European Actuarial Academy, Hotel Modul, Vienna

Technical aspects of modelling longevity risk

Stephen J. Richards, BSc, FFA, PhD

20th June 2013

Contents

1. About the speaker

1. About the speaker

- Consultant on longevity risk since 2005
- Founded longevity-related software businesses in 2006:

LONGEVITAS

mortalityrating.com

- Joint venture with Heriot-Watt in 2009:

PROJECTIONS TOOLKIT

2. Why model longevity risk?

2. Why model longevity risk?

We want to build a model for longevity risk so we can:

— understand the risk in a portfolio,

— know all the financially significant risk factors,

— manage existing risks, and

— correctly price new risks (underwriting).

2. What does a good model look like?

A good model will:

— closely match reality,

— make full use of all available data, but

— summarise the important features about the risk.

3. Data

3. Data

Typically actuaries are faced with portfolios with:

— separate policies with individual lives at risk,

— policies which start on different dates at different ages, and

— policyholders with different combinations of risk factors.

4. Model requirements and challenges

4. Requirements and challenges

Some of the things we want from a model include:

— getting the shape of the risk correct by age,

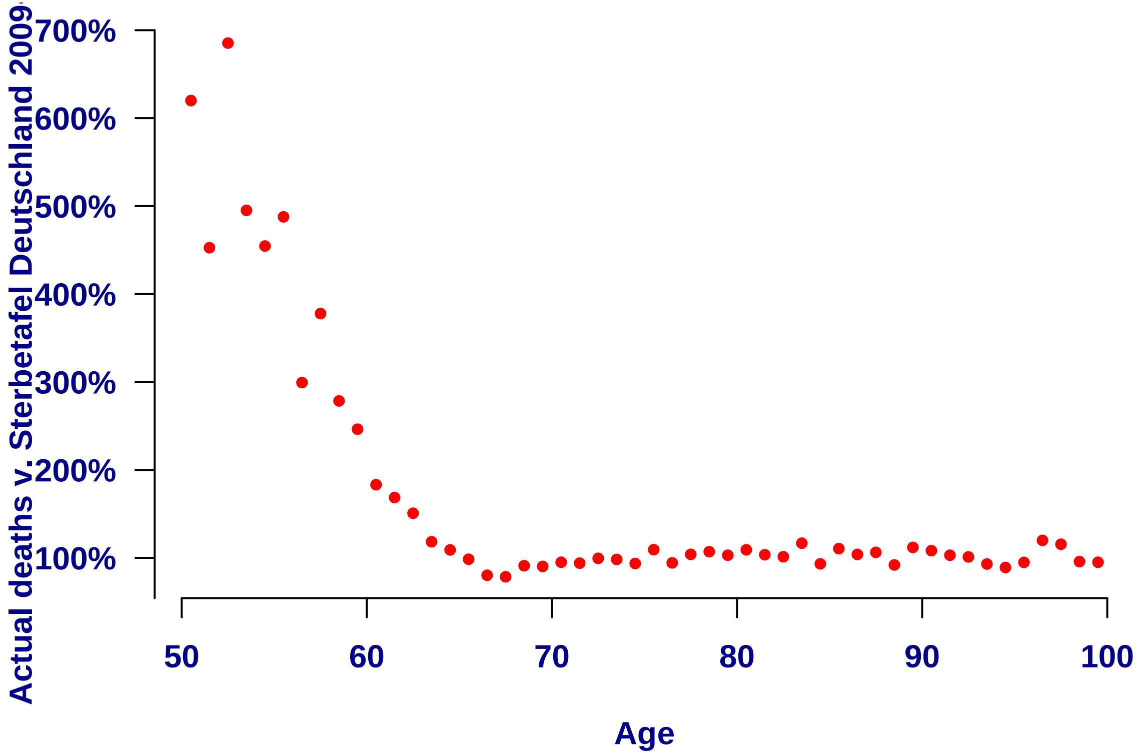
— identifying the main risk factors, and

— extrapolating to higher ages.

We also want a model to use all available data efficiently.

4. Getting the shape of the risk correct by age

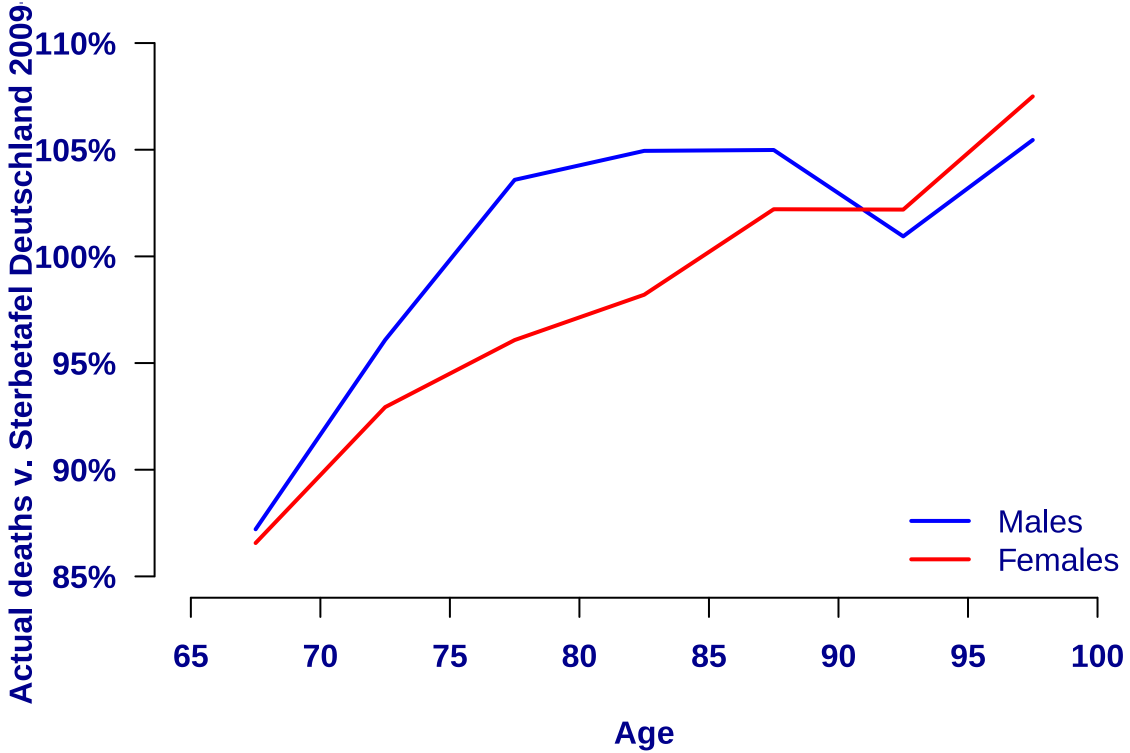
We need a model which follows the shape and pattern of our portfolio:



Source: Longevitas Ltd, using data from Richards, Kaufhold and Rosenbusch (2013). Ratio of observed deaths to the expected deaths according to German population mortality tables for 2009–2011 (males only).

4. Getting the shape of the risk correct by age

We need a model which follows the shape and pattern of our portfolio:



Source: Longevitas Ltd, using data from Richards, Kaufhold and Rosenbusch (2013). Ratio of observed deaths to the expected deaths according to German population mortality tables for 2009–2011.

4. Identifying the effect of risk factors

In the data set in Richards, Kaufhold and Rosenbusch (2013):

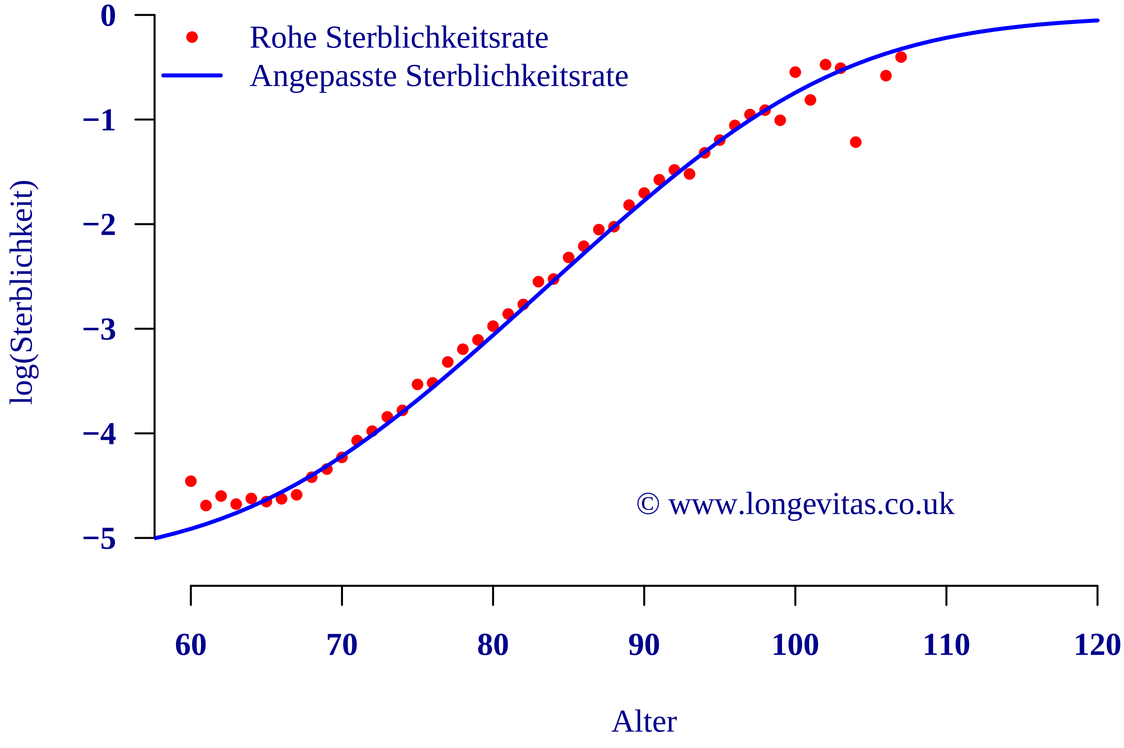
— 34.5% of lives are male, but

— 59.7% of lives with largest pensions are male.

- How do you separate the effects of gender and pension size?
- We need models which can do this without double counting.

4. Extrapolating to higher ages

We need mortality rates at ages where data are sparse or non-existent:



Source: Longevitas Ltd, using data from Richards, Kaufhold and Rosenbusch (2013). See also http://www.longevitas.co.uk/site/informationmatrix/graduation.html.

4. Inefficient uses of your data

- Splitting a data set (stratification) weakens a data set.
- Grouping individuals loses information on which lives actually died.
- Models for q_x:

  (i) lose information on when someone died during the year,

  (ii) lose partial years of exposure, and

  (iii) cannot easily handle competing risks.

5. Model types available

5. Model types available

We will consider five types of model:

— A/E comparisons,

— Whittaker-style graduation,

— Kaplan-Meier analysis,

— Generalized Linear Models (GLMs), and

— survival models.

5. A/E comparisons

Ratio of deaths to the number expected according to a table:

$$\frac{\text{Actual number of deaths}}{\sum_{i=1}^{n} \int_0^{t_i} \mu_{x_i+s} ds}$$

where:

- there are n lives,
- each life i is observed from age x_i to age $x_i + t_i$,
- μ_x is the mortality hazard at age x, and
- μ_x is approximated from a table with $\mu_{x+\frac{1}{2}} \approx -\log(1 - q_x)$

5. A/E comparisons

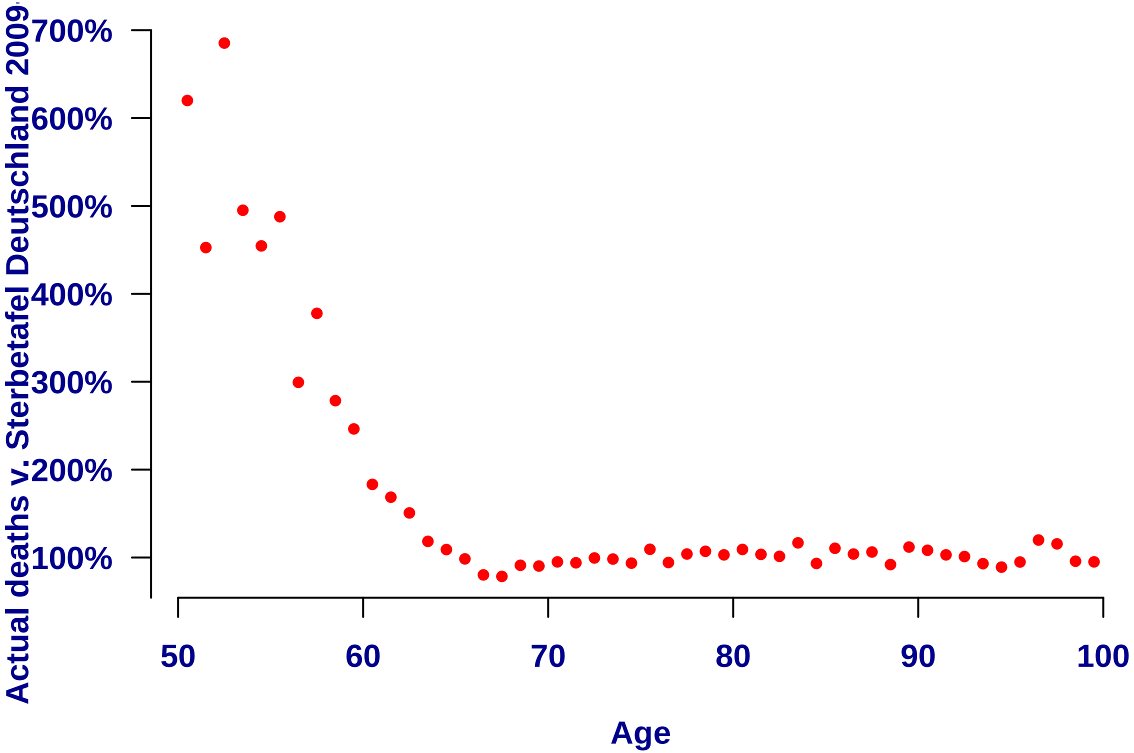
+ Simple — can be done in a spreadsheet
+ Robust when people have multiple policies
+ Provides extrapolated rates via existing table structure

But:

– Cannot handle multiple risk factors without stratification
– Assumes the risk is a constant proportion of the table…

5. A/E comparisons

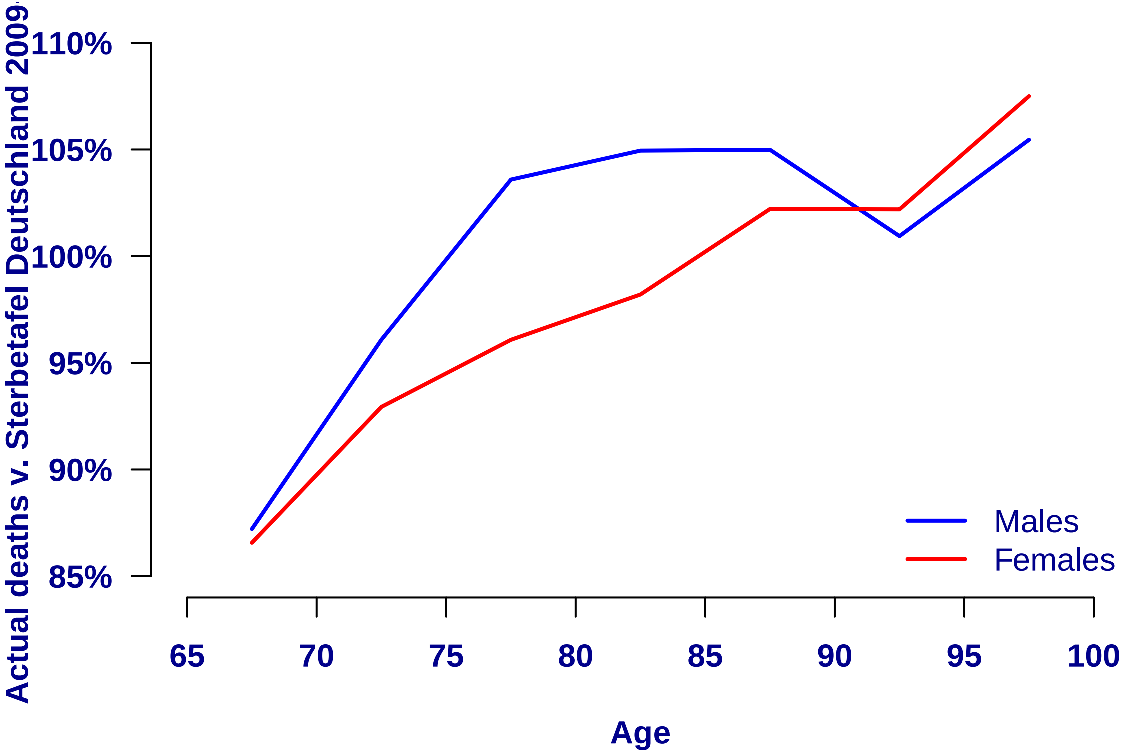
Risk is not a constant proportion of this table:



Source: Longevitas Ltd, using data from Richards, Kaufhold and Rosenbusch (2013). Ratio of observed deaths to the expected deaths according to German population mortality tables for 2009–2011 (males only).

5. A/E comparisons

Restricting the age range does not help much:



Source: Longevitas Ltd, using data from Richards, Kaufhold and Rosenbusch (2013). Ratio of observed deaths to the expected deaths according to German population mortality tables for 2009–2011.

5. Whittaker-style graduation

Find a set of rates m_x^{smooth} which minimizes:

$$\sum(\Delta^3 m_x^{\text{smooth}})^2 + h\sum\left(\frac{d_x}{E_c^c} - m_x^{\text{smooth}}\right)^2$$

where:

— d_x is the number of deaths observed at age x,

— E_x^c is the corresponding central exposed to risk (time lived), and

— h is set arbitrarily to balance the smoothness of the m_x^{smooth} rates against the closeness of fit to the observed deaths.

Source: Whittaker (1919).

5. Whittaker-style graduation

+ Relatively simple — can be done in R

+ Better fit to shape of your risk than A/E comparison

But:

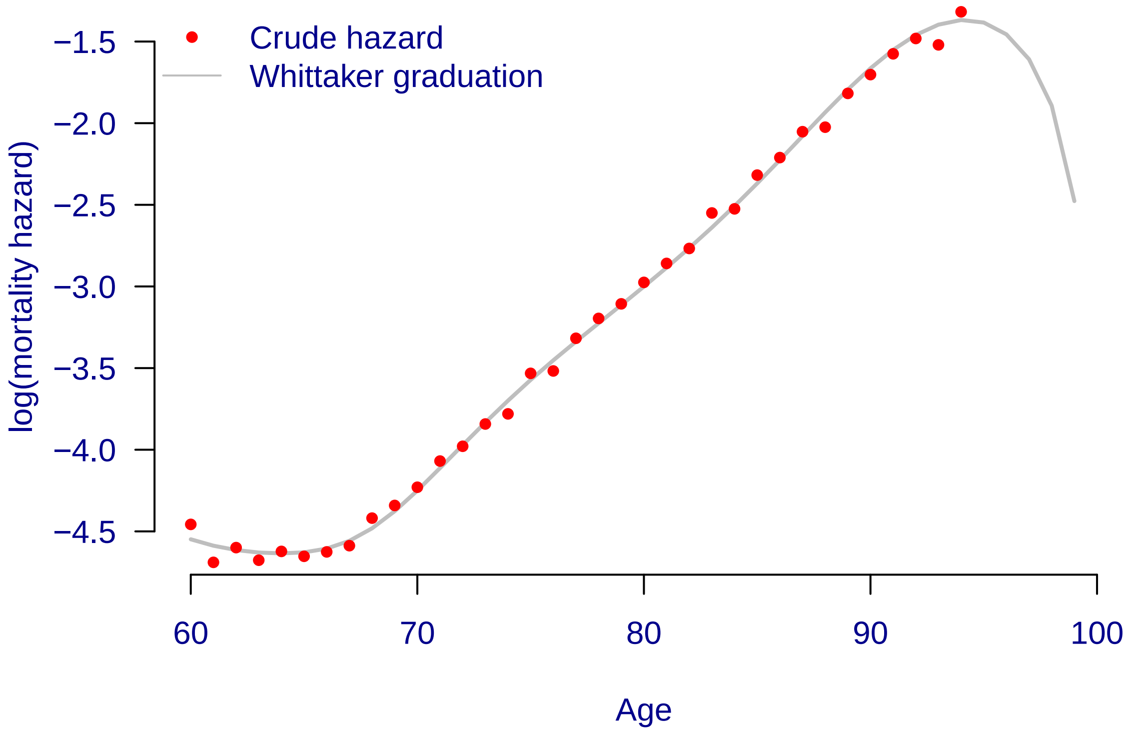
– Cannot handle multiple risk factors without stratification,

– Vulnerable to sparse data, and

– Poor at extrapolation. . .

5. Whittaker-style graduation

Whittaker graduation works well in the region of the data only:



Source: Longevitas Ltd, using data for males from Richards, Kaufhold and Rosenbusch (2013) and h=0.01.

5. Kaplan-Meier

Calculate the empirical survival curve as follows:

$$_{t_j}p_x = \prod_{i=1}^{j \leq n} \left(1 - \frac{d_{x+t_i}}{l_{x+t_i^-}}\right)$$

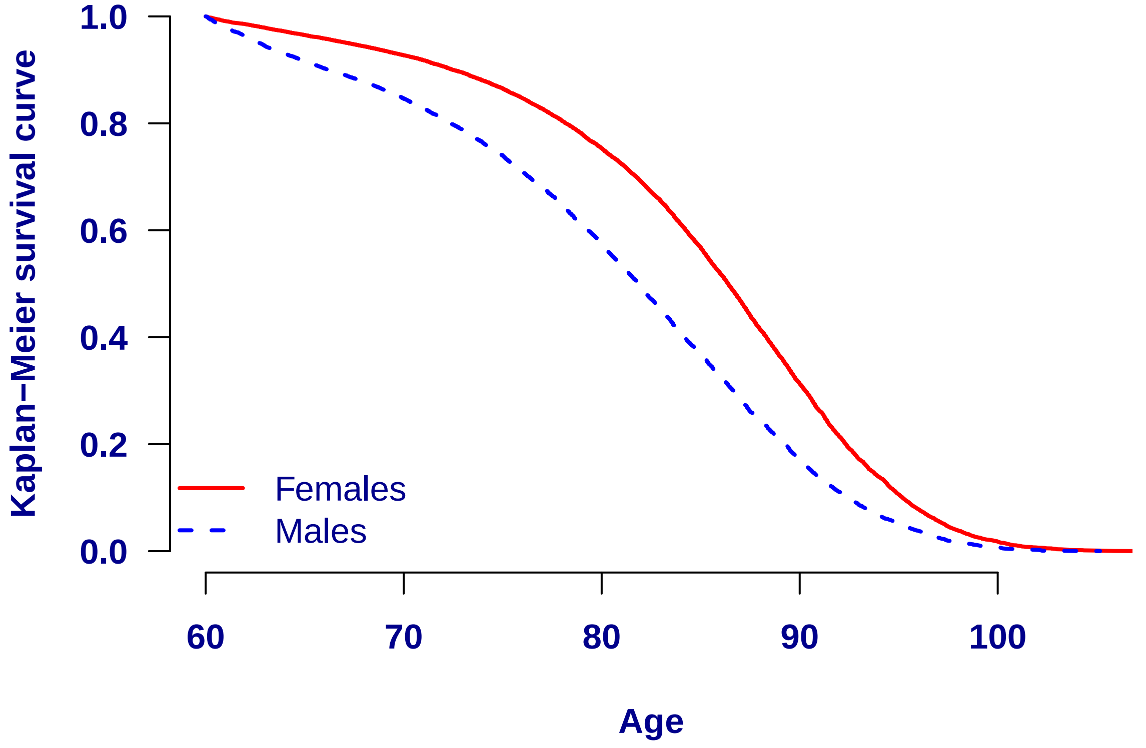
where:

- x is the outset age for the survival curve,
- $\{x + t_i\}$ is the set of n distinct ages at death,
- $l_{x+t_i^-}$ is the number of lives alive immediately before age $x + t_i$,
- d_{x+t_i} is the number of deaths dying at age $x + t_i$.

Source: Richards (2012), an adaptation from the concept from Kaplan und Maier (1958).

5. Kaplan-Meier curve

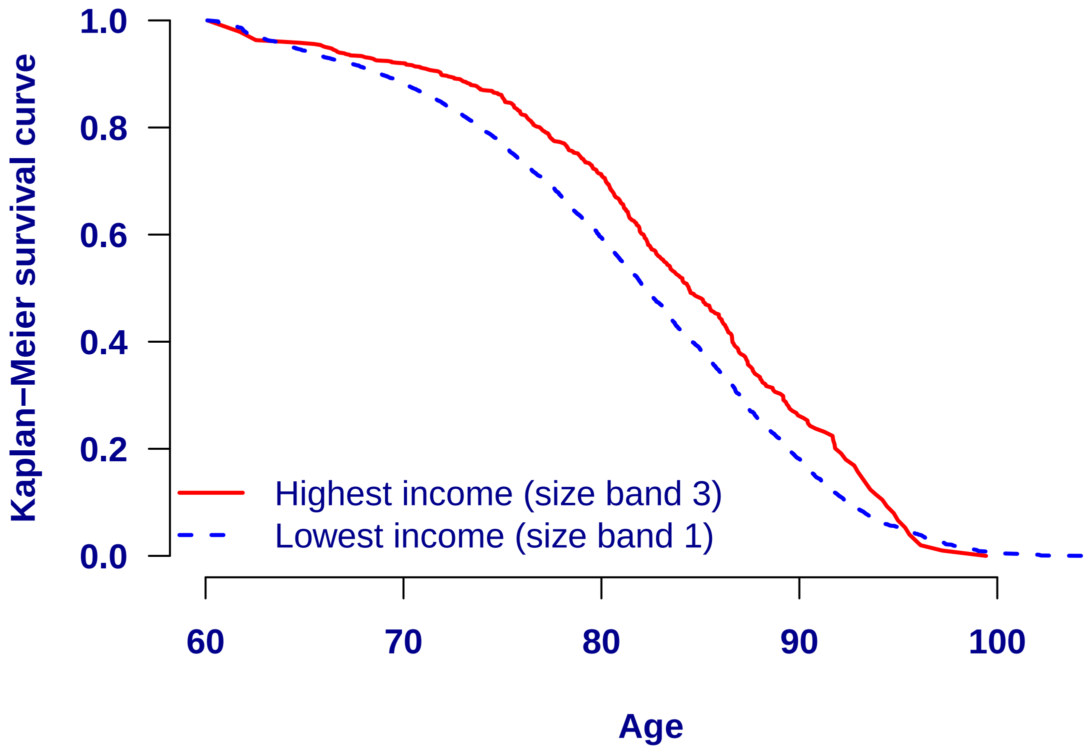
Actually a step function, but it looks smooth for large numbers of deaths:



Source: Richards, Kaufhold and Rosenbusch (2013).

5. Kaplan-Meier

A very useful tool for exploratory data analysis:



Source: Richards, Kaufhold and Rosenbusch (2013).
See also http://www.longevitas.co.uk/site/informationmatrix/doyouhatestatisticalmodels.html

5. Kaplan-Meier

+ Simple concept, supported in most statistical packages including R
+ Fits the data well

But:

– Cannot handle multiple risk factors without stratification, and
– Not a summary of the data, just a restatement of it.

5. GLMs for grouped counts

We assume a statistical model as follows:

$$D_x \sim \text{Binomial}(n_x, q_x)$$

or else:

$$D_x \sim \text{Poisson}(E_x^c \mu_x)$$

where:

— D_x is the number of observed deaths,

— n_x is the number of lives aged x,

— q_x is the mortality rate for age x,

— μ_x is the mortality hazard for age x, and

— E_x^c is the time lived exposed to risk of death at age x.

5. GLMs for grouped counts

+ Available in standard statistical software, such as R
+ Good at extrapolation
+ Can handle multiple risk factors

But:

– Loses information through grouping

– Binomial model loses further information through modelling q_x

– Poisson model requires minimum expected number of deaths per cell, which limits number of risk factors

5. GLMs for individual lives

We build a model for the individual probability of death, q_{x_i}, as follows:

$$\log\left(\frac{q_{x_i}}{1-q_{x_i}}\right) = \sum_j \alpha_j z_{i,j} + x_i \sum_j \beta_j z_{i,j}$$

where:

— each life i starts the year of observation ages x_i,

— there are j risk factors with main effects α_j,

— the main effects interact with age with β_j, and

— the indicator variable $z_{i,j}$ takes the value 1 when life i has risk factor j, and zero otherwise.

5. GLMs for individual lives

+ Available in standard statistical software, such as R
+ Good at extrapolation
+ Can handle unlimited number of risk factors
+ No stratification

But:

– Cannot easily handle competing risks
– Failure of independence assumption across multiple years…

5. Common mistakes with GLMs for individuals

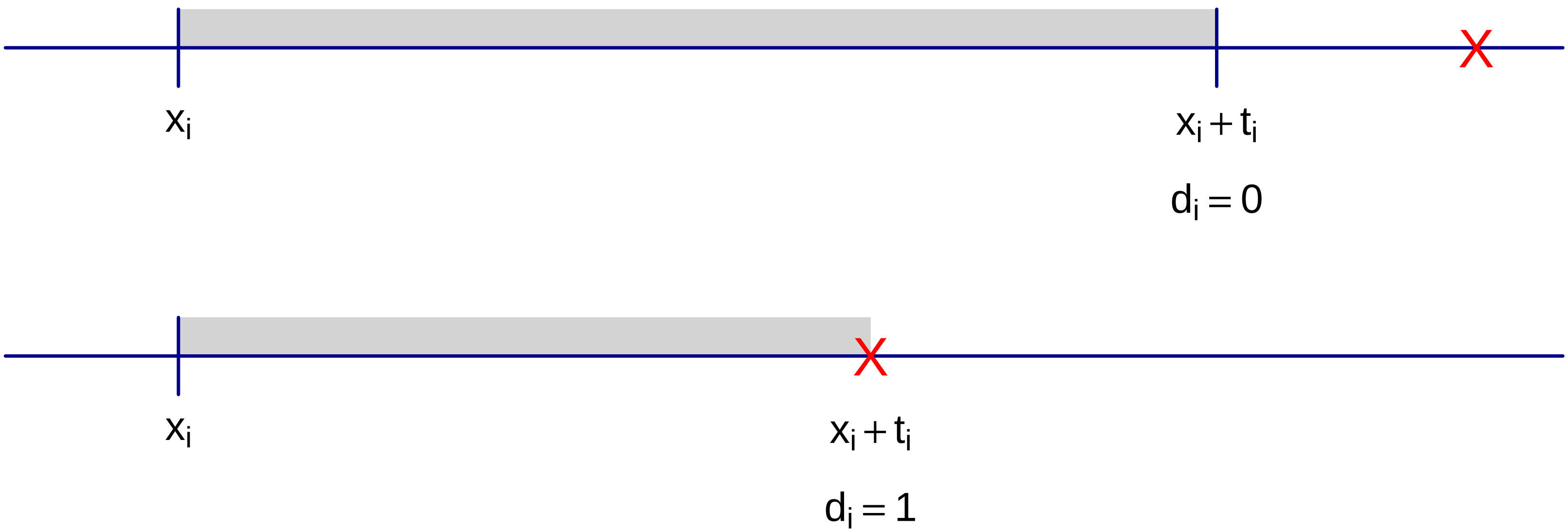
- Having each individual appear several times[1]
- Incorrectly allowing for partial years of exposure[2]
- Modelling ${}_tq_x$ — not linear when $t > 1$

[1] See http://www.longevitas.co.uk/site/informationmatrix/logisticalnightmares.html

[2] See http://www.longevitas.co.uk/site/informationmatrix/partofthestory.html

5. Survival models

Simple observational structure as longitudinal study:



Time observed, t_i, is shown in grey, while deaths are marked $\times$.

5. Survival models

- Time observed, t_i, is *waiting time* (*central exposed-to-risk* to actuaries).
- d_i is the event indicator: 1 for dead, 0 for alive.
- t_i and d_i not independent, so considered as a pair $\{t_i, d_i\}$.
- Not all lives are dead, so survival times are *right-censored*.
- Lives enter at age $x_i > 0$, so data is also *left-truncated*.

5. Survival models

- Survival models are ideal for actuarial work — Richards (2008, 2012).
- A portfolio of risks is like a medical study with continuous recruitment.
- The future lifetime of an individual aged x is a random variable, T_x.
- T_x has a probability density function ${}_tp_x\mu_{x+t}$ for $t > 0$.

3. Overview of some common models

$$\begin{aligned}
\text{Gompertz} \quad & \mu_x = e^{\alpha+\beta x} \\
\text{Makeham} \quad & \mu_x = e^{\epsilon} + e^{\alpha+\beta x} \\
\text{Perks} \quad & \mu_x = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} \\
\text{Beard} \quad & \mu_x = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}} \\
\text{Makeham} - \text{Perks} \quad & \mu_x = \frac{e^{\epsilon} + e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} \\
\text{Makeham} - \text{Beard} \quad & \mu_x = \frac{e^{\epsilon} + e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}
\end{aligned}$$

Source: Richards (2008, 2012).

5. Survival models

+ Good at extrapolation
+ Can handle unlimited risk factors
+ No stratification
+ Independence assumption respected
+ Can handle competing risks

6. Conclusions

6. Conclusions

- Kaplan-Meier curves useful for exploratory data analysis.
- Statistical models are best for:
  - — summarizing main risk features,
  - — separating the effect of risk factors, and
  - — extrapolating to ages with sparse data.
- Statistical models for individuals avoid stratification.
- Survival models most closely match the reality of individual risk.
- Example application to follow in second presentation...

References

Kaplan, E. L. and Meier, P. **1958** *Nonparametric estimation from incomplete observations*, Journal of the American Statistical Association **53**, 457–481.

Richards, S. J. **2012** *A handbook of parametric survival models for actuarial use*, Scandinavian Actuarial Journal, 2012 (**4**), pages 233–257.

Richards, S. J., Kaufhold, K. and Rosenbusch, S. **2013** *Creating portfolio-specific mortality tables: a case study*, Longevitas working paper

Whittaker, E. T. **1919** *A New Basis for Graduation*, Proceedings of the Edinburgh Mathematical Society, XLI, 63

LONGEVITAS