Estimating and modelling mortality rates in the absence of population denominators

Andrés M. Villegas and Munir Hiabu

UNSW Business School | AGSM

Longevity 17 Conference
Monday 12th and Tuesday 13th September 2022

Agenda

- Motivation
- Calculating the (forward) mortality rate (the usual way – if population data is available)
- The backward mortality rate
    - ▷ From backward mortality rate to forward mortality rate
    - ▷ Modelling the backward mortality rate
- Illustration using HMD data for England and Wales males
- Conclusion

Motivation

- Mortality rate (heuristic):

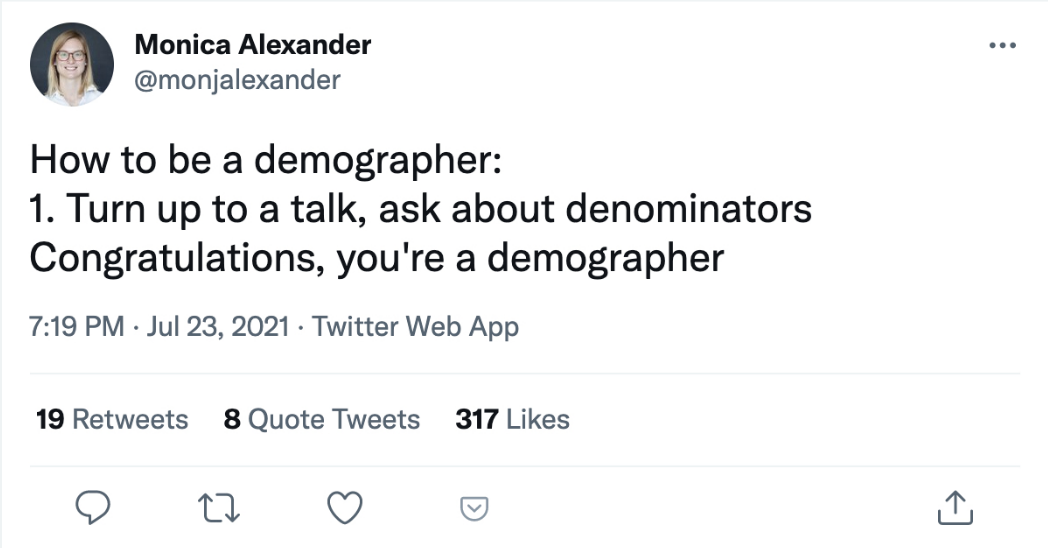
$$\frac{\#\text{ occurred deaths}}{\text{population size}}$$

- Problem: Denominator often poor quality or not known at all.
- Population size of older age groups (85+) often not well known or unavailable.

Motivation

Demographers know the struggle with the denominator

Motivation



Monica Alexander
@monjalexander

How to be a demographer:
1. Turn up to a talk, ask about denominators
Congratulations, you're a demographer

7:19 PM · Jul 23, 2021 · Twitter Web App

19 Retweets **8** Quote Tweets **317** Likes

Motivation - Quality of population data is sometimes doubtful

Colombia's population was overestimated by 5 million: Instead of the projected 50 millions population expected in 2018 in the Census 2005 projections, the population in 2018 was 45.5 million

Motivation - Imprecise population estimates at older ages

The Human Mortality Database overview says:

> *"Age exaggeration, on the other hand, is a more insidious problem. Our approach is guided by the conventional wisdom that* ***age reporting in death registration systems is typically more reliable than in census counts or official population estimates. For this reason, we derive population estimates at older ages from the death counts themselves****, employing extinct cohort methods. Such methods eliminate some, but certainly not all, of the biases in old-age mortality estimates due to age exaggeration."*

Calculating the mortality rate (the usual way – if population data is available)

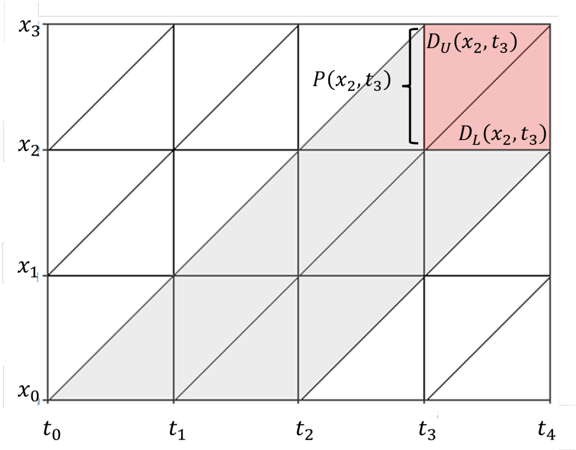
Calculating the mortality rate (the usual way – if population data is available)

We are interested in estimating the (forward) mortality rate

$$\mu(x|t) = \lim_{h \downarrow 0} h^{-1} \mathrm{Pr}\{\underbrace{X \in [x, x+h) |\ X \geq x}_{\substack{\text{Die in the next instant} \\ \text{given survival to age } x}}, T - X = t - x\}.$$

- X is age of death
- T is date (also called period) of death
- $C = T - X$ is cohort; known before death

Calculating the mortality rate – The Lexis diagram



$$D_U(x_j,t_k) = \sum_i I\{X_i \in [x_j, x_{j+1}),\, T_i - X_i \in [t_k - x_{j+1}, t_k - x_j)\}$$

$$D_L(x_j,t_k) = \sum_i I\{X_i \in [x_j, x_{j+1}),\, T_i - X_i \in [t_k - x_j, t_{k+1} - x_j)\}$$

$$P(x_j,t_k) = \sum_i I\{T_i > t_k,\, T_i - X_i \in [t_k - x_{j+1}, t_k - x_j)$$

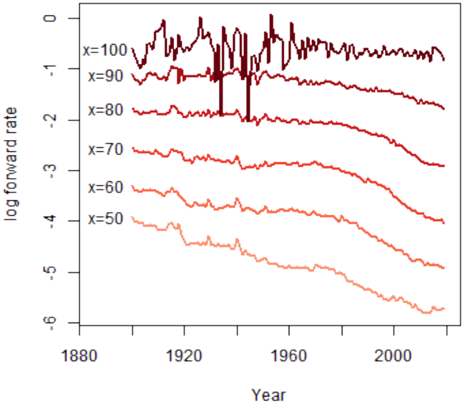
Calculating the mortality rate (the usual way – if population data is available) – central mortality rate

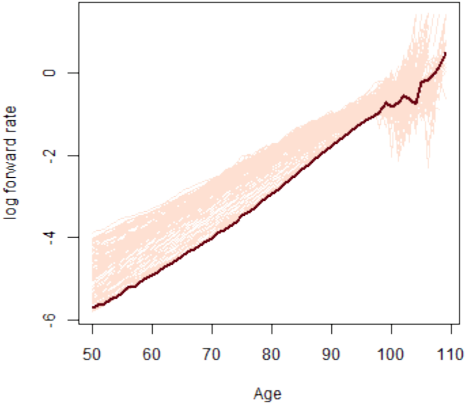
The central mortality rate is defined as

$$m(x_j|t_k) = \frac{D(x_j, t_k)}{E(x_j, t_k)} = \frac{D(x_j, t_k)}{\frac{1}{2}\{P(x_j, t_k) + P(x_j, t_{k+1})\} + \frac{1}{3}\{D_L(x_j, t_k) - D_U(x_j, t_k)\}}.$$

Under appropriate assumptions, $m(x_j|t_k)$ is an unbiased estimator of the expected value of $\mu(X|T)$ for (X, T) conditioned on the square $[x_j, x_j + 1) \times [t_k, t_k + 1)$.

The central mortality rate (England and Wales males)


Time trends


Age profile in 2019

Back to the original question:

Can we estimate the mortality rate without using information on the size of the population P?

- Maybe.
- Let's look at the backward mortality first.

The backward mortality rate

We aim to estimate

$$\mu^R(x|t) = \lim_{h\downarrow 0} h^{-1}\Pr\{\underbrace{X \in (x-h, x]|\ X \leq x,}_{\substack{\text{Die in the previous instant}\\ \text{given dead by age } x}}\ C = c\}.$$

The backward central mortality rate is given as

$$m^R(x_j|t_k) = \frac{D(x_j, t_k)}{E^R(x_j, t_k)}.$$

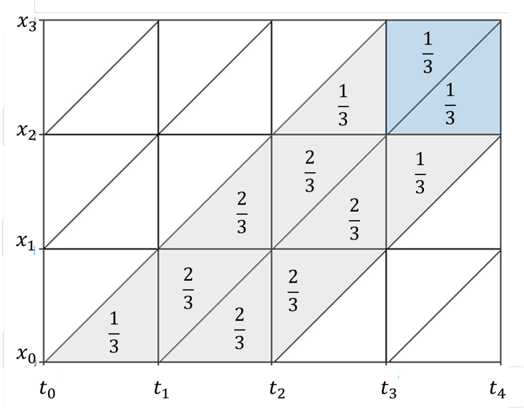
- The deaths counts $D(x_j, t_k)$ are the same as before
- Now: $\underline{\text{How to calculate } E^R(x_j, t_k)}$?

Exposed to risk (Under the assumption of closed population)

- Forward rate: The number of people at risk of dying in next instant is the number of all future deaths. But this number is only known for extinct cohorts.
- Backward rate: The number of people at risk of having died in the previous instant is the number of all people who have already died. **This number can be counted from death data**.

The backward mortality rate – how to calculate $E^R(x_j, t_k)$

The figures below show the weights of the deaths when calculating the exposure.



$$\widehat{E}^R(x_j, t_k) =$$
$$\frac{1}{3}\{D_U(x_j, t_k) + D_L(x_j, t_k)\}$$
$$+\frac{1}{3}\sum_{l=0}^{j} D_L(x_{j-l}, t_{k-l-1}) + D_U(x_{j-l-1}, t_{k-l})$$
$$+\frac{2}{3}\sum_{l=1}^{j} D_U(x_{j-l}, t_{k-l}) + D_L(x_{j-l}, t_{k-l})$$

The backward mortality rate

Under appropriate assumption,

$$m^R(x_j|t_k) = \frac{D(x_j, t_k)}{E^R(x_j, t_k)},$$

is an unbiased estimator of the expected value of $\mu^R(X|T)$ for (X, T) conditioned on the square $[x_j, x_j + 1) \times [t_k, t_k + 1)$.

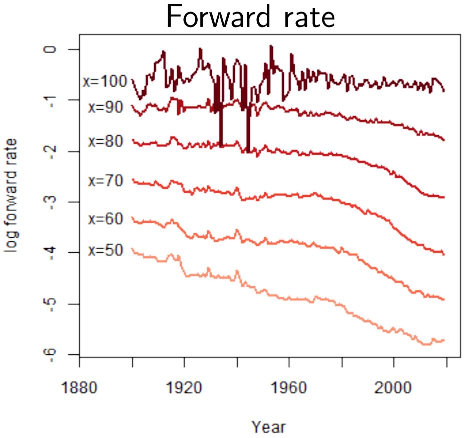
Is the backward mortality rate useful?

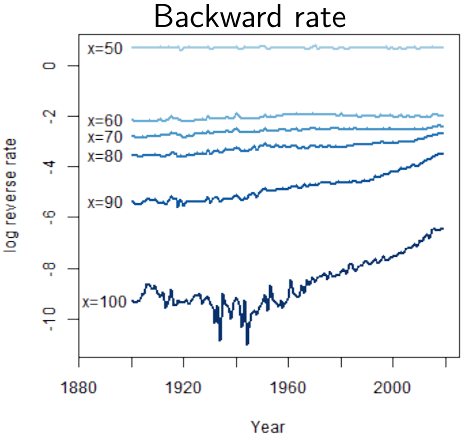
1. The backward mortality rate can be interesting in its own right.
2. We can use the backward mortality rate to estimate the forward mortality rate, and all other common statistics, e.g. expected age of death.
3. Modelling the backward rate can give a new perspective on mortality forecasting.

Illustration using HMD data for England and Wales males

Backward mortality rate can be interesting in its own right

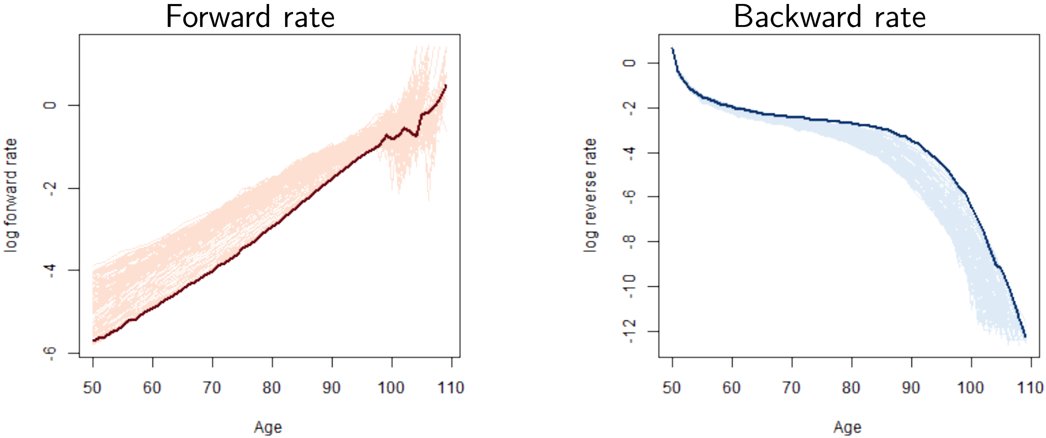
Time trends





Backward mortality rate can be interesting in its own right

Age profile in 2019

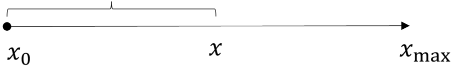


We can use the backward mortality rate to estimate the forward mortality rate, and all other common statistics, e.g. expected age of death.

Relationship between backward mortality rate and forward mortality rate

Forward time

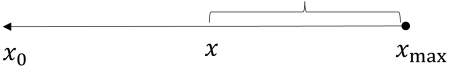
Survive to age x and die in the next instant



- $\mu(x) = \lim_{h \downarrow 0} h^{-1} \mathrm{Pr}\{X \in [x, x+h)\}$
- $f(x) = S(x)\mu(x)$
- $S(x) = e^{-\int_{x_0}^{x} \mu(v)\mathrm{d}v}$

Backward time

Be dead by age x and have died in the previous instant



- $\mu^R(x) = \lim_{h \downarrow 0} h^{-1} \mathrm{Pr}\{X \in (x-h, x]\}$
- $f(x) = F(x)\mu^R(x)$
- $F(x) = e^{-\int_{x}^{x_{\max}} \mu^R(v)\mathrm{d}v}$

$$\boxed{S(x)\mu(x) = F(x)\mu^R(x)}$$

From backward mortality rate to forward mortality rate

We have then

$$\mu(x|t) = \mu^R(x|t) \frac{e^{-\int_x^{x_{\max}} \mu^R(v|t-x+v)\mathrm{d}v}}{1 - e^{-\int_x^{x_{\max}} \mu^R(v|t-x+v)\mathrm{d}v}},$$

where

$$\frac{e^{-\int_x^{x_{\max}} \mu^R(v|t-x+v)\mathrm{d}v}}{1 - e^{-\int_x^{x_{\max}} \mu^R(v|t-x+v)\mathrm{d}v}} = \frac{\text{Probability of dying before } x}{\text{Probability of dying after } x}$$

Problem: The integral runs over **unobserved ages for non-extinct cohorts**.

Solution: **Extrapolate the backward mortality rate** to complete data for non-extinct cohort

Forecasting the backward mortality rate

We can use generalised age-period-cohort models to forecast the backward mortality rate

$$\log m^R(x|t) \quad = \quad \alpha_x \quad + \quad \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} \quad + \quad \gamma_{t-x}$$

Extrapolate period indexes $\kappa_t^{(1)}, \ldots, \kappa_t^{(N)}$ using time series models (e.g. ARIMA models)

From backward mortality rate to forward mortality rate

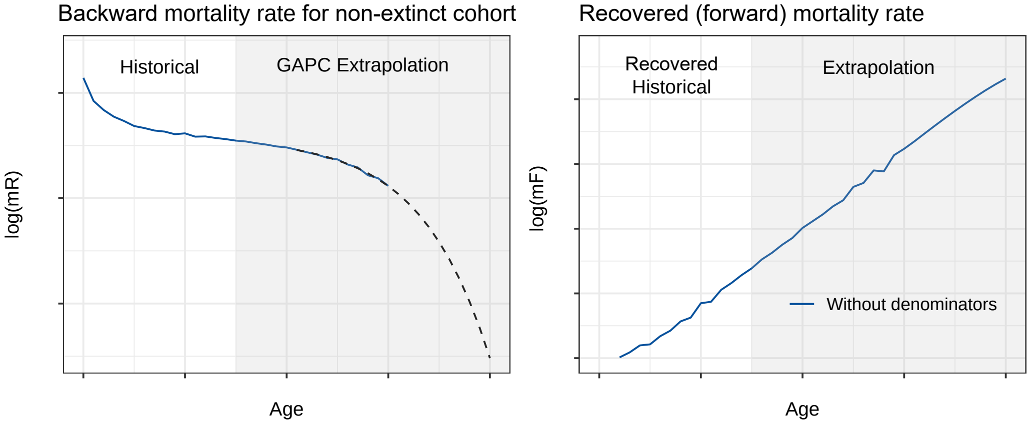
- Then the discrete time estimate is,

$$m^F(x_j|t_k) = m^R(x_j|t_k)\frac{\prod_{l=j}^{J}\{1 - q^R(x_l|t_k - x_j + l)\}}{1 - \prod_{l=j}^{J}\{1 - q^R(x_l|t_k - x_j + l)\}},$$

where $q^R(x|t)$ is an estimate of the conditional probability of dying in the previous year of age given death by age x:

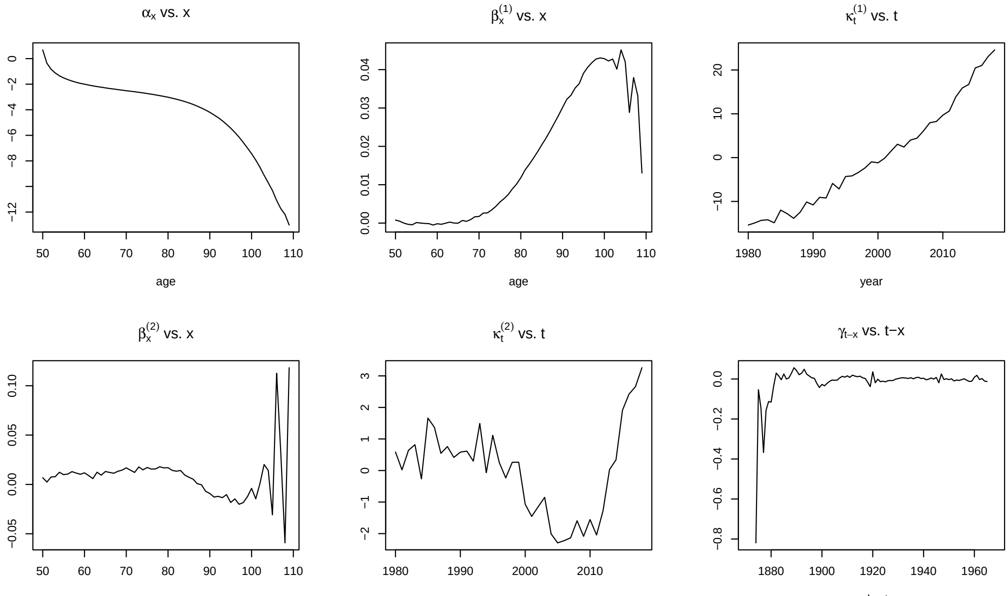
$$q^R(x_l|t_k) = \frac{2m^R(x_l|t_k)}{2 + m^R(x_l|t_k)}$$

From backward mortality rate to forward mortality rate



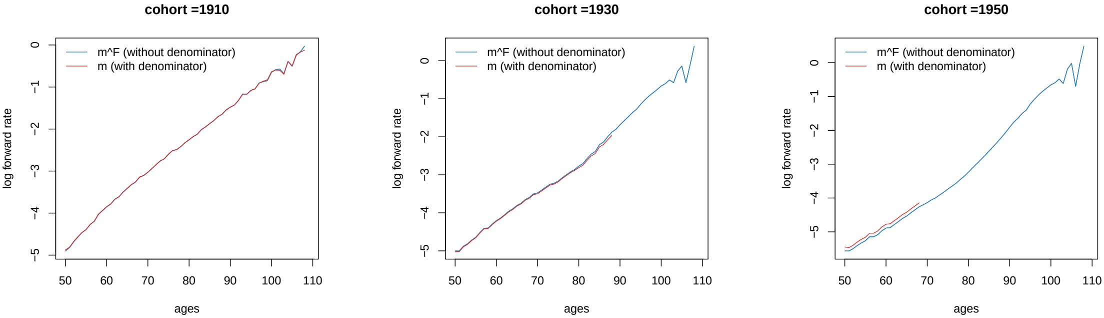
$$q^R(x_l) = \frac{2m^R(x_l)}{2 + m^R(x_l)}, \quad m^F(x_j) = m^R(x_j)\frac{\prod_{l=j}^{J}\{1 - q^R(x_l)\}}{1 - \prod_{l=j}^{J}\{1 - q^R(x_l)\}}$$

RH2, $\log m^R(x|t) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \gamma_{t-x}$



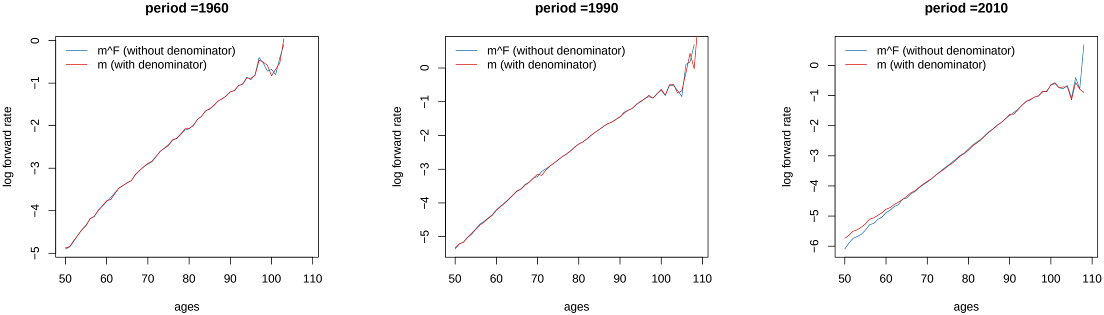
How does the traditional central mortality rate $m(x_j, t_k)$ which uses population data, compare to $m^F(x_j, t_k)$ which only uses death counts?

Mortality along age for fixed cohorts - England and Wales males



The mortality rates estimated without the denominators align well with the ones with denominators

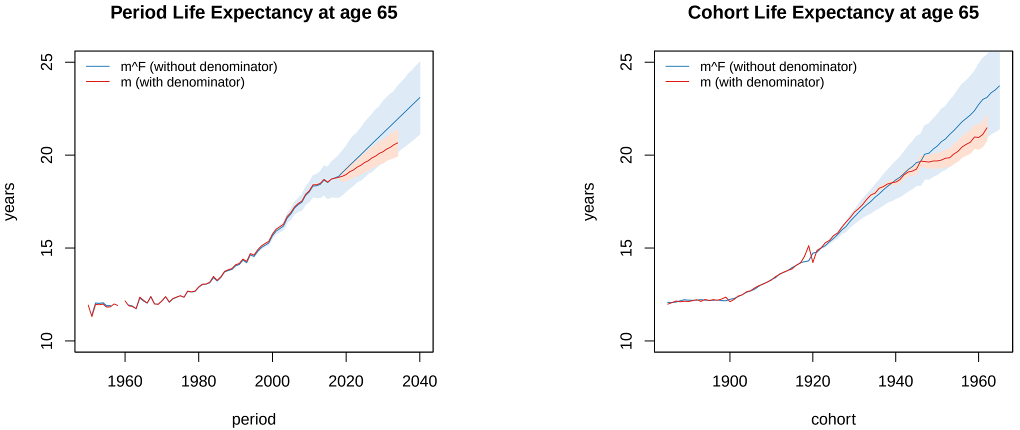
Mortality along age for fixed periods - England and Wales males



The mortality rates estimated without the denominators align well with the ones with denominators but deteriorate for more recent periods

How do projections based on backward rates compare with those based on traditional models which use population data?

Life expectancy projections at age 65 - EW males



Forecast with denominators use the model $\log m(x|t) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \gamma_{t-x}$

Conclusion

- The actual size of the population of interest, if available at all, can often be poor quality
- Propose a way to estimate mortality rates by using death counts only
- The propose approach is reasonably accurate
  - Good fit of rates along both period and cohort
  - Good estimates and projections of life expectancies
- Useful new perspective for projection of mortality at older ages
  - Explore out-of-sample forecast accuracy
  - Check consistency of projections using population sizes
  - Add diversity of projections → model ensembles

Thank you!

a.villegas@unsw.edu.au