Coherent Cause-Specific Mortality Forecasting via Constrained Penalized Regression Models

Giancarlo Camarda
Institut National d'Etudes Démographiques, Aubervilliers, France

INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES

in collaboration with María Durbán (Universidad Carlos III de Madrid)

SCOR Chair on mortality research
Workshop 1

April 4-5, 2024

Motivation

- Overall mortality trends are the summation of cause-specific mortality experiences

Motivation

- Overall mortality trends are the summation of cause-specific mortality experiences

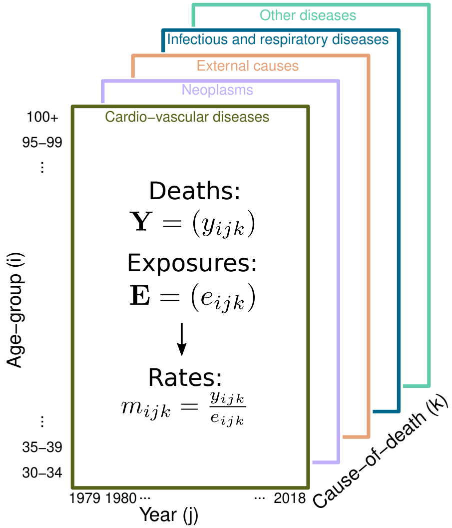
- Modelling and forecasting cause-specific mortality patterns allows us to
  - recognize drivers of all-cause mortality
  - identify emerging health challenges

Motivation

- Overall mortality trends are the summation of cause-specific mortality experiences
- Modelling and forecasting cause-specific mortality patterns allows us to
  - recognize drivers of all-cause mortality
  - identify emerging health challenges
- For long time it has been argued that all-cause mortality projections based on cause-specific mortality present serious drawbacks (Wilmoth, MPS, 1995)

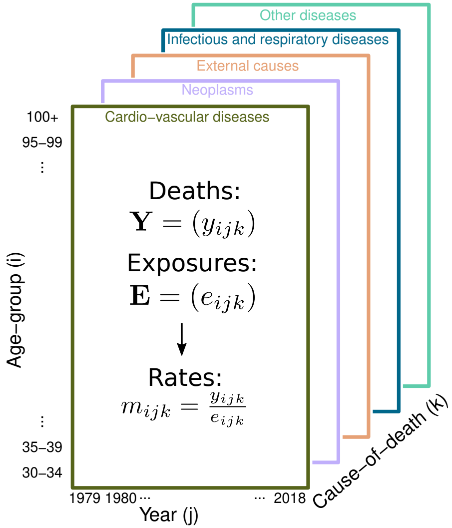
Motivation

- Overall mortality trends are the summation of cause-specific mortality experiences
- Modelling and forecasting cause-specific mortality patterns allows us to
  - recognize drivers of all-cause mortality
  - identify emerging health challenges
- For long time it has been argued that all-cause mortality projections based on cause-specific mortality present serious drawbacks (Wilmoth, MPS, 1995)
- Recent developments:
  - Lee-Carter model for specific cause (Kjærgaard et al., JRSS-C, 2019)
  - Bayesian hierarchical model for cause-specific death rates in geographic subunits (Foreman et al., JRSS-C, 2017)

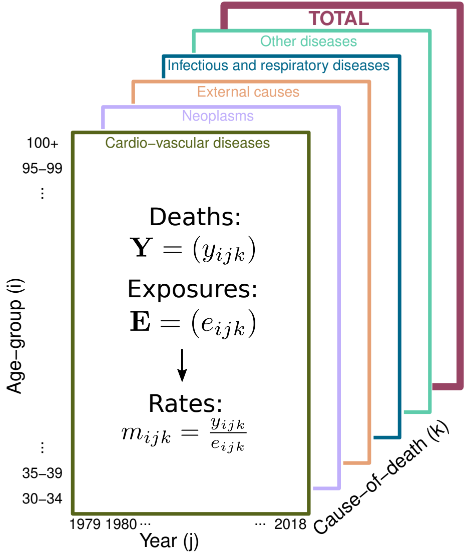


- Data are two 3-D arrays

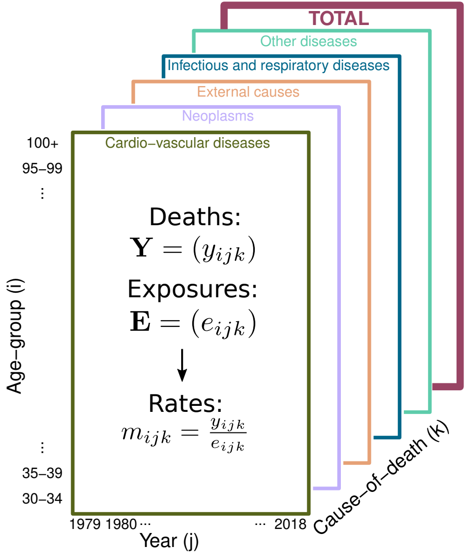
Data structure



- Data are two 3-D arrays
- Classified by age ($\boldsymbol{a}$), year ($\boldsymbol{t}$) and cause of death ($\boldsymbol{c}$)

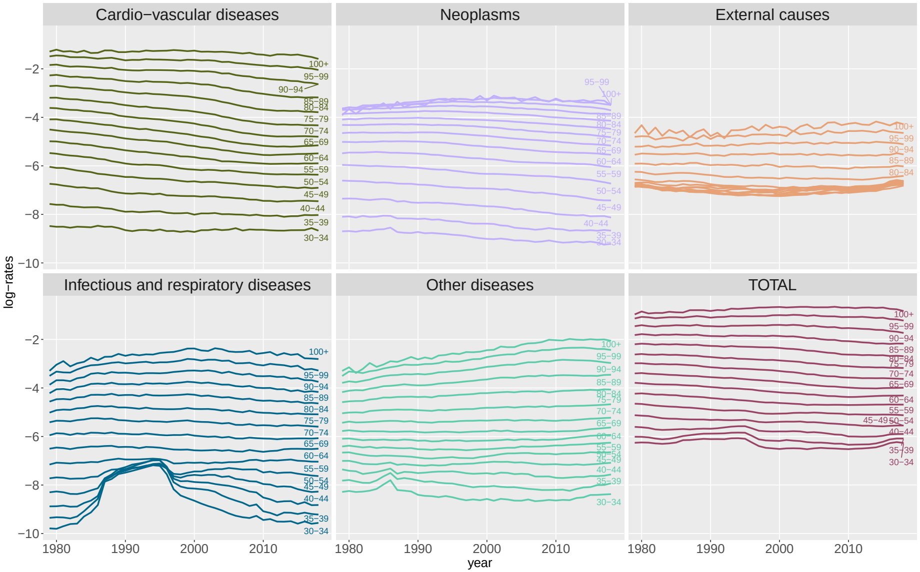


- Data are two 3-D arrays
- Classified by age ($\boldsymbol{a}$), year ($\boldsymbol{t}$) and cause of death ($\boldsymbol{c}$)
- Final layer of $\boldsymbol{Y}$ contains total number of deaths

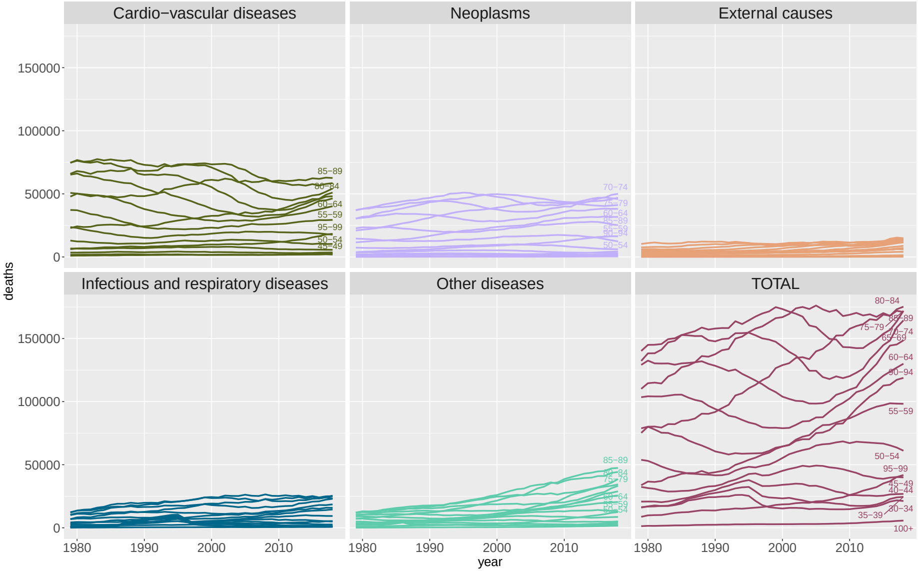


- Data are two 3-D arrays
- Classified by age ($\boldsymbol{a}$), year ($\boldsymbol{t}$) and cause of death ($\boldsymbol{c}$)
- Final layer of $\boldsymbol{Y}$ contains total number of deaths
- Each layer in $\boldsymbol{E}$ includes the same age-year matrix

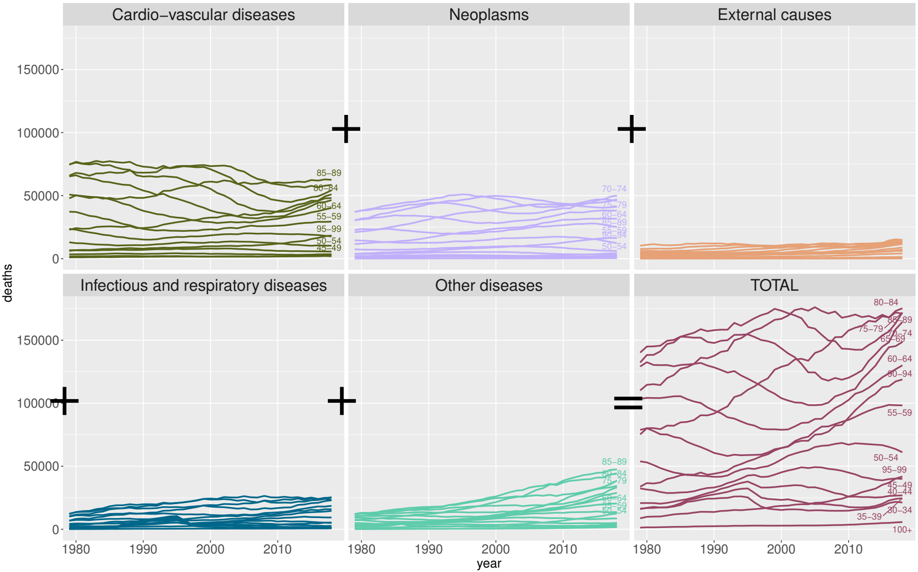
Example: USA mortality, males. Log-rates



Example: USA mortality, males. Deaths



Example: USA mortality, males. Deaths



The model

- Stochastic assumption:

$$y_{ijk} \sim \mathcal{P}(\mu_{ijk} e_{ij})$$

The model

- Stochastic assumption:

$$y_{ijk} \sim \mathcal{P}(\mu_{ijk} e_{ij})$$

- We model the linear predictor, vectorized:

$$\ln(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{B}\,\boldsymbol{\alpha}$$

where

$$\boldsymbol{B} = \boldsymbol{I}_k \otimes \boldsymbol{B}_t \otimes \boldsymbol{B}_a = \begin{bmatrix} \boldsymbol{B}_t \otimes \boldsymbol{B}_a & & & \\ & \boldsymbol{B}_t \otimes \boldsymbol{B}_a & & \\ & & \ddots & \\ & & & \boldsymbol{B}_t \otimes \boldsymbol{B}_a \end{bmatrix}$$

The model

- Stochastic assumption:

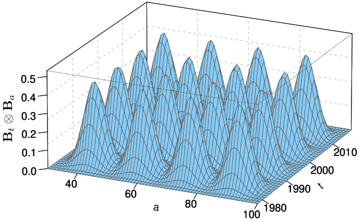
$$y_{ijk} \sim \mathcal{P}(\mu_{ijk} e_{ij})$$

- We model the linear predictor, vectorized:

$$\ln(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{B}\,\boldsymbol{\alpha}$$

where

$$\boldsymbol{B} = \boldsymbol{I}_k \otimes \boldsymbol{B}_t \otimes \boldsymbol{B}_a = \begin{bmatrix} \boldsymbol{B}_t \otimes \boldsymbol{B}_a & & & \\ & \boldsymbol{B}_t \otimes \boldsymbol{B}_a & & \\ & & \ddots & \\ & & & \boldsymbol{B}_t \otimes \boldsymbol{B}_a \end{bmatrix}$$



Adding constraints on death counts

- Additional constraints needed as the deaths-level:

$$\boldsymbol{C}\,\mathbb{E}[\boldsymbol{y}] = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{C}\,(\boldsymbol{e} * \boldsymbol{\mu}) = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B}\boldsymbol{\alpha})) = \mathbf{0}$$

  where $\boldsymbol{C}$ sums up cause-specific deaths and then subtract the associated total number of deaths

Adding constraints on death counts

- Additional constraints needed as the deaths-level:

$$\boldsymbol{C}\,\mathbb{E}[\boldsymbol{y}] = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{C}\,(\boldsymbol{e} * \boldsymbol{\mu}) = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B}\boldsymbol{\alpha})) = \boldsymbol{0}$$

  where $\boldsymbol{C}$ sums up cause-specific deaths and then subtract the associated total number of deaths

$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & -1 \end{bmatrix} \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_n = \boldsymbol{C}_c \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_n$$

Adding constraints on death counts

- Additional constraints needed as the deaths-level:

$$\boldsymbol{C}\,\mathbb{E}[\boldsymbol{y}] = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{C}\,(\boldsymbol{e} * \boldsymbol{\mu}) = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B}\boldsymbol{\alpha})) = \boldsymbol{0}$$

where $\boldsymbol{C}$ sums up cause-specific deaths and then subtract the associated total number of deaths

$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & -1 \end{bmatrix} \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_n = \boldsymbol{C}_c \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_n$$

- $\boldsymbol{C}$ can be written as a Kronecker product

Adding constraints on death counts

- Additional constraints needed as the deaths-level:

$$\boldsymbol{C}\,\mathbb{E}[\boldsymbol{y}] = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{C}\,(\boldsymbol{e} * \boldsymbol{\mu}) = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B\alpha})) = \boldsymbol{0}$$

where $\boldsymbol{C}$ sums up cause-specific deaths and then subtract the associated total number of deaths

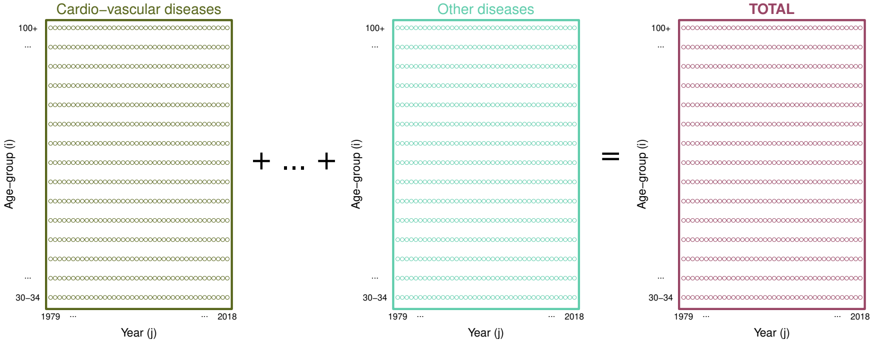
$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & -1 \end{bmatrix} \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_n = \boldsymbol{C}_c \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_n$$

- $\boldsymbol{C}$ can be written as a Kronecker product

- $\Rightarrow$ All inner operations can be embedded in a Generalized Linear Array Model (GLAM) framework (Currie et al., JRSS-B, 2006)

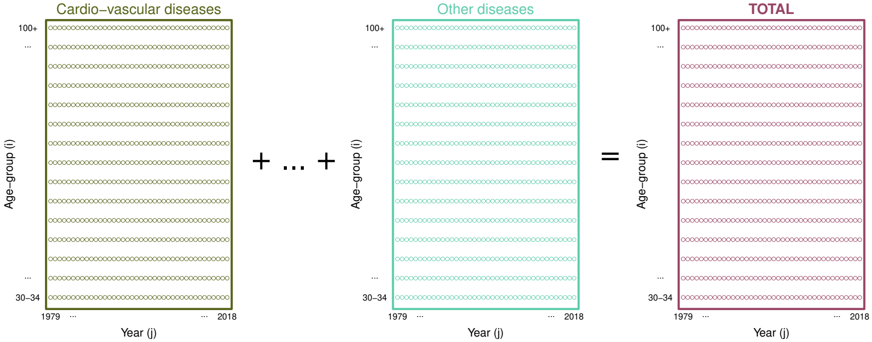
About constraints and singularity

- Many coefficients + all possible age-time constraints
  $\Rightarrow$ singular system



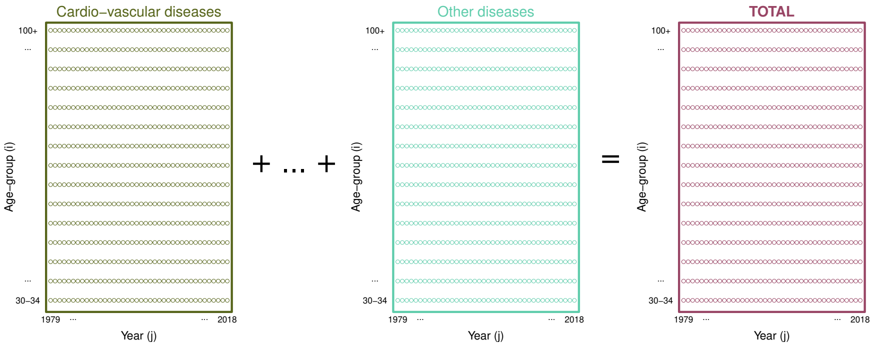
About constraints and singularity

- Many coefficients + all possible age-time constraints
  $\Rightarrow$ singular system
- A compromise between # of coefficients and # of constraints



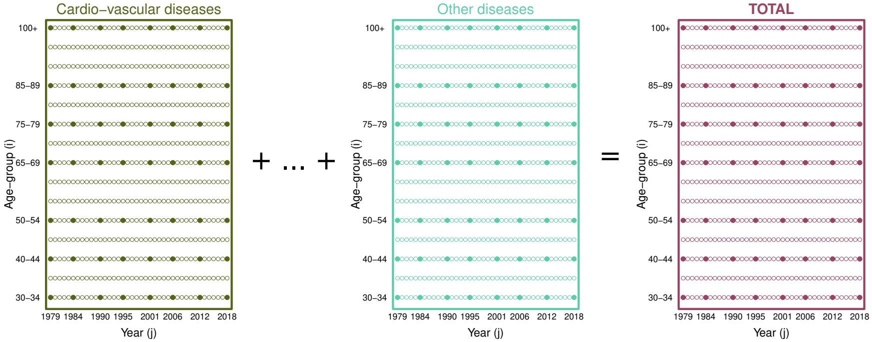
About constraints and singularity

- Many coefficients + all possible age-time constraints
  $\Rightarrow$ singular system
- A compromise between # of coefficients and # of constraints
- Rich bases in $\boldsymbol{B}_a$ and $\boldsymbol{B}_t$ to ensure a good data-description



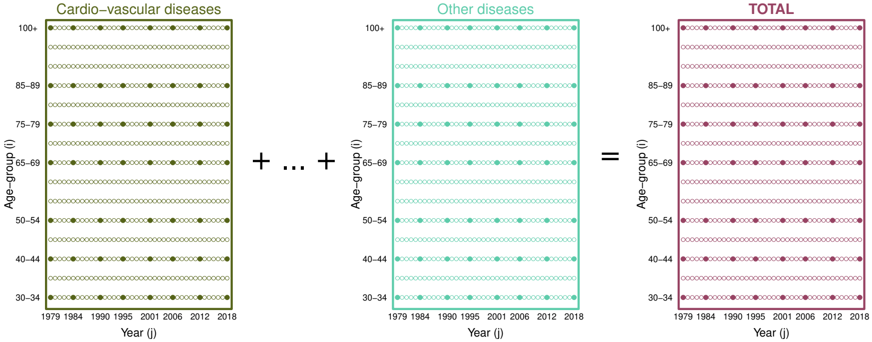
About constraints and singularity

- Many coefficients + all possible age-time constraints
  $\Rightarrow$ singular system
- A compromise between # of coefficients and # of constraints
- Rich bases in $\boldsymbol{B}_a$ and $\boldsymbol{B}_t$ to ensure a good data-description
- Enforce constraints for a number of equally-spaced data-points



About constraints and singularity

- Many coefficients + all possible age-time constraints
  $\Rightarrow$ singular system
- A compromise between # of coefficients and # of constraints
- Rich bases in $\boldsymbol{B}_a$ and $\boldsymbol{B}_t$ to ensure a good data-description
- Enforce constraints for a number of equally-spaced data-points
- $\boldsymbol{C} = \cancel{\boldsymbol{C}_c \otimes \boldsymbol{I}_m \otimes \boldsymbol{I}_n} = \boldsymbol{C}_c \otimes \boldsymbol{C}_t \otimes \boldsymbol{C}_a$



Likelihood function

- We maximize a Poisson likelihood:

$$\ell = \boldsymbol{y}' \boldsymbol{B}\boldsymbol{\alpha} - \boldsymbol{e}' \exp(\boldsymbol{B}\boldsymbol{\alpha})$$

Likelihood function

- We maximize a penalized Poisson likelihood:

$$\ell_P = \boldsymbol{y}'\boldsymbol{B}\boldsymbol{\alpha} - \boldsymbol{e}' \exp(\boldsymbol{B}\boldsymbol{\alpha}) - \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{P}\boldsymbol{\alpha}$$

- Smoothness over age and time for each cause is ensured by $\boldsymbol{P}$ (block-diagonal structure)

- We maximize a penalized Poisson likelihood:

$$\ell_P = \boldsymbol{y}'\boldsymbol{B}\boldsymbol{\alpha} - \boldsymbol{e}'\exp(\boldsymbol{B}\boldsymbol{\alpha}) - \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{P}\boldsymbol{\alpha}$$

- Smoothness over age and time for each cause is ensured by $\boldsymbol{P}$ (block-diagonal structure)
- We use smoothing parameters optimized by Bayesian Information Criterion (BIC) when estimating each cause-specific age-time matrix independently

Likelihood function

- We maximize a constrained penalized Poisson likelihood:

$$\ell_P = \boldsymbol{y}'\boldsymbol{B}\boldsymbol{\alpha} - \boldsymbol{e}'\exp(\boldsymbol{B}\boldsymbol{\alpha}) - \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{P}\boldsymbol{\alpha} - \boldsymbol{\omega}'\boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B}\boldsymbol{\alpha}))$$

- Smoothness over age and time for each cause is ensured by $\boldsymbol{P}$ (block-diagonal structure)
- We use smoothing parameters optimized by Bayesian Information Criterion (BIC) when estimating each cause-specific age-time matrix independently
- Constraints are enforced by Lagrange multipliers $\boldsymbol{\omega}$

Estimation procedure (for equation enthusiasts)

- To maximize

$$\ell_P = \boldsymbol{y}'\boldsymbol{B\alpha} - \boldsymbol{e}'\exp(\boldsymbol{B\alpha}) - \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{P\alpha} - \boldsymbol{\omega}'\boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B\alpha}))$$

Estimation procedure (for equation enthusiasts)

- To maximize

$$\ell_P = \boldsymbol{y}'\boldsymbol{B}\boldsymbol{\alpha} - \boldsymbol{e}'\exp(\boldsymbol{B}\boldsymbol{\alpha}) - \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{P}\boldsymbol{\alpha} - \boldsymbol{\omega}'\boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B}\boldsymbol{\alpha}))$$

- We derive the following scoring algorithm:

$$\begin{bmatrix} \boldsymbol{B}'\tilde{\boldsymbol{W}}\boldsymbol{B} + \boldsymbol{P} + \boldsymbol{B}'\mathtt{diag}(\boldsymbol{C}'\tilde{\omega})\tilde{\boldsymbol{V}}\boldsymbol{B} & \boldsymbol{B}'\tilde{\boldsymbol{V}}\boldsymbol{C}' \\ \boldsymbol{C}\tilde{\boldsymbol{V}}\boldsymbol{B} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\alpha}} \\ \tilde{\omega} \end{bmatrix} =$$

$$\begin{bmatrix} \boldsymbol{B}'\tilde{\boldsymbol{W}}\tilde{\boldsymbol{z}} + \boldsymbol{B}'\mathtt{diag}(\boldsymbol{C}'\tilde{\omega})\tilde{\boldsymbol{V}}\boldsymbol{B}\tilde{\alpha} \\ \boldsymbol{C}\tilde{\boldsymbol{V}}\boldsymbol{B}\tilde{\alpha} - \boldsymbol{C}\gamma \end{bmatrix}$$

where

- $\gamma = \exp\boldsymbol{\eta}$, $\boldsymbol{V} = \mathtt{diag}(\gamma)$
- $\boldsymbol{W}$ and $\boldsymbol{z}$ Poisson regression weights and working response

- To maximize

$$\ell_P = \boldsymbol{y}'\boldsymbol{B}\boldsymbol{\alpha} - \boldsymbol{e}'\exp(\boldsymbol{B}\boldsymbol{\alpha}) - \frac{1}{2}\boldsymbol{\alpha}'\boldsymbol{P}\boldsymbol{\alpha} - \omega'\boldsymbol{C}(\boldsymbol{e} * \exp(\boldsymbol{B}\alpha))$$

- We derive the following scoring algorithm:

$$\begin{bmatrix} \boldsymbol{B}'\tilde{\boldsymbol{W}}\boldsymbol{B} + \boldsymbol{P} + \boldsymbol{B}'\mathtt{diag}(\boldsymbol{C}'\tilde{\omega})\tilde{\boldsymbol{V}}\boldsymbol{B} & \boldsymbol{B}'\tilde{\boldsymbol{V}}\boldsymbol{C}' \\ \boldsymbol{C}\tilde{\boldsymbol{V}}\boldsymbol{B} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\alpha}} \\ \tilde{\omega} \end{bmatrix} =$$

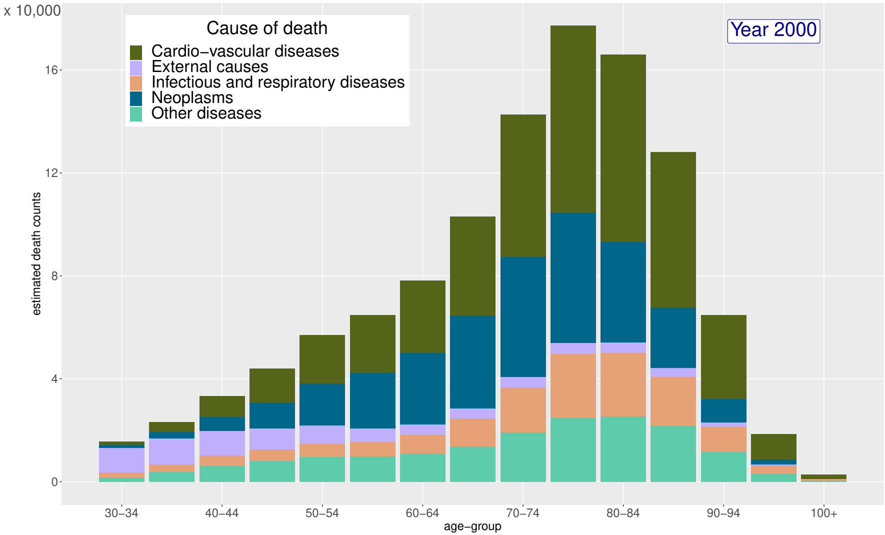
$$\begin{bmatrix} \boldsymbol{B}'\tilde{\boldsymbol{W}}\tilde{\boldsymbol{z}} + \boldsymbol{B}'\mathtt{diag}(\boldsymbol{C}'\tilde{\omega})\tilde{\boldsymbol{V}}\boldsymbol{B}\tilde{\alpha} \\ \boldsymbol{C}\tilde{\boldsymbol{V}}\boldsymbol{B}\tilde{\alpha} - \boldsymbol{C}\gamma \end{bmatrix}$$

where

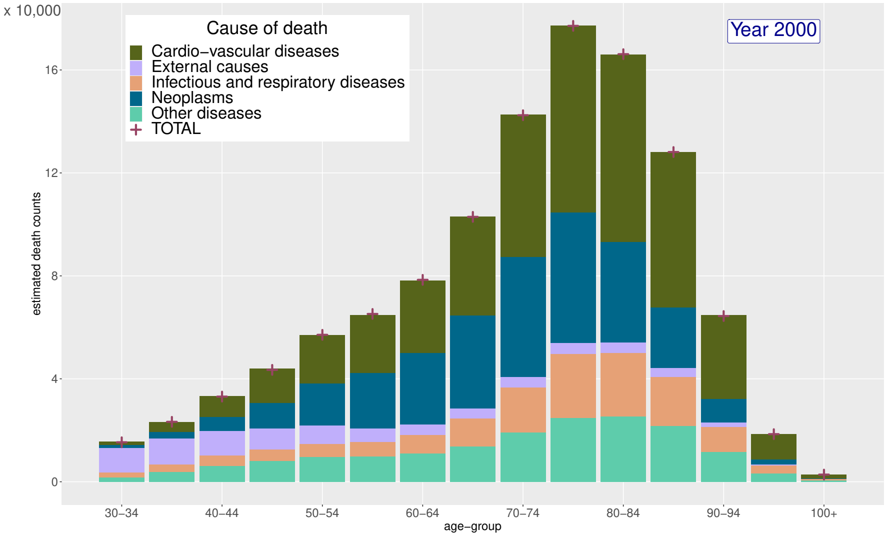
- $\gamma = \exp\boldsymbol{\eta}$, $\boldsymbol{V} = \mathtt{diag}(\gamma)$
- $\boldsymbol{W}$ and $\boldsymbol{z}$ Poisson regression weights and working response

- With no summation constraint the model simply reduces to a series of two-dimensional GLAMs

Checking constraints



Checking constraints

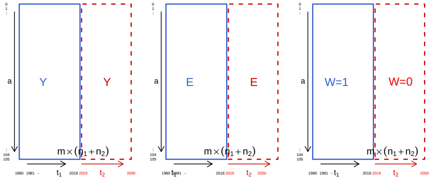


Forecasting mortality for a single cause

- Future is a missing value problem (Currie et al., StatMod, 2004)

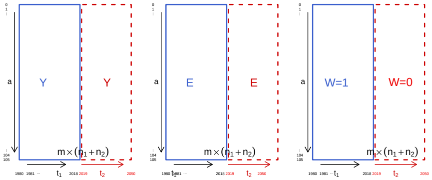
Forecasting mortality for a single cause

- Future is a missing value problem (Currie et al., StatMod, 2004)
- Data are augmented with arbitrary values for future years



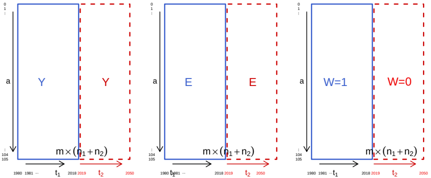
Forecasting mortality for a single cause

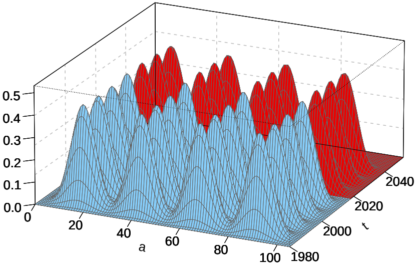
- Future is a missing value problem (Currie et al., StatMod, 2004)
- Data are augmented with arbitrary values for future years
- Define a 0/1 weight matrix ($\boldsymbol{W}^A$)



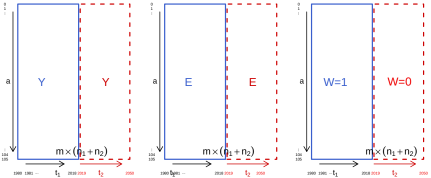
Forecasting mortality for a single cause

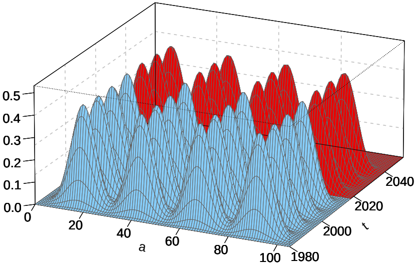
- Future is a missing value problem (Currie et al., StatMod, 2004)
- Data are augmented with arbitrary values for future years
- Define a 0/1 weight matrix ($\boldsymbol{W}^A$)
- B-spline bases are augmented too: $\boldsymbol{B} = [\boldsymbol{B}_{t_1} : \boldsymbol{B}_{t_2}] \otimes \boldsymbol{B}_a$





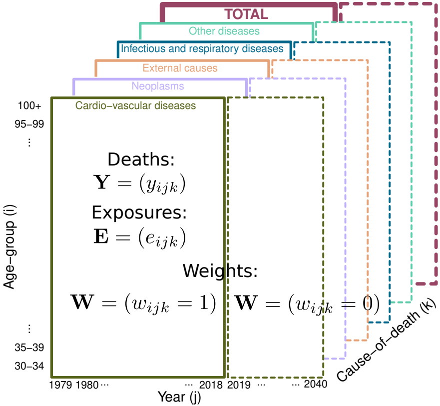
- Future is a missing value problem (Currie et al., StatMod, 2004)
- Data are augmented with arbitrary values for future years
- Define a 0/1 weight matrix ($\boldsymbol{W}^A$)
- B-spline bases are augmented too: $\boldsymbol{B} = [\boldsymbol{B}_{t_1} : \boldsymbol{B}_{t_2}] \otimes \boldsymbol{B}_a$
- We add 0/1 weights in the scoring algorithm



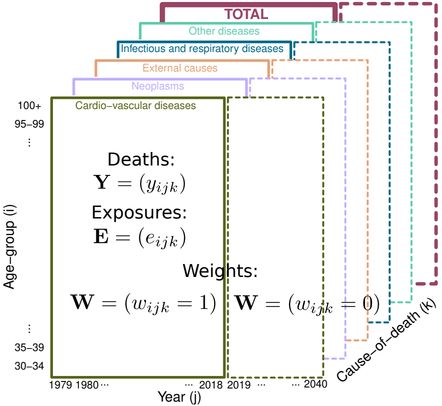


Forecasting mortality for multiple cause

- We augment both deaths and exposures
- We define a 0/1 weight 3D array
- "Arbitrary" values for future years *must* comply constraints



- We augment both deaths and exposures
- We define a 0/1 weight 3D array
- "Arbitrary" values for future years *must* comply constraints

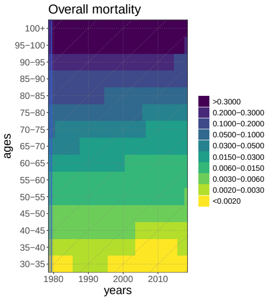


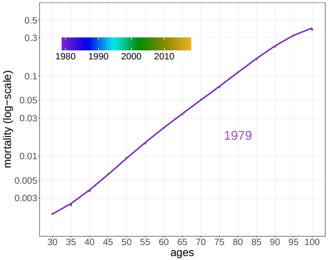
- We generalize in 3D the expanded B-spline bases:

$$\boldsymbol{B} = \boldsymbol{I}_k \otimes [\boldsymbol{B}_{t_1} : \boldsymbol{B}_{t_2}] \otimes \boldsymbol{B}_a = \begin{bmatrix} [\boldsymbol{B}_{t_1} : \boldsymbol{B}_{t_2}] \otimes \boldsymbol{B}_a & & \\ & \ddots & \\ & & [\boldsymbol{B}_{t_1} : \boldsymbol{B}_{t_2}] \otimes \boldsymbol{B}_a \end{bmatrix}$$

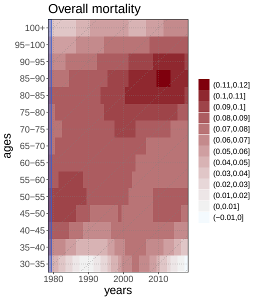
Observing mortality patterns (over ages)

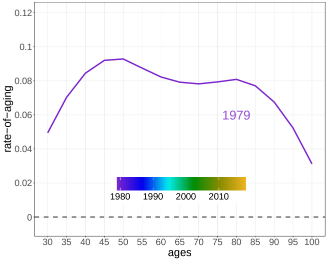
Actual and Smooth log-mortality





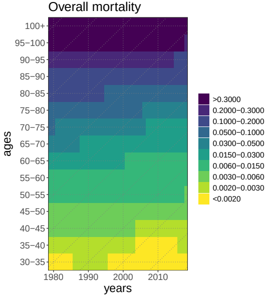
Rate-of-aging from smooth mortality

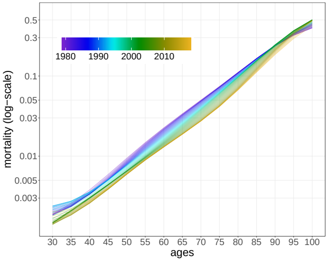




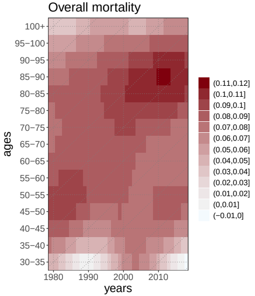
Observing mortality patterns (over ages)

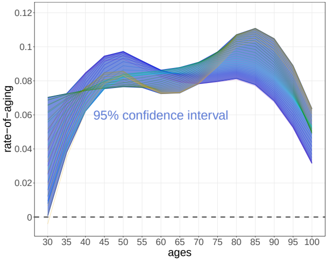
Actual and Smooth log-mortality





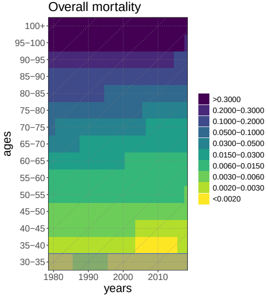
Rate-of-aging from smooth mortality

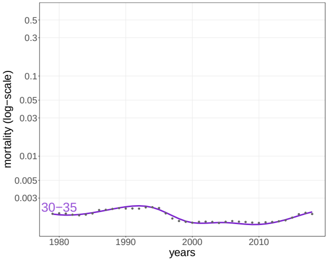




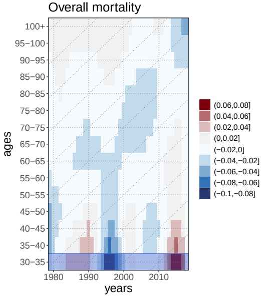
Observing mortality patterns (over years)

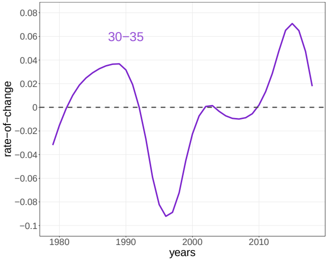
Actual and Smooth log-mortality





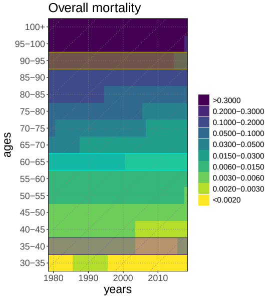
Rate-of-change from smooth mortality

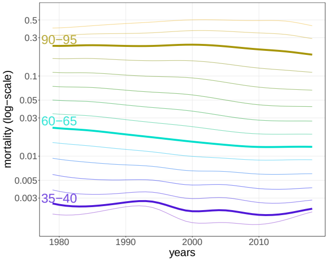




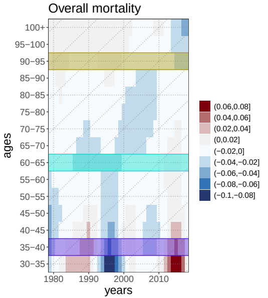
Observing mortality patterns (over years)

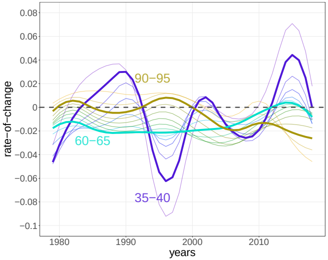
Actual and Smooth log-mortality



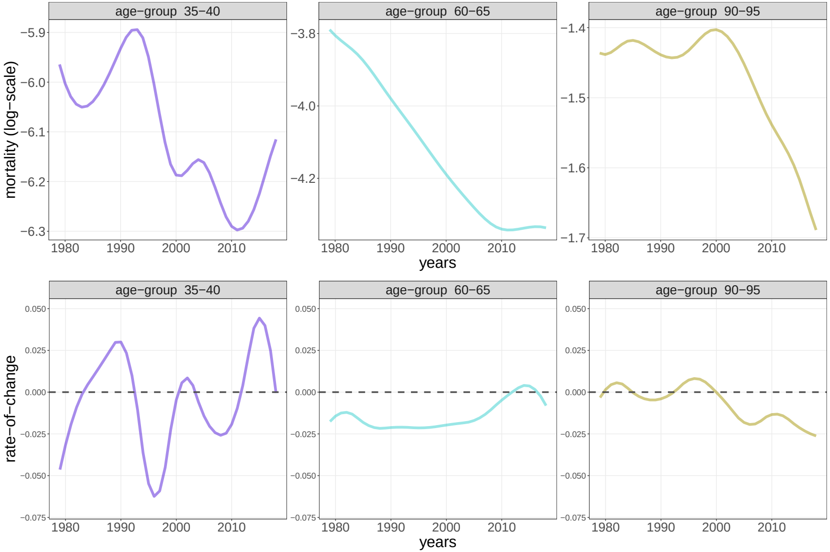


Rate-of-change from smooth mortality

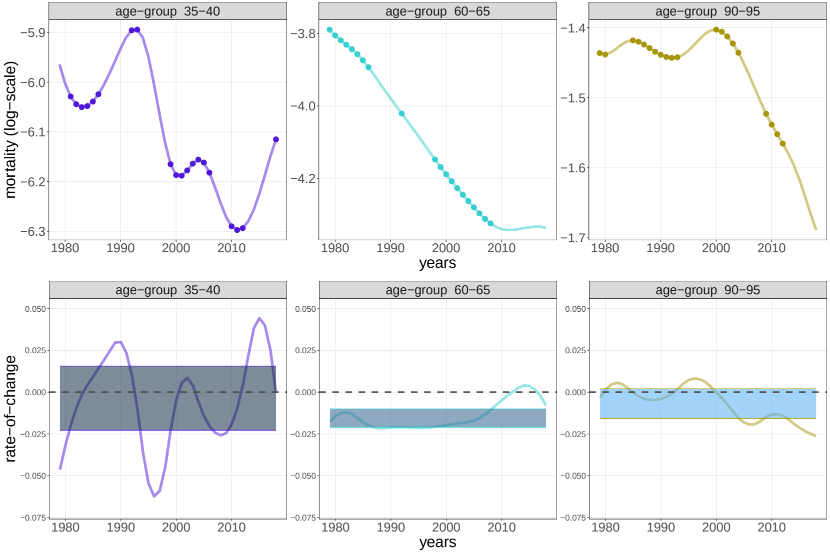




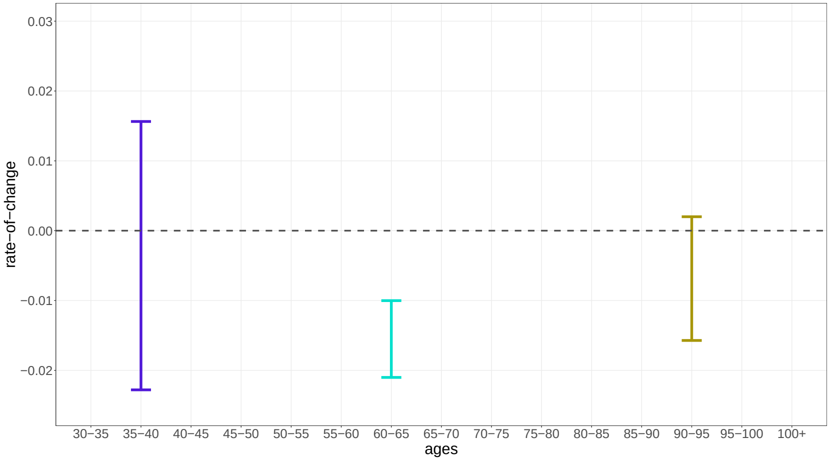
Rate-of-change, selected ages



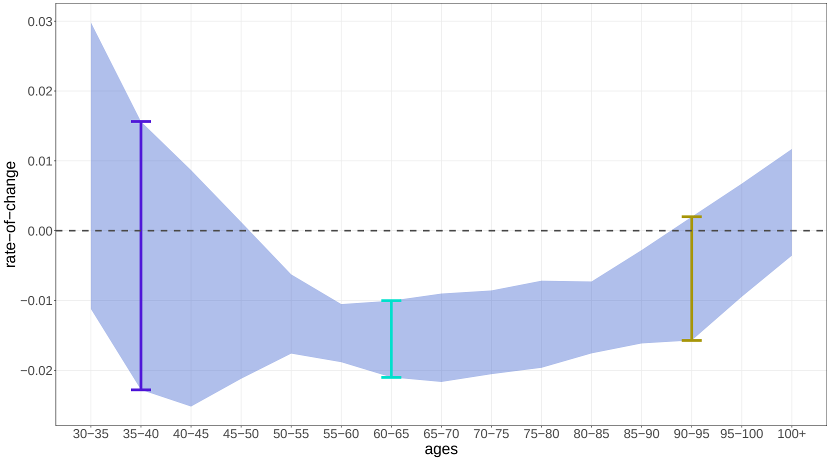
Rate-of-change, selected ages

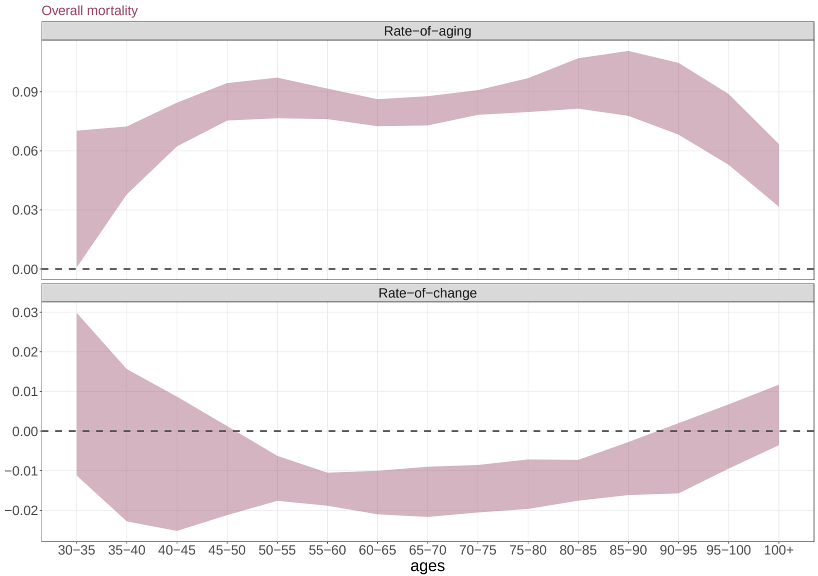


Rate-of-change, selected ages

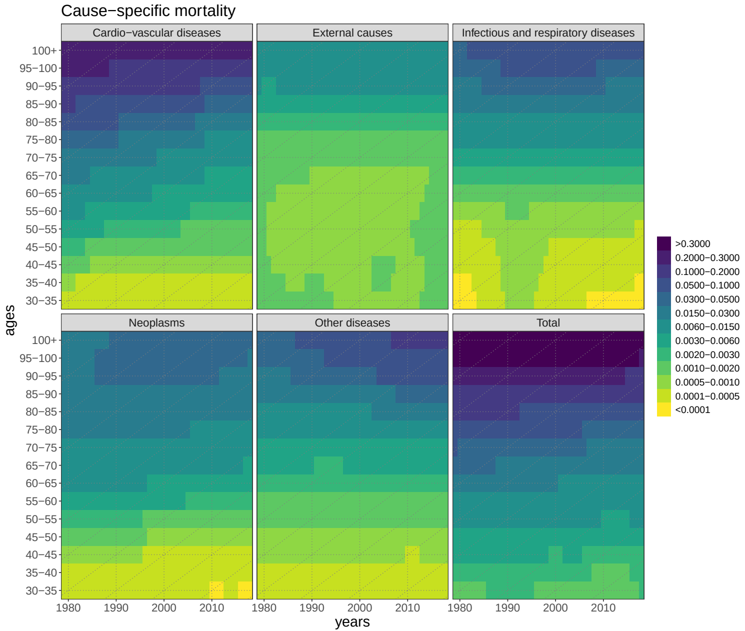


Rate-of-change, all ages

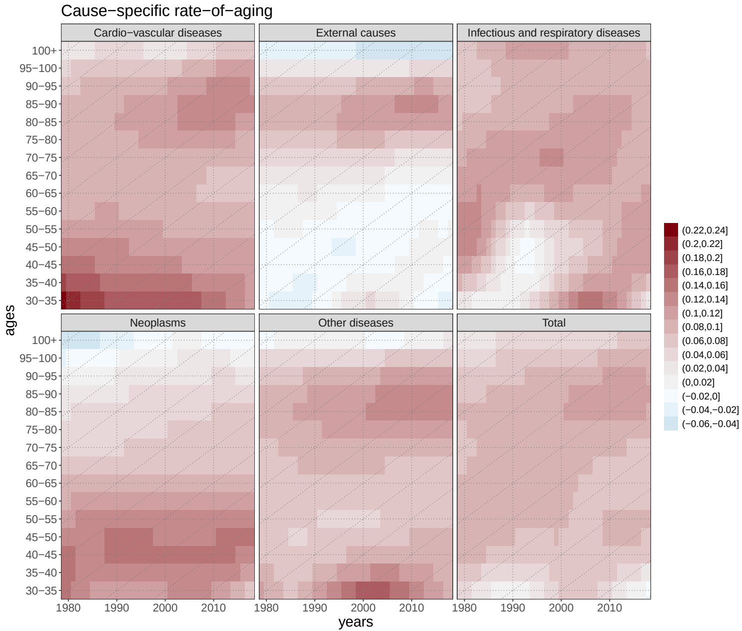



Overall mortality
Rate−of−aging
0.09
0.06
0.03
0.00
Rate−of−change
0.03
0.02
0.01
0.00
−0.01
−0.02
30−35
35−40
40−45
45−50
50−55
55−60
60−65
65−70
70−75
75−80
80−85
85−90
90−95
95−100
100+
ages

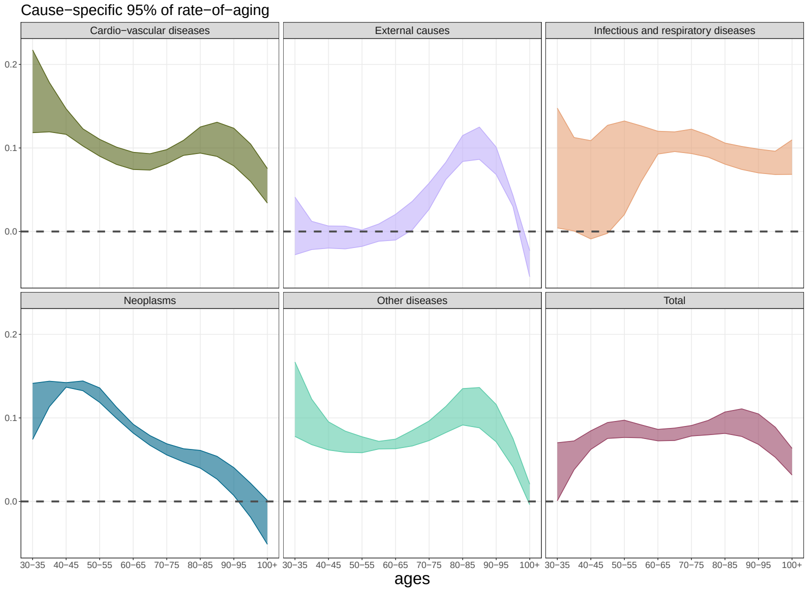
Looking at cause-specific mortality



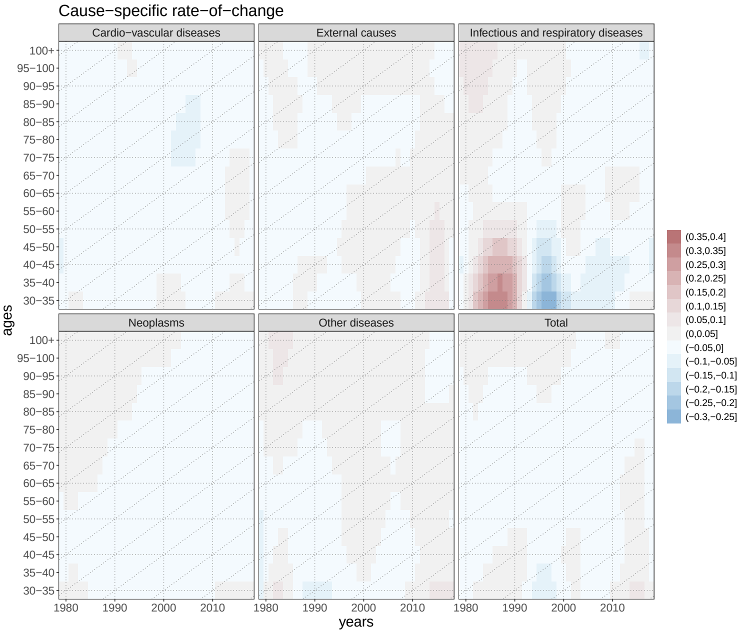
Looking at cause-specific mortality



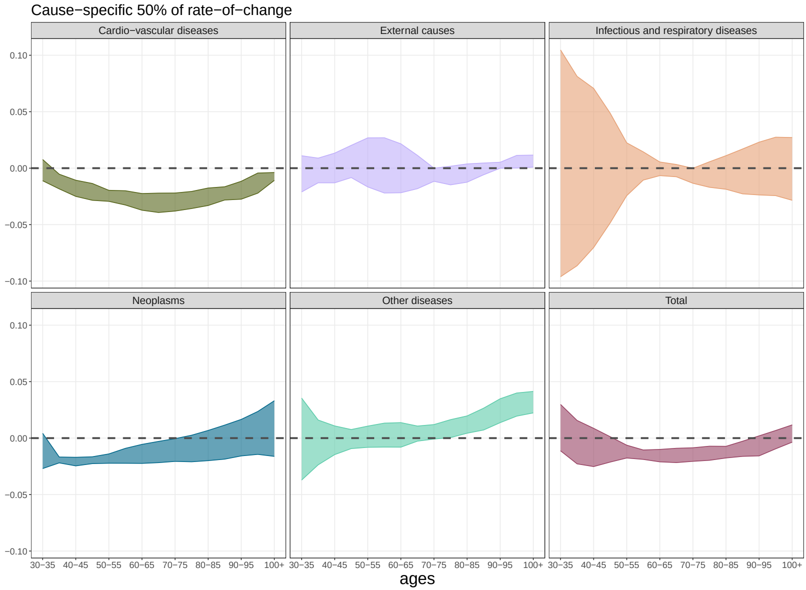
Looking at cause-specific mortality



Looking at cause-specific mortality



Looking at cause-specific mortality



Working on the shape

- Disregarding information about shapes seems unreasonable

Working on the shape

- Disregarding information about shapes seems unreasonable
- Future mortality *must* follow data-driven age-profiles and rate-of-change

Working on the shape

- Disregarding information about shapes seems unreasonable
- Future mortality *must* follow data-driven age-profiles and rate-of-change
- We constrain derivatives of future mortality to lay within certain confidence intervals of observed derivatives

Working on the shape

- Disregarding information about shapes seems unreasonable
- Future mortality *must* follow data-driven age-profiles and rate-of-change
- We constrain derivatives of future mortality to lay within certain confidence intervals of observed derivatives
- Asymmetric penalties are employed for this purpose:
    - within each iteration, whenever current estimations present derivatives (in future years) out of the desired intervals, a penalty intervenes

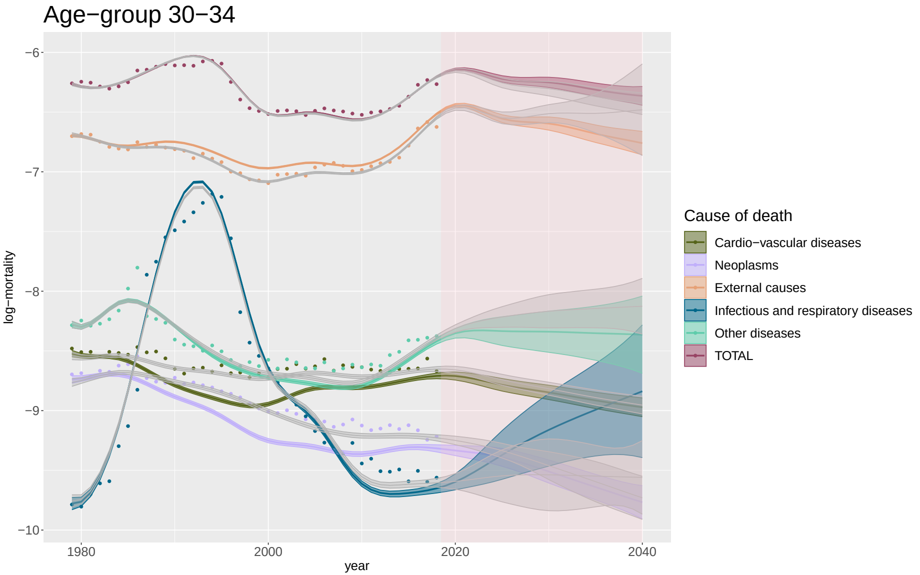
Working on the shape

- Disregarding information about shapes seems unreasonable
- Future mortality *must* follow data-driven age-profiles and rate-of-change
- We constrain derivatives of future mortality to lay within certain confidence intervals of observed derivatives
- Asymmetric penalties are employed for this purpose:
  - within each iteration, whenever current estimations present derivatives (in future years) out of the desired intervals, a penalty intervenes
- We use *CP*-splines (Camarda, DemRes, 2019)

Working on the shape

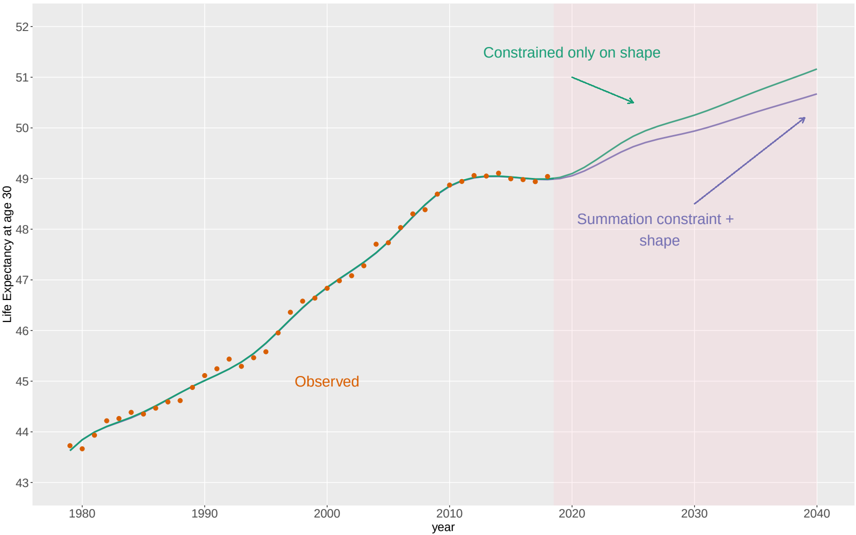
- Disregarding information about shapes seems unreasonable
- Future mortality *must* follow data-driven age-profiles and rate-of-change
- We constrain derivatives of future mortality to lay within certain confidence intervals of observed derivatives
- Asymmetric penalties are employed for this purpose:
  - within each iteration, whenever current estimations present derivatives (in future years) out of the desired intervals, a penalty intervenes
- We use *CP*-splines (Camarda, DemRes, 2019)
- We generalize *CP*-splines in a 3D setting (age+time+cause)

Outcomes: log-Rates

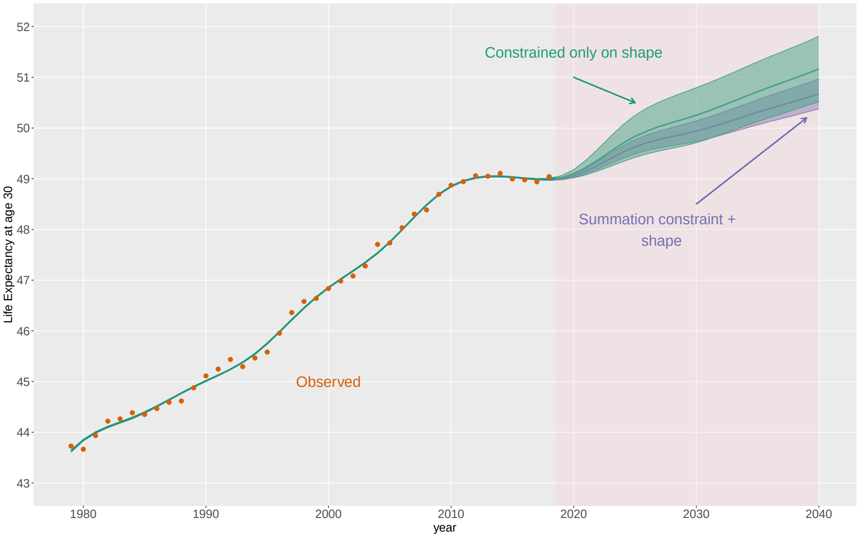


Grey lines and areas depict unconstrained model

Life expectancy at age 30 for overall mortality



Life expectancy at age 30 for overall mortality



Concluding remarks

- Our proposed approach smooths and forecasts cause-specific and overall mortality simultaneously

Concluding remarks

- Our proposed approach smooths and forecasts cause-specific and overall mortality simultaneously

- Non-linear summation constraints at the level death counts are incorporated by Lagrange multipliers

Concluding remarks

- Our proposed approach smooths and forecasts cause-specific and overall mortality simultaneously
- Non-linear summation constraints at the level death counts are incorporated by Lagrange multipliers
- Scoring algorithm is derived and embedded in a GLAM framework

Concluding remarks

- Our proposed approach smooths and forecasts cause-specific and overall mortality simultaneously
- Non-linear summation constraints at the level death counts are incorporated by Lagrange multipliers
- Scoring algorithm is derived and embedded in a GLAM framework
- For US males: More pessimistic future trends with narrower confidence intervals

Concluding remarks

- Our proposed approach smooths and forecasts cause-specific and overall mortality simultaneously
- Non-linear summation constraints at the level death counts are incorporated by Lagrange multipliers
- Scoring algorithm is derived and embedded in a GLAM framework
- For US males: More pessimistic future trends with narrower confidence intervals
- Future research:
  - Selection of the $2 \times (k+1)$ smoothing parameters
  - Back-testing for model performance
  - Extension to female/male and regional mortality settings

Giancarlo Camarda
in collaboration with María Durbán

Coherent Cause-Specific Mortality Forecasting via Constrained Penalized Regression Models

Thanks for your attention.
Comments and questions?

Main bibliography

- Camarda, C.G. (2019). Smooth Constrained Mortality Forecasting. *Demographic Research*
- Camarda, C.G. and Durban, M. (2023). Coherent cause-specific mortality forecasting via constrained penalized regression models. *Proc. International Workshop Stat Mod*
- Currie, I.D., Durban, M. and Eilers, P.H.C. (2004). Smoothing and Forecasting Mortality Rates. *Statistical Modelling*
- Currie, I.D., Durban, M. and Eilers, P.H.C. (2006). Generalized Linear Array Models with Applications to Multidimensional Smoothing. *Journal of Royal Statistical Society. Series B*
- Foreman, K.J., Li, G., Best, N. and Ezzati, M. (2017). Small Area Forecasts of Cause-Specific Mortality: Application of a Bayesian Hierarchical Model to US Vital Registration Data. *Journal of the Royal Statistical Society: Series C*
- Kjærgaard, S., Ergemen, Y.E., Kallestrup-Lamb, M. et al. (2019). Forecasting causes of death by using compositional data analysis: the case of cancer deaths. *Journal of the Royal Statistical Society. Series C*
- Wilmoth, J.R. (1995). Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies*