September 13, 2019

Coherent cause specific life table closure : a compositional data approach

Samuel Piveteau, Julien Tomas

Longevity R&D Centre | Scor SE Life
SAF | ISFA | Université Claude Bernard Lyon 1

SCOR The Art & Science of Risk | Life

Fifteenth International Longevity and Risk Capital Markets Solutions Conference

15 2019 LONGEVITY washington, d.c.

Contents

SCOR The Art & Science of Risk | Life

Introduction

The uncertainty regarding mortality extrapolation is increasing when considering cause-specific mortality. This is especially relevant when observing mortality behaviour at high ages.

The lack of data requires extrapolation based on specific assumptions regarding the high ages mortality shape.

These hypotheses lead to particular mathematical models, combining both statistics and analytical assumptions.

Extrapolating the cause specific mortality without taking into account them may generate inconsistent results.

We propose to combine Compositional Data Analysis framework with usual life table closure methods to derive cause specific mortality after after 90.

We present an application on US Female data for 2017.

SCOR The Art & Science of Risk | Life

Theorical framework

We consider a cohort or a population observed on a limited period.

- Age range of observation is $[\omega_0, ..., \omega_1]$;
- I causes of death are identified ;
- X^k the lifespan of individual k ;
- C^k the cause of death of individual k if dead ;
- μ_x^i the force of mortality for cause i at age x ;
- $\mu_x = \sum_{i=1}^{I} \mu_x^i$ the force of mortality at age x ;
- E_x^k the time spent by individual k at age x ;
- $D_x^{k,i}$ the indicator function equal to 1 if the individual k is dead from cause i at age x.
- $D_x^k = \sum_i D_x^{k,i}$ the indicator function equal to 1 if the death of the individual k occurs between ages x and $x+1$.

We assume piecewise mortality for all causes.

SCOR The Art & Science of Risk | Life

Theorical framework

The individual likelihood for k is :

$$L^k((\mu_x^i)_{i\in\{1,\dots,I\}}) = \prod_{x=\omega_0}^{\omega_1} exp(-E_x^k\mu_x)(\mu_x^1)^{D_x^{k,1}}\dots(\mu_x^I)^{D_x^{k,I}}.$$

Introducing $\pi_x^i = \frac{\mu_x^i}{\mu_x}$, the individual likelihood becomes :

$$L^k(\mu_x,(\pi_x^i)_{i\in\{1,\dots,I\}}) = \prod_{x=\omega_0}^{\omega_1} exp(-E_x^k\mu_x)\mu_x^{D_x^k}\prod_{i=1}^{I}(\pi_x^i)^{D_x^{k,i}}.$$

Considering the N individuals, we obtain :

$$L(\mu_x,(\pi_x^i)_{i\in\{1,\dots,I\}}) = \underbrace{\prod_{x=\omega_0}^{\omega_1} exp(-E_x\mu_x)\mu_x^{D_x}}_{L_1}\underbrace{\prod_{x=\omega_0}^{\omega_1}\prod_{i=1}^{I}(\pi_x^i)^{D_x^i}}_{L_2};$$

where $D_x = \sum_k D_x^k$,$E_x = \sum_k E_x^k$ and $D_x^i = \sum_k D_x^{k,i}$.

SCOR The Art & Science of Risk | Life

Theorical framework

The likelihood is splitted in two parts : L_1 is relative to the global force of mortality and L_2 measures the relative contribution of each cause to the mortality.

Furthemore, we can see that D_x can be considered as $\mathcal{P}(E_x\mu_x)$ distributed, and $(D_x^i)_i|D_x$ as a $MN(D_x, (\pi_x^i)_i)$.

We assume that the relationship between the coefficients $(\mu_x)_x$ and $(\pi_x^i)_{x,i}$ and ages can respectively be measured through parameters β^μ and β^π.

Hence, the log-likelihood is :

$$l(\beta^\mu, \beta^\pi) = l_1(\beta^\mu) + l_2(\beta^\pi).$$

In consequence, we can fit separately the l_1 and l_2.

SCOR The Art & Science of Risk | Life

Extrapolation of global mortality at high ages

1 **Denuit and Goderniaux :**

$$log(\hat{q}_x)(\beta^\mu) = \beta_1^\mu + \beta_2^\mu x + \beta_3^\mu x^2 + \epsilon_x$$

$$u.c\ q_{x_{\omega_2}} = 1\,;\ \frac{\partial q_x}{\partial x}_{x=x_{\omega_2}} = 0$$

2 **Gompertz :**

$$\mu_x(\beta^\mu) = \beta_1^\mu exp\left(-\beta_2^\mu x\right) + \beta_3^\mu$$

3 **Kannisto :**

$$\mu_x(\beta_1^\mu, \beta_2^\mu) = \frac{\beta_1^\mu exp(\beta_2^\mu x)}{1 + \beta_1^\mu exp(\beta_2^\mu x)}$$

SCOR The Art & Science of Risk | Life

Extrapolation of global mortality at high ages

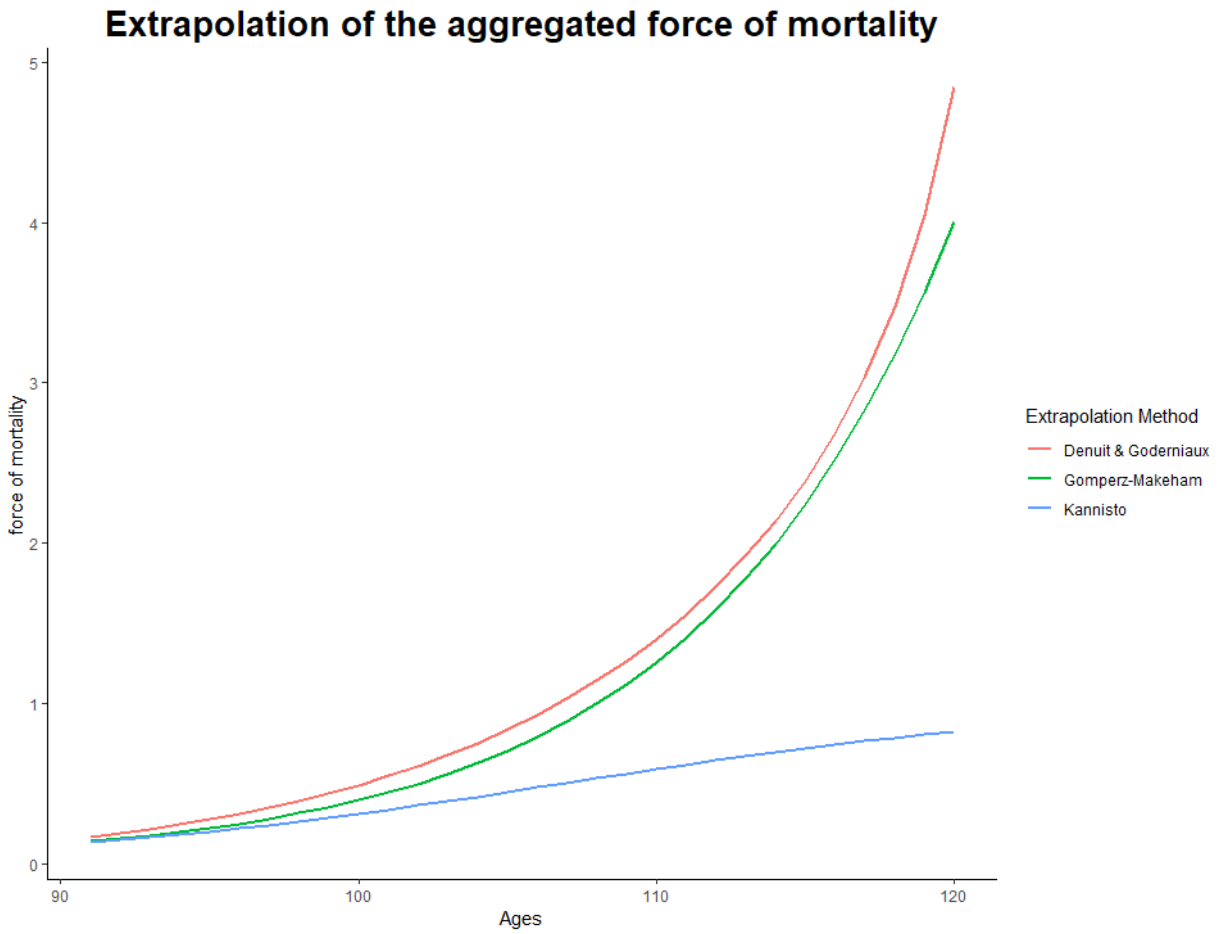

Figure – Extrapolation application on US female population for 2017

SCOR The Art & Science of Risk | Life

Cause specific contribution : a GLM framework

The π_x vectors belong to the I-dimension simplex space, and are called compositional data (cf. Aitchison(1986)).

We have realisations of these compositional vectors $(\frac{D_x^i}{D_x})_i$ and the denominators D_x. So a common and appropriate framework is the multinomial regression.

The standard transformation for the parameters is the additive log-ratio (alr), which corresponds to the usual multi logistic regression.

- $alr(\pi_x) = \left(log\left(\frac{\pi_x^1}{\pi_x^I}\right), ..., log\left(\frac{\pi_x^{I-1}}{\pi_x^I}\right)\right)$

Two other transformations can be drawn from the compositional data theory.

- $clr(\pi_x) = \left(log\left(\frac{\pi_x^1}{(\prod_i^I \pi_x^i)^{\frac{1}{I}}}\right), ..., log\left(\frac{\pi_x^I}{(\prod_i^I \pi_x^i)^{\frac{1}{I}}}\right)\right)$
- $ilr(\pi_x) = log(V^T \pi_x)$ whith V an orthonormal base of vectors from the simplex space $\mathbb{S}^I$.

SCOR The Art & Science of Risk | Life

Cause specific contribution : a GLM framework

Denoting $\boldsymbol{D_x} = (D_x^1, ..., D_x^I)$,we rewrite the multinomial model under its exponential expression :

$$\mathbb{P}(\boldsymbol{D_x}|D_x) = G(D_x^1, ..., D_x^I) exp\left(\theta_x^t T(\boldsymbol{D_x}) - A(\theta_x)\right).$$

Regarding the log-ratio transformation used, we can express the different components of the exponential model :

Method	$T(\boldsymbol{D_x})$	θ_x	$A(\theta_x)$
alr	$(\boldsymbol{D_x})^{-I}$	$alr(\pi_x)$	$D_x log(1 + \sum_{i=1}^{I-1} exp(\theta_x^i))$
clr	$(\boldsymbol{D_x})^{-I} - D_x^I$	$clr(\pi_x)^{-I}$	$D_x log\left(\sum_{i=1}^{I-1} exp(\theta_x^i) + exp(-\sum_{i=1}^{I-1} \theta_x^i)\right)$
ilr	$V^T \boldsymbol{D_x}$	$ilr(\pi_x)$	$D_x log\left(\sum_i^I exp\left(V_{i,.}^T \theta_x\right)\right)$

Table – CoDa Interpretation for GLMs

A P-spline approach for multi logistic extrapolation

We estimate the cause specific contribution to mortality considering the age as a variable in order to extrapolate them by including ages beyond ω_1.

Due to some trend changes, we use smooth methods to catch the behaviour in the curves and extrapolate the last trend observed.

We propose a P-splines method applied on the 3 multinomial GLMs defined above. The method is an adaptation for the P-splines method (Marx and Eilers (1996)) to the multi logit regression. The extrapolation process is the same as described in Currie et al (2004).

SCOR The Art & Science of Risk | Life

A P-spline approach for multi logistic extrapolation

We denote $B(x) = (B_{1,2}(x), B_{2,2}(x), ..., B_{n,2}(x))^T$ the column vector containing the n B-spline values in x.

$$M(x) = \begin{pmatrix} 1 & 0 & ... & 0 & B(x)^T & 0 & ... & 0 \\ 0 & 1 & ... & 0 & 0 & B(x)^T & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 0 & 0 & ... & B(x)^T \end{pmatrix}$$

We pool all the age information into a single matrix $\overline{M}$

$$\overline{M} = (M(\omega_0)^T, M(\omega_0 + 1)^T, ..., M(\omega_1)^T)^T$$

This leads to the following maximization program :

$$max_{\beta^\pi} \sum^{\omega_1} (T(\boldsymbol{D_x})^T(\overline{M}M(x)\beta^\pi) - A(\overline{M}(x)\beta^\pi)) - \frac{\lambda}{2}\sum_i (\Delta^2\beta_i^\pi)^2$$

SCOR The Art & Science of Risk | Life

A P-spline approach for multi logistic extrapolation

All the observations $T(\boldsymbol{D_x})$ are stacked into Y, then the program can be solved using the IWLS algorithm :

$$\beta^{\pi,(k+1)} \leftarrow (\overline{M}^T W(\beta^{\pi,(k)})\overline{M} + \lambda Df_2^T Df_2)^{-1}\overline{M}^T W(\beta^{\pi,(k)})Z(\beta^{\pi,(k)})$$

where :

- $Z(\beta^\pi) = W^{-1}(\beta^\pi)(Y - g^{-1}(\theta)) + \overline{M}\beta^\pi$;
- $W(\beta^{(k)})$ the weight matrix formed by applying the blockdiag operation on the hessian matrices of $A(\theta_x)$ for all ages x ;
- Df_2 the difference matrix of order 2 adapted for the multi logit scheme.

SCOR The Art & Science of Risk | Life

A P-spline approach for multi logistic extrapolation

Method	Penality Parameter	Log Likelihood	AIC
alr	0.07879205	1673811	3347735
clr	0.1502377	1673811	3347737
ilr	0.1176795	1673811	3347736

Table – Information regarding the parameters for the GLMs component

As seen in the table, the three GLM methods present similar results regarding the fitting and the AIC.

SCOR The Art & Science of Risk | Life

Analysis of proportion of deaths by cause

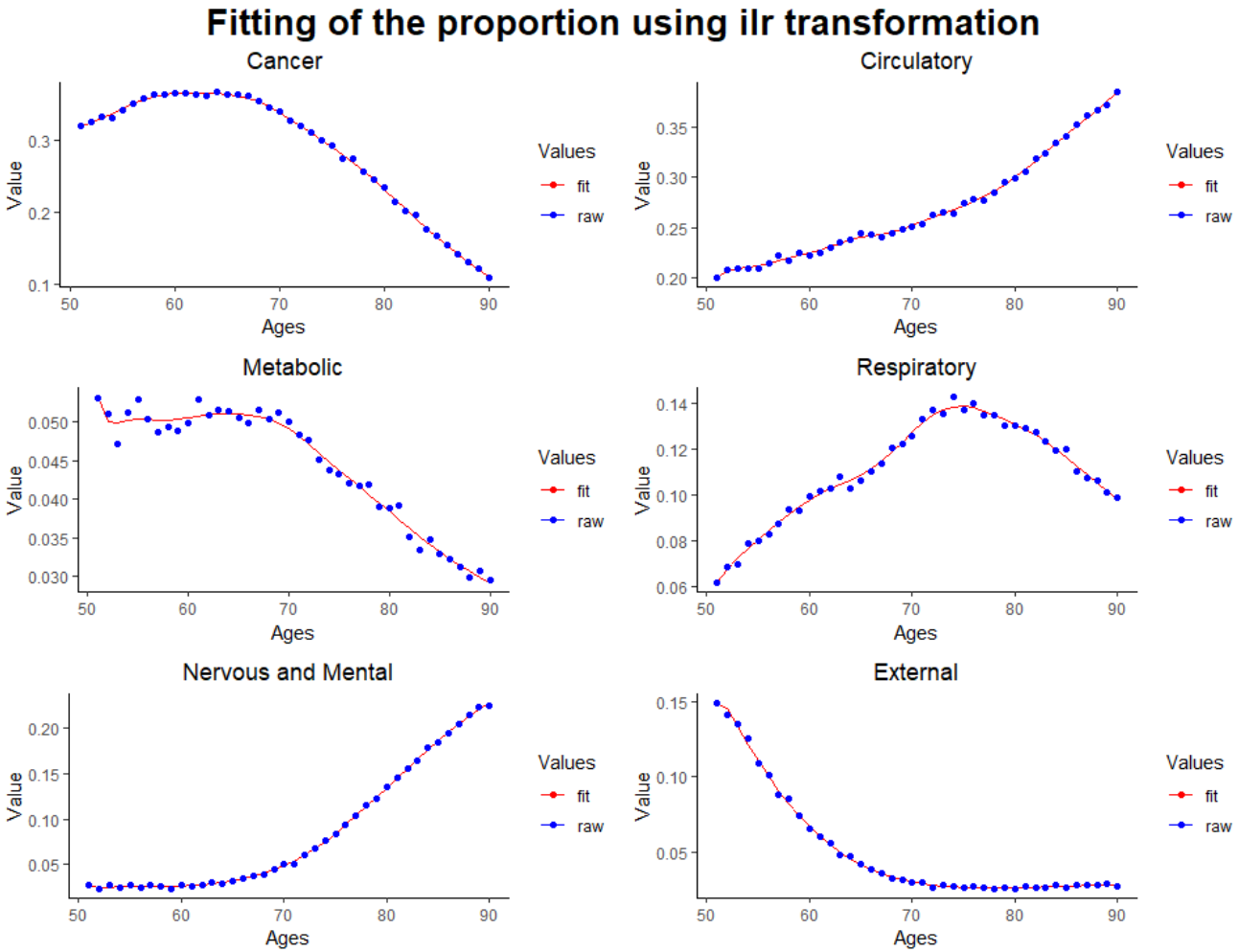

Figure – P-splines fitting using ilr transformation

Analysis of proportion of deaths by cause

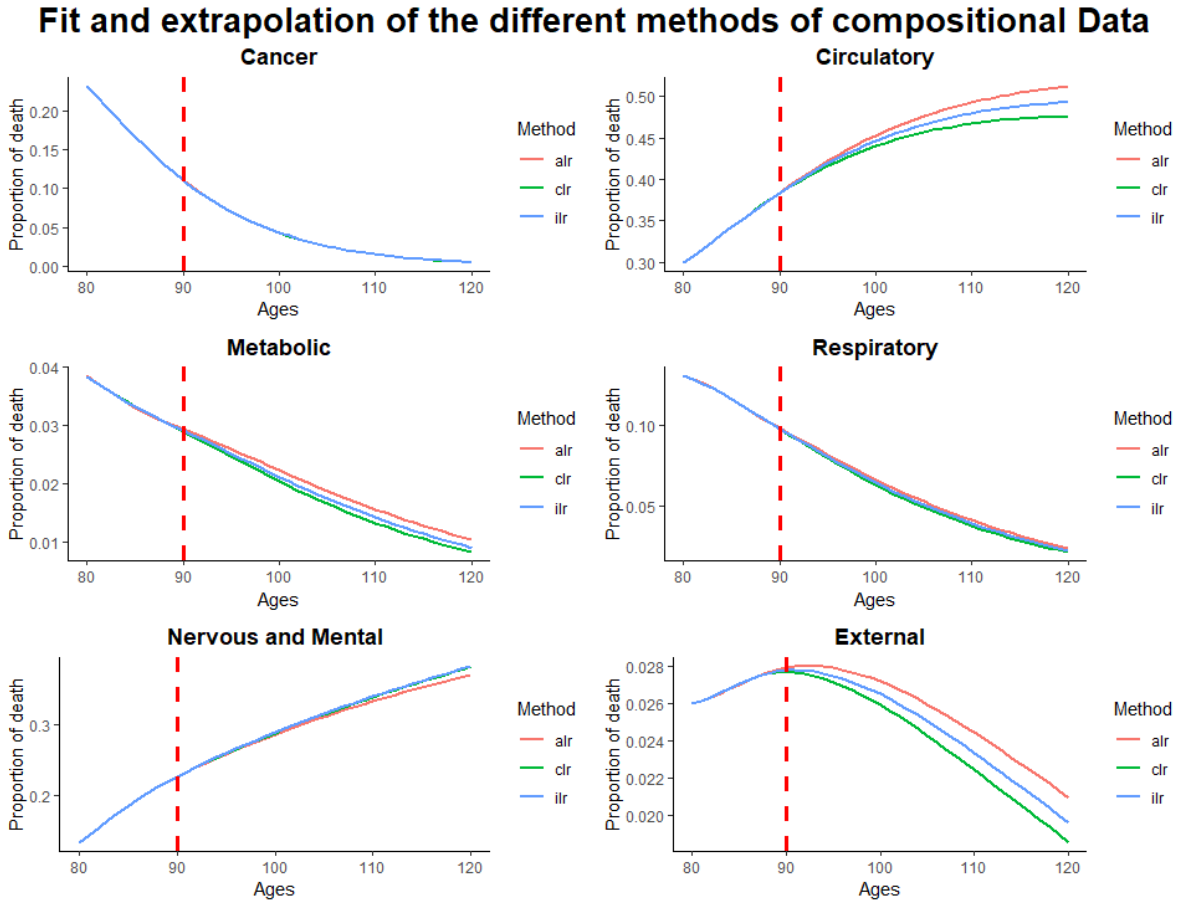

Figure – Extrapolation of the Cause of death contribution to the aggregated force of mortality

Analysis of proportion of deaths by cause

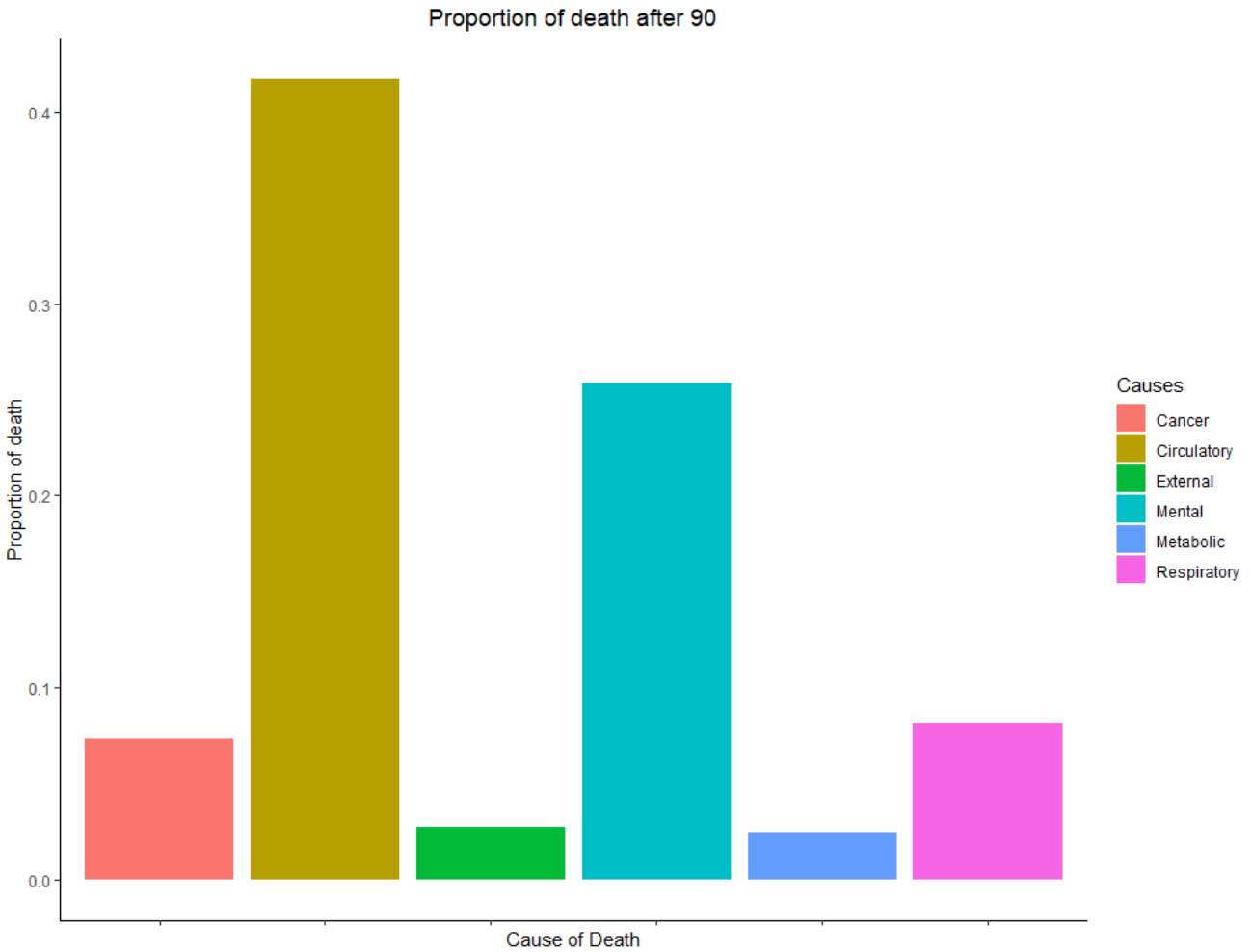

Figure – View on the proportion of death by cause after 90, by averaging

SCOR The Art & Science of Risk | Life

Analysis of proportion of deaths by cause

- The result regarding the repartition of deceased among causes of death slightly varies depending both completion approaches and the choice of the GLMs.
- Some completion methods will delay the age at death of the population. Then causes of death which relatively increase will regroup larger amount of people.
- The relative dynamic of the cause of death can also vary regarding the choice of the GLM transformation. This effect cumulates with the impact of the completion approaches.

SCOR The Art & Science of Risk | Life

Analysis of proportion of deaths by cause

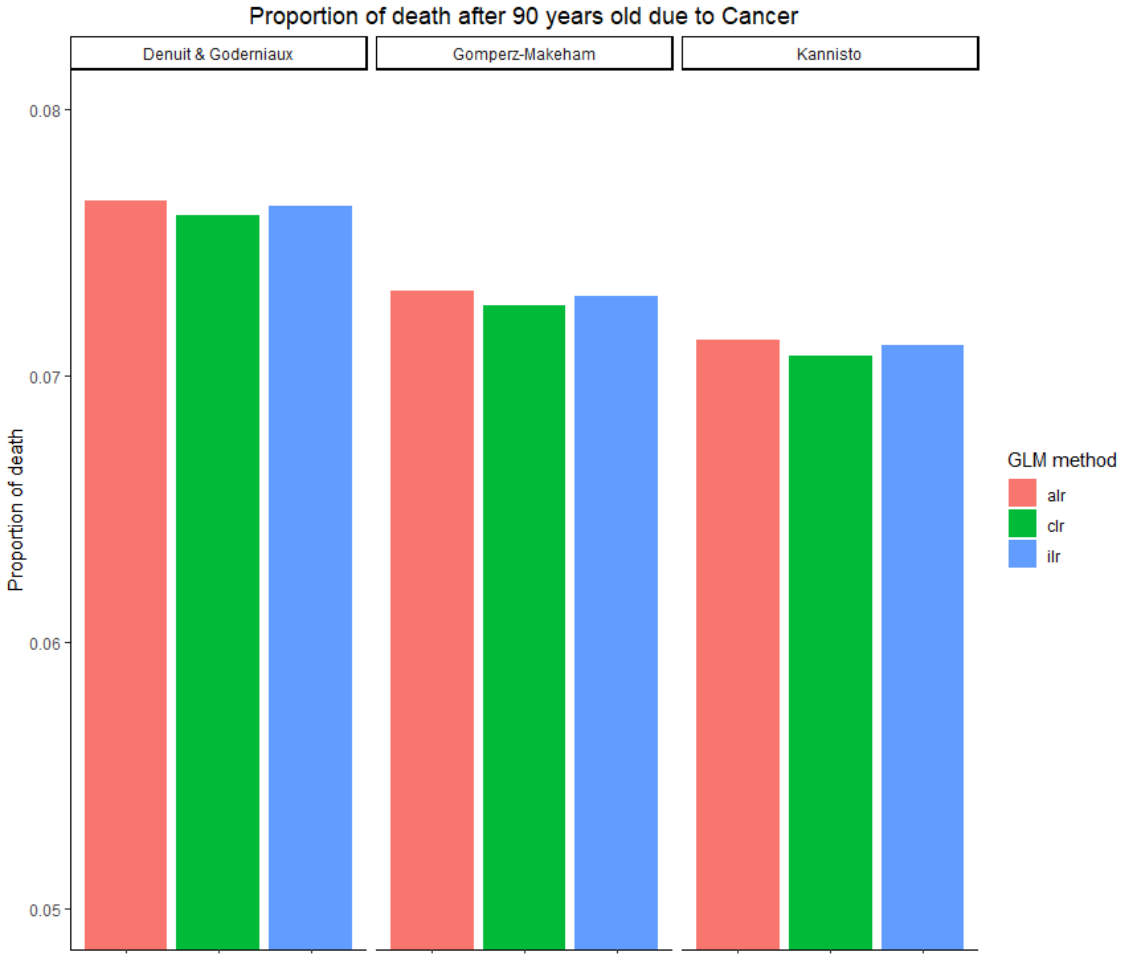

Figure – Cancers proportion after 90 for all methods

SCOR The Art & Science of Risk | Life

Analysis of proportion of deaths by cause

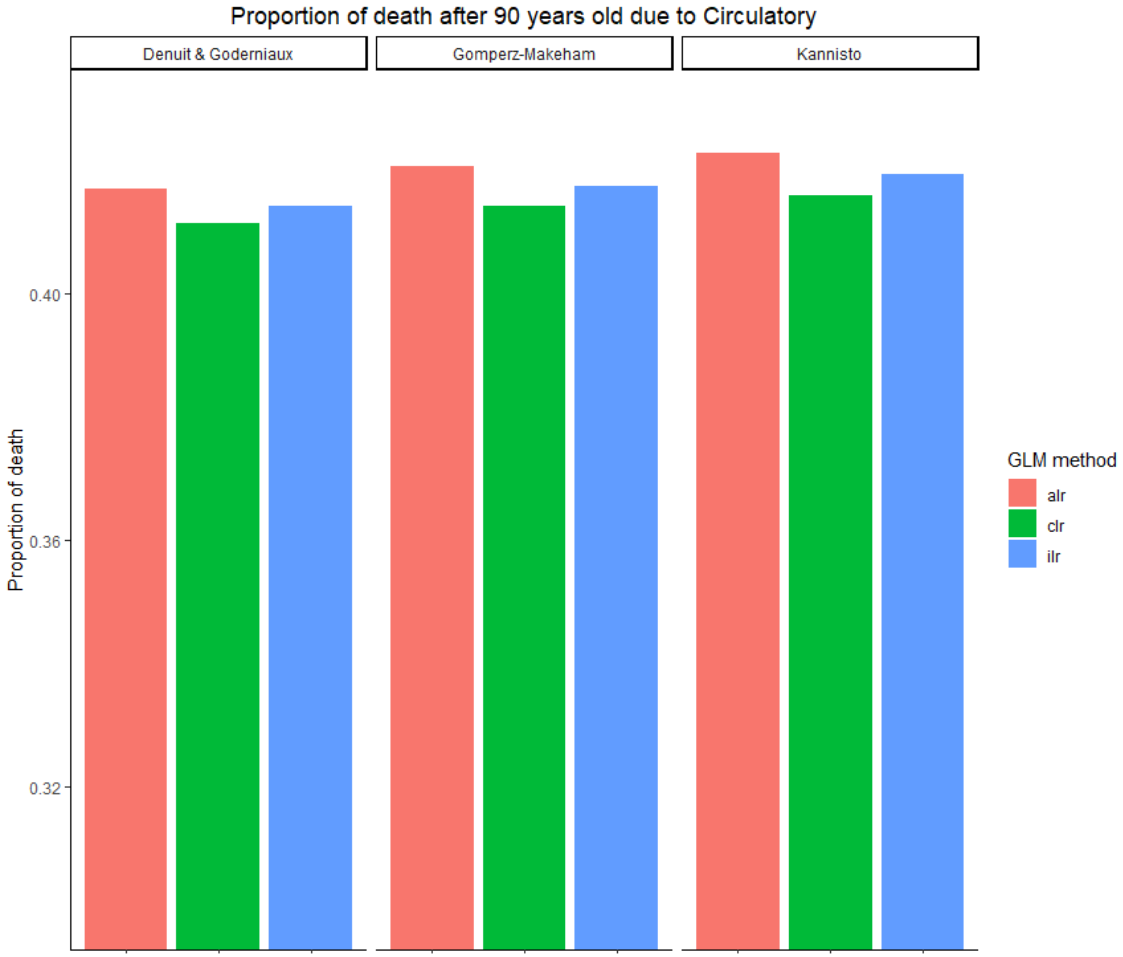

Figure – Circulatory proportion after 90 for all methods

SCOR The Art & Science of Risk | Life

Discussion

- We have experimented the cause specific life table closure with three different GLMs and completion methods.
- Cause specific extrapolation is consistent with the global mortality assumptions.
- The fact that likelihood is the same for the different models does not imply similar extrapolations.
- Only small differences between the methods have been observed. This is partly explained by the low amount of survivors after 90.
- A work in progress attempts to explain the differences in extrapolation between the three GLMs. In a different work, we extend this method to extrapolate the mortality by cause over time.

SCOR The Art & Science of Risk | Life

Thank you for your attention !