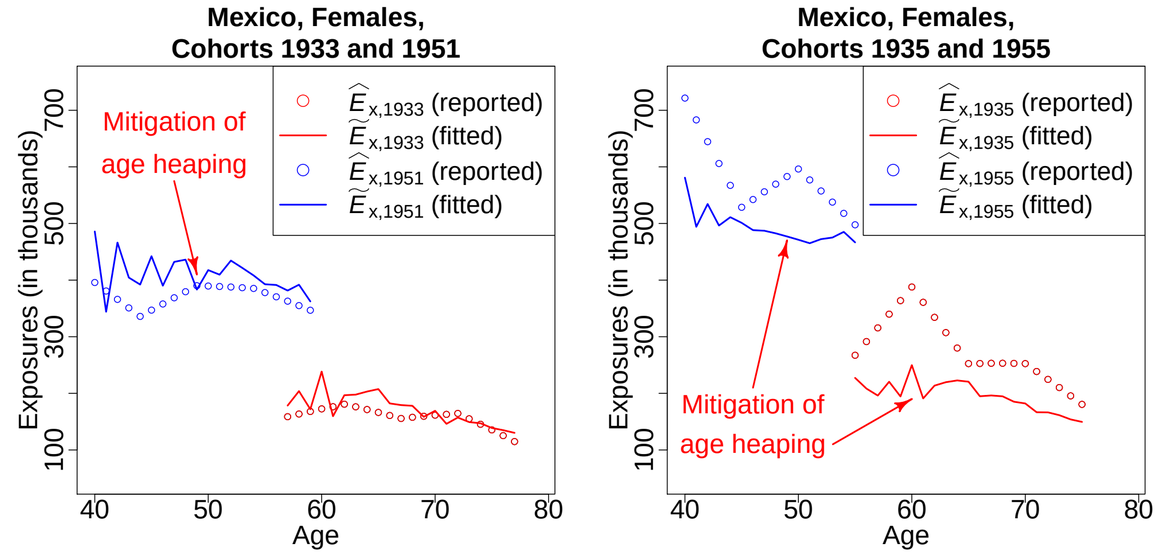


Figure 3: Reported $\widehat{E}_{x,y}$ and fitted exposures $\widetilde{E}_{x,y}$ different cohorts, Mexico. The big reduction for 1935 and 1955 cohorts, reflects the significant age heaping at ages 40, 50, 60, 70 in the census years.

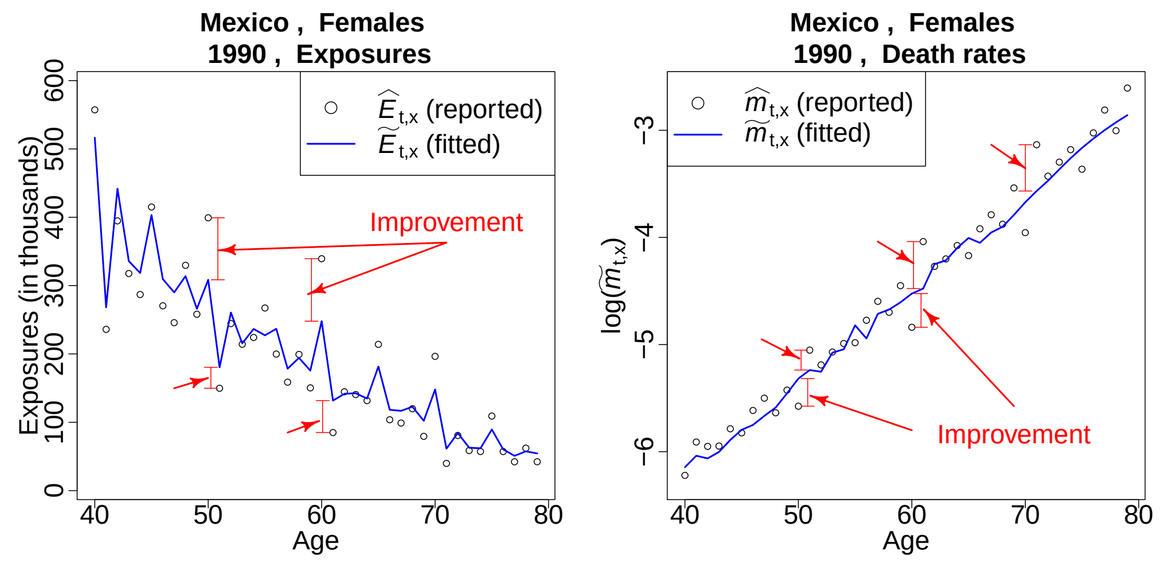


Figure 4: Fitted exposures $\widetilde{E}_{t,x}$ and death rates $\widetilde{m}_{t,x}$, Mexico 1990.

Conclusions

- We obtain much improved exposures $\widetilde{E}_{x,y}$, and death rates $\widetilde{m}_{x,y}$ when $\lambda_i = 10^6$. Therefore, this model improves the quality of the Mexican data by reducing *age heaping* across all cohorts.
- The remaining volatility in the fitted exposures comes from the death counts.

Forthcoming Research

- Specify a prior for all parameters in order to apply Bayesian methods and compute credible intervals for all parameters.
- Include constraints on death counts to reduce the volatility in the fitted exposures.
- Collaborate with HMD on Mexican data.

Modelling Mortality Rates by Cause of Death

Alexander Yiu, Dr. George Streftaris, Dr. Torsten Kleinow

Heriot-Watt University, UK.

Email: amy4@hw.ac.uk / G.Streftaris@hw.ac.uk / T.Kleinow@hw.ac.uk

Introduction

Historically, mortality rate modelling has focused on modelling mortality due to all causes. However, including the cause of death in the mortality models may provide improvements in the models and better mortality rate projections.

Problem: Cause-of-Death Mortality

Information about the cause of death may provide improved models and projections of mortality rates.

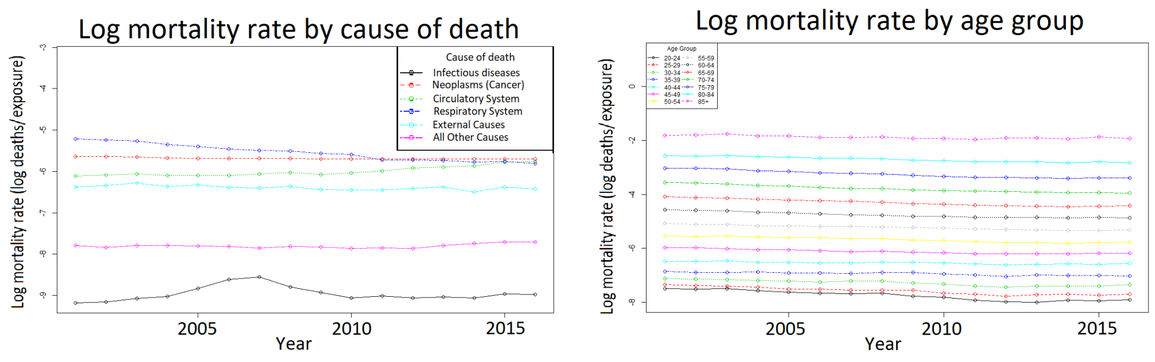


Figure 1: Observed log mortality rates by cause of death and by age group from years 2001 to 2016

Data

The data for England and Wales is obtained from the Office of National Statistics (2001 - 2016)

- Age in five-year age intervals
- Gender
- Population at June 30 of the given year (central exposures)
- Reported death counts grouped by ICD-10 classification
- Year of observation

Death counts are grouped into six main causes of death groups:

1. Infectious diseases
2. Cancers
3. Diseases of the circulatory system
4. Diseases of the respiratory system
5. External causes of death
6. All other causes of death

Objectives

1. Investigate mortality rates by cause of death
2. Predict future cause-specific mortality rates
3. Compare models for distributional assumptions

Model

We will model the death counts using Poisson and Negative Binomial generalised linear models. The analysis will focus on mortality rates for individuals aged 50 and above.

Notation

i = index for a unique combination of age group a_i, gender group s_i, and observation year y_i

$C_{(c,i)}$ = indicator variable for cause of death c for combination i

$D_{(c,i)}$ = reported death counts for cause of death c for individuals in combination i

E_i^C = central exposure for combination i

$\mathbf{x}_i$ = vector of covariates for combination i

$\beta_{(c,i)}$ = vector of parameters for cause of death c and combination i

The observed cause-specific mortality rate is given by $m_{(c,i)} = \frac{D_{(c,i)}}{E_i^C}$

Fitted cause-specific death counts

$$\widehat{D}_{(c,i)} = E_i^C \cdot \exp(\widehat{\beta}_{(c,i)}\mathbf{x}_i)$$

Model selection is performed using AIC. Interaction terms between two covariates are considered.

Results

The Negative Binomial model is chosen to the model the death counts where the age and year covariates are standardised and the baseline is male deaths due to infectious diseases. This model has interaction terms for each of age, gender, and year with the cause of death, as well as an interaction between age and gender.

$$\begin{aligned}\log \widehat{D}_{(c,i)} = &-8.108 + 0.441a_i - 0.274s_i - 0.003y_i + 3.620C_{(2,i)} + 3.902C_{(3,i)} \\ &+ 2.461C_{(4,i)} + 1.455C_{(5,i)} + 2.271C_{(6,i)} + 0.063a_is_i - 0.132a_iC_{(2,i)} \\ &+ 0.080a_iC_{(3,i)} + 0.155a_iC_{(4,i)} - 0.183a_iC_{(5,i)} + 0.041a_iC_{(6,i)} \\ &- 0.051s_iC_{(2,i)} - 0.389s_iC_{(3,i)} - 0.068s_iC_{(4,i)} - 0.302s_iC_{(5,i)} \\ &+ 0.082s_iC_{(6,i)} - 0.010y_iC_{(2,i)} - 0.047y_iC_{(3,i)} - 0.011y_iC_{(4,i)} \\ &+ 0.009y_iC_{(5,i)} + 0.008y_iC_{(6,i)} + \log E_i^C\end{aligned} \quad (1)$$

Fitted cause-specific log mortality rates

$$\log \widehat{m}_{(c,i)} = \log\left(\frac{\widehat{D}_{(c,i)}}{E_i^C}\right)$$

where $\widehat{D}_{(c,i)}$ is the fitted death count for cause of death c and combination i.

Fitted Log Mortality Rates

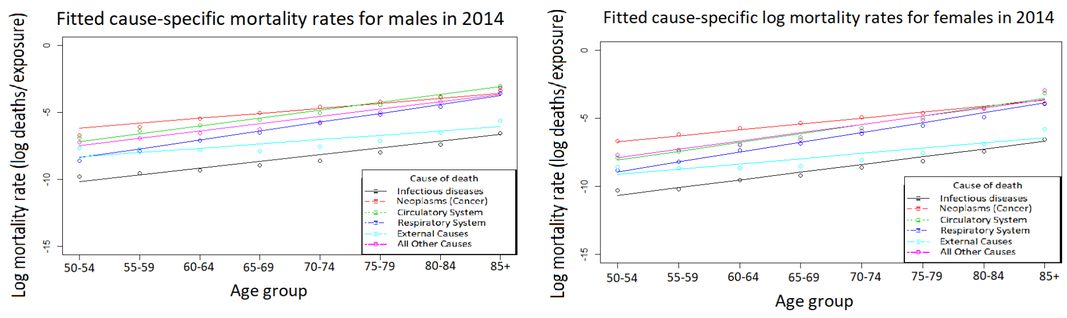


Figure 2: Fitted cause-specific log mortality rates in 2014

Prediction Intervals

Models are fitted to the first 15 years of data from 2001 to 2015 and used to predict the mortality rates in 2016. Bootstrap methods are used to construct prediction intervals.

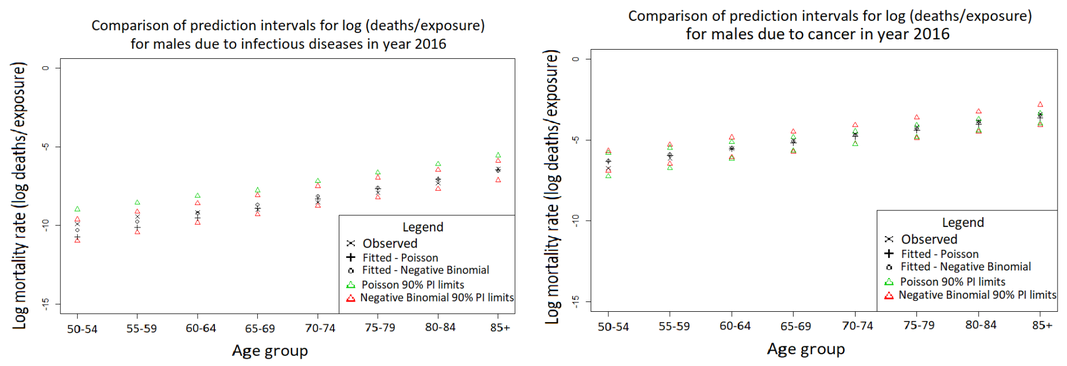


Figure 3: Prediction intervals for cause-specific log mortality rates in 2016

Conclusions

- The Poisson and Negative Binomial models provide similar fitted values for the number of deaths
- The Negative Binomial model provides better modelling of overdispersion in the data and consistent prediction interval widths for log mortality rates
- The interaction terms between cause of death and the age, gender, and year are important for modelling the mortality rates by cause of death

Forthcoming Research

- Investigate cause-of-death mortality under a Bayesian framework.
- Investigate correlation between different causes of death and their joint effects on the rate of mortality.
- Investigate cause-of-death mortality experiences of other countries.

References

Davison, A. C., and Hinkley, D. V. (1997). Bootstrap methods and their application. Cambridge university press.

Office for National Statistics (2017). "The 21st century mortality files - deaths dataset, England and Wales", [Electronic Dataset] Available at https://www.ons.gov.uk/people population andcommunity/births deathsandmarriages /deaths/datasets/the21stcentury mortalityfilesdeathsdataset [Accessed 13 October, 2017]

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth edition. Springer.

Genetics, Insurance, and Hypertrophic Cardiomyopathy (HCM)

Oytun Haçarız, Dr. Torsten Kleinow and Prof. Angus S. Macdonald

Actuarial Mathematics & Statistics, Heriot-Watt University, Edinburgh, UK.

Email: oh6@hw.ac.uk / t.kleinow@hw.ac.uk / a.s.macdonald@hw.ac.uk

Introduction

Since DNA testing became available in the 1990s, the use of genetic information by insurers has been disputed. *Several studies estimated increases in life insurance premium rates under undisclosed genetic test results to insurers:*

- **Macdonald & Yu (2011)** modelled six major single-gene disorders and estimated the increases of up to 0.6% .
- **Howard (2014)** modelled thirteen genetic disorders, **"cardiomyopathies"** different than Macdonald & Yu (2011), and estimated the increases of 12% .

Cardiomyopathies are a large group of disorders affecting the heart muscle. Hypertrophic Cardiomyopathy (HCM) is the most prevalent of these disorders.

- **We aim to model the impact of genetic testing in HCM in a life insurance market.**

Hypertrophic Cardiomyopathy (HCM)

HCM is the thickening (hypertrophying) of the heart muscle.

- **Onset:** It generally exists at early adulthood.
- **Diagnosis:** Echocardiography and Cardiac Magnetic Resonance.
- **Symptoms:** Chest pain, shortness of breath, syncope, palpitations.
- **Genetics:** Autosomal dominant mutations in over 8 genes. Mutations in the MYBPC3 gene are associated with late-onset HCM while mutations in other genes are associated with early-onset HCM.
- **Mortality:** It is not related to gender and race. The causes of HCM death:
  a. Sudden Cardiac Arrest (SCA) (common at young ages),
  b. Heart Failure (HF),
  c. Stroke (common at older ages).

Modelling HCM for Life Insurance Market

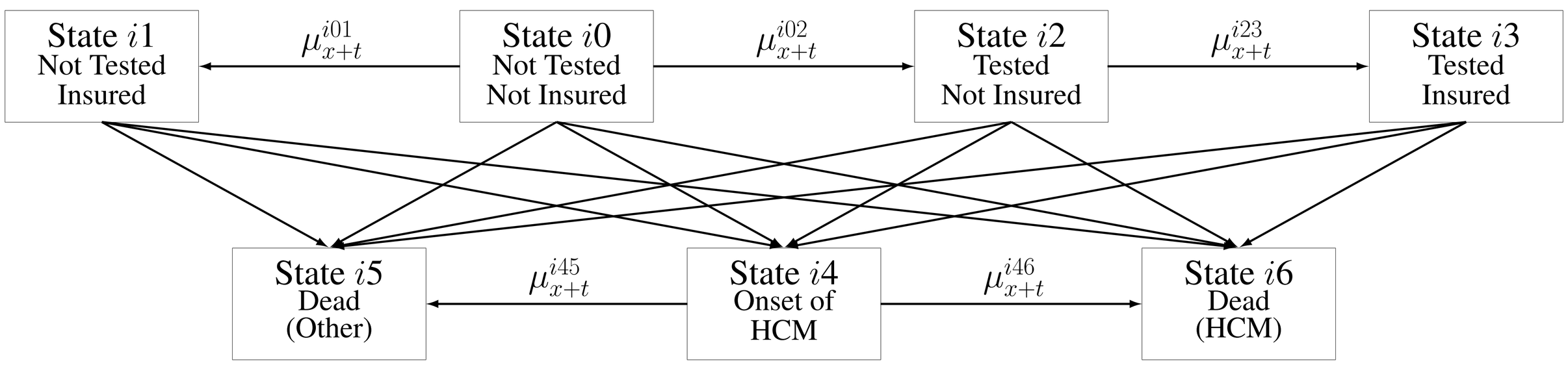


Figure 1: A multi-state Markov model of genetic testing in HCM for a person in ith risk sub-population in a life insurance market.

Formulating the Model

Our model is the discrete-state continuous-time Markov model. i label represents the respective sub-populations to the published epidemiology of HCM:

$i : 0$, Not At Risk of HCM

$i : 1$, A known Early-Onset HCM Mutation Absent	$i : 2$, A known Early-Onset HCM Mutation Present
$i : 3$, A known Late-Onset HCM Mutation Absent	$i : 4$, A known Late-Onset HCM Mutation Present
$i : 5$, An unknown Early-Onset HCM Mutation Absent	$i : 6$, An unknown Early-Onset HCM Mutation Present
$i : 7$, An unknown Late-Onset HCM Mutation Absent	$i : 8$, An unknown Late-Onset HCM Mutation Present

We define the *occupancy probabilities* of the model as follows:

$$ {}_{dt}p^{ijk}_{x+t} = P[\text{In state } ik \text{ at age } x+t+dt \mid \text{In state } ij \text{ at age } x+t], \quad j \neq k. \quad (1) $$

Then, we can obtain the *transition intensities* as follows:

$$ \mu^{ijk}_{x+t} = \lim_{dt \to 0} \frac{{}_{dt}p^{ijk}_{x+t}}{dt} \quad (2) $$

Parametrising the Model I: Rates at Birth

- **Prevalence rate in the general population:** $\sim 0.2\%$ (Maron et al. 1995).
- **Frequency of the gene mutations in HCM Population:** 40-60%, *most frequently in the MYBPC3 gene (15-30%) and MYH7 gene (10-20%)*, of HCM patients, tested with a mutation in a known gene (Elliott et al. 2014).

Parametrising the Model II: Hazard Rates

- **Uptake of genetic testing:** Rate of uptake of the testing at risk first-degree relatives is $\sim$70% in the first year.
- **Insurance purchase:** Adverse rate of the purchase is assumed 25% per year.
- **Onset:** Rate of onset is defined as $\mu^{\text{Onset}}_x = F'(x)/(1-F(x))$ where $F(x) = P[\text{Phenotype present at age x}]$. See Figure 2.
- **Mortality:** Rate of HCM death is $\sim$0.5% per year (Maron et al. 2016).

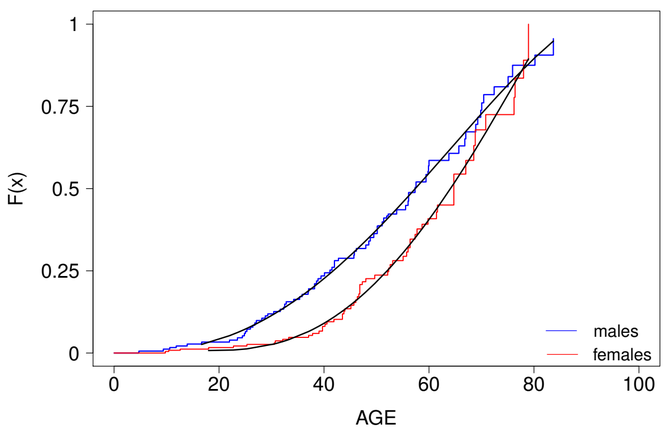


Figure 2: Late-Onset Penetrance Rate of HCM from Christiaans et al. (2011).

Simulation and Computation of the Model

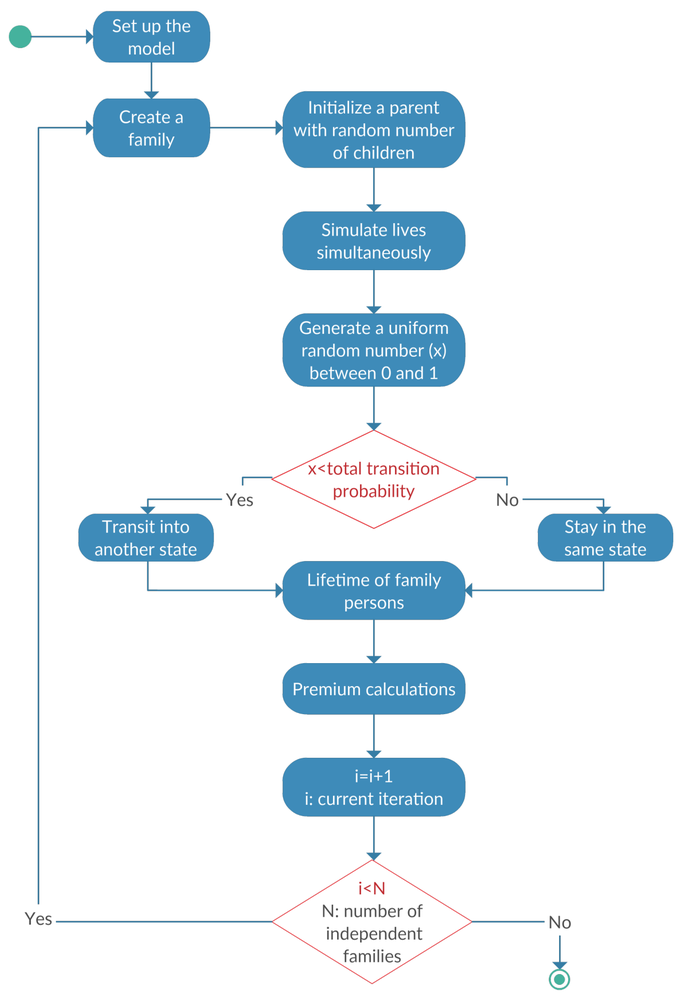


Figure 3: An algorithmic implementation of the model, in C++ programming language, for HCM population in a life insurance market.

Further Work

We wrote the simulation code of lifetimes of family persons; our goal now is to

- implement the code of premium calculations of the simulated population,
- have results and compare them with the earlier studies.

References

Christiaans, I., Birnie, E., Bonsel, G. J., Mannens, M. M., Michels, M., Majoor-Krakauer, D., Dooijes, D., van Tintelen, J. P., van den Berg, M. P., Volders, P. G., Arens, Y. H., van den Wijngaard, A., Atsma, D. E., Helderman-van den Enden, A. T., Houweling, A. C., de Boer, K., van der Smagt, J. J., Hauer, R. N., Marcelis, C. L., Timmermans, J., van Langen, I. M. & Wilde, A. A. (2011), 'Manifest disease, risk factors for sudden cardiac death, and cardiac events in a large nationwide cohort of predictively tested hypertrophic cardiomyopathy mutation carriers: determining the best cardiological screening strategy', *European Heart Journal* **32**(9), 1161.

Elliott, P. M., Anastasakis, A., Borger, M. A., Borggrefe, M., Cecchi, F., Charron, P., Hagege, A. A., Lafont, A., Limongelli, G., Mahrholdt, H., McKenna, W. J., Mogensen, J., Nihoyannopoulos, P., Nistri, S., Pieper, P. G., Pieske, B., Rapezzi, C., Rutten, F. H., Tillmanns, C. & Watkins, H. (2014), '2014 esc guidelines on diagnosis and management of hypertrophic cardiomyopathy', *European Heart Journal* .

Howard, R. C. W. (2014), 'Report to cia research committee: Genetic testing model: If the underwriters had no access to known results', *Canadian Institute of Actuaries (CIA)* .

Macdonald, A. & Yu, F. (2011), 'The impact of genetic information on the insurance industry: Conclusions from the 'bottom-up'modelling programme', *Astin Bulletin* **41**(02), 343–376.

Maron, B. J., Gardin, J. M., Flack, J. M., Gidding, S. S., Kurosaki, T. T. & Bild, D. E. (1995), 'Prevalence of hypertrophic cardiomyopathy in a general population of young adults echocardiographic analysis of 4111 subjects in the cardia study', *Circulation* **92**(4), 785–789.

Maron, B. J., Rowin, E. J., Casey, S. A., Lesser, J. R., Garberich, R. F., McGriff, D. M. & Maron, M. S. (2016), 'Hypertrophic cardiomyopathy in children, adolescents and young adults associated with low cardiovascular mortality with contemporary management strategies', *Circulation* **133**, 62–73.