Scandinavian Actuarial Journal
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713690025

A handbook of parametric survival models for actuarial use

S. J. Richards[a]
[a] Richards Consulting, Edinburgh, UK

First published on: 19 August 2010

Scandinavian Actuarial Journal
2010, 1–25, iFirst article

Original Article

A handbook of parametric survival models for actuarial use

S. J. RICHARDS

Richards Consulting, 4 Caledonian Place, Edinburgh EH11 2AS, UK

(*Accepted 20 June 2010*)

Traditional actuarial techniques for mortality analysis are being supplanted by statistical models. Chief amongst these are survival models, which model mortality continuously at the level of the individual. An assumption of a mathematical form for the hazard function or, equivalently, the assumption of a continuous distribution for an individual's lifetime, leads automatically to smooth fitted mortality rates. This note gives an overview of the survival models commonly found in statistical packages and compares their suitability for actuarial work with the mortality 'laws' proposed by actuaries over the past two centuries. We find that the actuarial laws provide substantially better fits at post-retirement ages. We also give a common structure of parameterisation which gives consistent behaviour and interpretation of risk factors across all 16 survival models listed here. Finally, we consider the benefits of working directly with the log-likelihood function, including making allowance for the left truncation which is common for the data with which actuaries work.

Keywords: Survival models; Mortality laws; Left truncation

1. Introduction

Richards (2008) compared of the effectiveness of six actuarial mortality 'laws' in explaining patterns of mortality in a pensioner data set. However, these laws are not all widely used outside the actuarial community as standard software often cannot handle them. Instead, such software often makes available other survival models which were not shown in Richards (2008). This paper provides a comparison of 16 different survival models and shows why actuaries use their mortality laws in preference to the models often used by other practitioners. The target audience is actuaries who want to know more about survival models outside the life-insurance world, and other practitioners who want to know why actuaries build survival models the way they do.

In this paper a survival model will be regarded as synonymous with a model for the continuous-time hazard function for an individual. The hazard function is known to actuaries as the force of mortality, and the hazard rate at age x, μ_x, is defined as:

E-mail: stephen@richardsconsulting.co.uk

Scandinavian Actuarial Journal
ISSN 0346-1238 print/ISSN 1651-2030 online © 2010 Taylor & Francis
http://www.tandf.no/saj
DOI: 10.1080/03461238.2010.506688

$$\mu_x = \lim_{h \to 0^+} \frac{1}{h} \Pr(\text{death before age } x+h | \text{alive at age } x)$$
$$= \lim_{h \to 0^+} \frac{{}_h q_x}{h} \tag{1}$$

where ${}_h q_x$ denotes the probability of a life currently aged x dying in the small interval of time h. Using μ_x contrasts with the historical actuarial habit of using discrete-time mortality rates, denoted q_x, which typically apply over a single year (i.e. ${}_1 q_x$). In the pre-computer era, q_x was preferred for reasons of expediency in calculation.

The models presented here deal with mortality at the level of the individual. However, it should be noted that a Poisson model for the number of deaths in a group is also a model for the hazard function, and could hence be viewed as a survival model. Similarly, a model for q_x can be used to approximate the survival curve. Both the Poisson and q_x models are concerned with the count of the number of events taking place, whereas the individual hazard models used in this paper deal with the time until an event occurs.

One immediate advantage of modelling the hazard rate is that it allows each and every piece of data to contribute to the model. In contrast, modelling the annualised mortality rate, q_x, involves throwing away data where the policyholder could not have completed a full year of exposure. While it is possible to make certain assumptions to enable q_x models to handle fractional years of exposure, these assumptions introduce unnecessary complications.

To illustrate this loss of information in q_x models, consider the following example from Richards (2008). Two groups each consist of four lives alive at the start of the year. During the course of the year one life dies in each group, making the estimated mortality rate, $\hat{q}_A = \hat{q}_B = \frac{1}{4}$ in both cases. If the death in Group A occurs at the end of January, the estimated force of mortality is $\hat{\mu}_A = \frac{1}{3\frac{1}{12}} = 0.324$. If the death in Group B occurs at the start of December the estimated force of mortality is $\hat{\mu}_B = \frac{1}{3\frac{11}{12}} = 0.255$. As this simple example shows, working with the force of mortality means we can use all the information available, and will usually result in a better model. In contrast, working with q-type rates throws away the information on time of death and is therefore less sophisticated.

Another reason for using μ_x is that it can be used to exactly derive q_x using Eq. (2):

$$q_x = 1 - \exp\left(-\int_0^t \mu_{x+s} ds\right) \tag{2}$$

Models for μ_x also lend themselves to multiple-decrement analysis or competing-risk problems without further adjustment. In contrast, a model for q_x cannot normally be used to derive μ_x without further assumptions or approximations. Furthermore, q_x models require more assumptions for each additional decrement simply to fit the model. It is for these reasons that actuaries are switching to survival models, 16 of which are surveyed in this paper for use in studying pensioner mortality.

2. Data, methodology and terminology

Survival models are widely used in the analysis of medical trials (Collett (2003)). A life-insurance portfolio or a pension scheme is similar in many ways to a medical trial with continuous recruitment as new lives join the existing portfolio. However, there are some important differences, the first of which is scale; a small medical trial might have only a few tens of observations, whereas a small annuity portfolio could have tens of thousands of policies. The largest portfolios of annuitant data in the UK can exceed a million records.

Medical trials are primarily interested in detecting differences between groups or treatments, but are less concerned with estimating the precise shape of the hazard function. Actuaries are also interested in differences between groups, but they are also crucially interested in the shape of the hazard function (and thus survivor function) for pricing liabilities. The time value of money makes actuaries more keenly interested in the precise shape of the hazard function than other researchers.

As with medical-trials data, when an extract of mortality data is taken from an administration system not all lives will be dead at the extract date. Such data are called *right-censored*, since all that can be said of the mortality process is that it will occur after the observation time. Right censorship is standard in survival models, and all software implementations can handle this easily enough. The upper example in Figure 1 shows a right-censored observation as the extract has taken place at age x_i+t_i before death has occurred (marked with a cross).

A particular feature of life-insurance contracts or pension benefits is that they commence when people are well into adult life. The lifetimes observed are called *left-truncated*, since observation starts at age x_i and we have no data on deaths and exposure prior this age. This poses a problem for many implementations of survival models which rely on dealing with age-varying mortality through a variable transformation. These survival models are often fitted using existing algorithms for Generalised Linear Models (GLMs) – see Aitken *et al*. (1989), who demonstrate how to fit Weibull and other survival models using a Poisson GLM. However, such an approach demands that the lives be

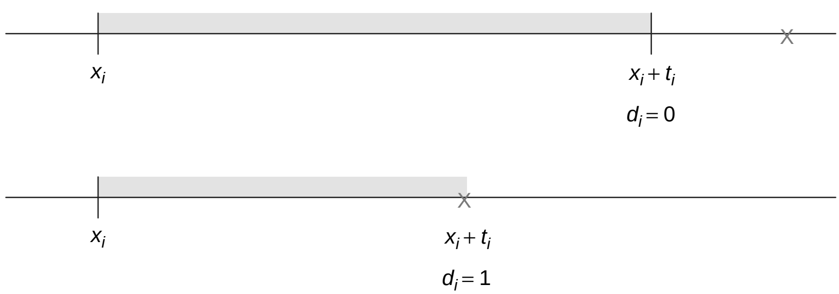

Figure 1. Diagram of survival-model setup. The time observed, t_i, is shown in grey, while deaths are marked with a cross, $\times$. Since people do not usually enter into life-insurance contracts at birth, observations are left-truncated, i.e. lives start being observed at age $x_i > 0$. The upper example is right-censored as death happens after the end of the observation period.

observed from outset, i.e. from birth if chronological age is to be used directly. Thus, actuaries working with typical life-insurance data cannot rely on standard implementations of survival models due to left truncation. Instead, they work directly with the log-likelihood in Eq. (6). This is computationally more intensive, but it frees the actuary to use a much wider choice of hazard functions. As we will see, this wider choice leads to some substantial improvements in model fit.

A feature of some medical trials is *interval censoring*, namely where death is known to have occurred between two dates, but the precise date of death is not known. This can happen where a patient was last examined on a given date (and was hence known to be alive), but who does not turn up for a later check and the researcher learns that the patient has died. The date of death is not known, but the interval in which death occurred is known. In actuarial work, however, the involvement of financial payments and legal processes means that a precise date of death typically is known, so interval censoring is rarely required in life-office data.

To fit a survival model we will need to specify the log-likelihood function. For each life i of n lives we have: (1) an entry age, x_i; (2) a time observed, t_i; and (3) an indicator variable, d_i, for the state of the life at age x_i+t_i. The variable d_i takes the value 0 on survival and 1 on the event of interest. This event can be death (as in this paper) or any other decrement of interest, such as critical-illness claim, lapse or surrender. The likelihood function, L, is therefore given by:

$$L \propto \prod_{i=1}^{n} {}_{t_i}p_{x_i}\mu_{x_i+t_i}^{d_i} \tag{3}$$

where ${}_tp_x$ is the probability of surviving from age x to age $x+t$ and is given by:

$${}_tp_x = e^{-H_x(t)} \tag{4}$$

where $H_x(t)$ is the integrated hazard function:

$$H_x(t) = \int_0^t \mu_{x+s}ds \tag{5}$$

We can therefore substitute Eq. (4) into Eq. (3) and take natural logarithms to get the log-likelihood function, ℓ:

$$\ell = \sum_{i=1}^{n} -H_{x_i}(t_i) + \sum_{i=1}^{n} d_i \log\mu_{x_i+t_i} \tag{6}$$

Thus, when applying survival models to individual data, it simply suffices to specify the structure of the hazard rate, μ_x, and subsequently derive $H_x(t)$. When fitting any model, we choose the parameter values to maximise the log-likelihood function in Eq. (6).

Another major difference between medical trials and actuarial work is that life-assurance data are of policies, not people; administration systems are normally set up to process policies, these being the legal liability of the insurer. Life-assurance work therefore requires an additional data-preparation stage not normally required elsewhere; deduplication, i.e. the identification of multiple annuities held by the same person. Failure to process policy data into lives would violate the independence assumption, since the

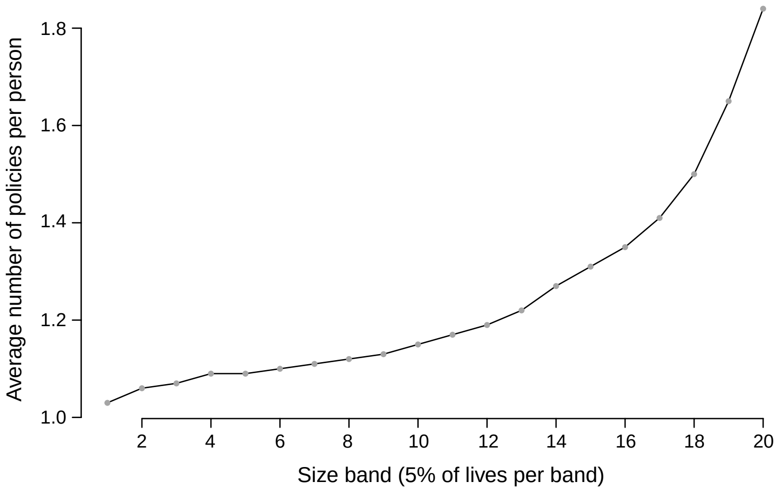

Figure 2. Average number of policies per person in each of equal-sized membership bands ordered by total annual annuity income. Band 1 is the 5% of lives with smallest annual pensions, through to band 20 which is the 5% of lives with the largest annual pensions. Figure reproduced from Richards and Currie (2009).

number of policies per person is correlated with some of the very risk factors actuaries need to investigate – see Figure 2. In the past actuaries have had to make corrections for over-dispersion in the absence of proper deduplication – see Daw (1951). However, actuaries nowadays use proper deduplication algorithms such as those described by Richards (2008).

To illustrate the practical points in this paper we will use a deduplicated data set of over 300,000 life-office pension annuities with over 40,000 deaths. Figure 3 shows the observed

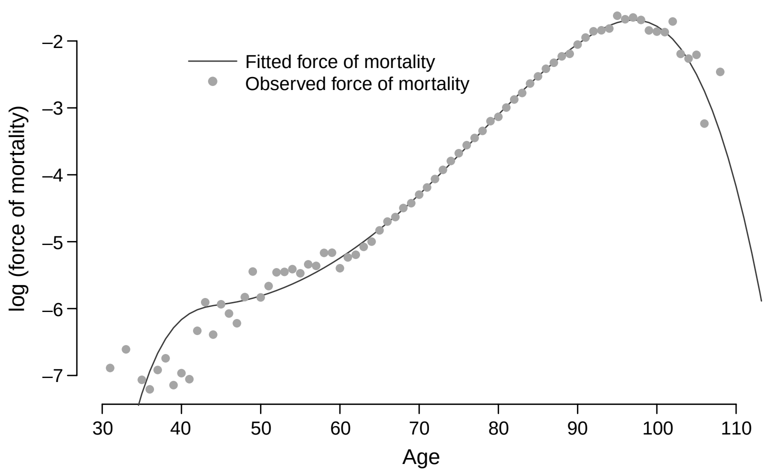

Figure 3. Force of mortality for pensioners between ages 30 and 110: observed crude force of mortality (•)together with fitted values from P-spline regression. There is evidence of data-quality problems above age 95, which is common for life-office data sets. Source: Richards (2008).

force of mortality by age on a logarithmic scale. Between ages 60 and 90 mortality increases in a roughly linear way, i.e. exponentially increasing mortality on the natural scale. Below age 60 this linearity breaks down as the non-age-related component of mortality makes itself felt. Above age 95 there is evidence of data-quality problems, which is common for life-office data sets such as this.

3. Mortality laws and distributions for future lifetime

A major advantage of fitting a formula for the force of mortality is that smoothness is built-in and there is no need to separately graduate (smooth) the resulting fitted rates. In this paper we will look at some actuarial mortality laws listed in Table 1. The parameterisations in Table 1 are often different from those used by the original authors, such as Gompertz (1825) who gave his law as $\mu_x = Bc^x$, with $B > 0$ and $c > 0$. The more modern exponential parameterisations mean we can dispense with any constraints on the range of parameters, allowing them to vary over the entire real line. This has practical advantages in optimising log-likelihood functions using computers.

The naming convention in Table 1 follows Richards (2008) and is different from what might be seen elsewhere. For example, the model labelled as Makeham–Beard was proposed by Perks (1932). We have opted: (1) to use the term Makeham wherever the constant e^ε appears; (2) to name the logistic form $\frac{e^a}{1+e^a}$ after Perks; and (3) to use the term Beard wherever the logistic form has a so-called heterogeneity parameter, ρ.

Some of the models in Table 1 are related to the proportional hazards model of Cox (1972). For example, the Gompertz model can be expressed as a proportion of a baseline hazard, albeit as a time- or age-varying proportion. The Makeham model, however, cannot be expressed in terms of a baseline hazard due to the non-multiplicative e^ε term. The models in Table 1 are mainly non-linear in their nature, although this does not cause

Table 1. Some actuarial mortality laws and their corresponding integrated hazard functions, $H_x(t)$.

Mortality law	μ_x	$H_x(t)$
Gompertz (1825)	$e^{\alpha+\beta x}$	$\frac{(e^{\beta t}-1)}{\beta}e^{\alpha+\beta x}$
Makeham (1859)	$e^\varepsilon + e^{\alpha+\beta x}$	$te^\varepsilon + \frac{(e^{\beta t}-1)}{\beta}e^{\alpha+\beta x}$
Perks (1932)	$\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$	$\frac{1}{\beta}\log\left(\frac{1+e^{\alpha+\beta(x+t)}}{1+e^{\alpha+\beta x}}\right)$
Beard (1959)	$\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$	$\frac{e^{-\rho}}{\beta}\log\left(\frac{1+e^{\alpha+\rho+\beta(x+t)}}{1+e^{\alpha+\rho+\beta x}}\right)$
Makeham–Perks (1932)	$\frac{e^\varepsilon + e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$	$te^\varepsilon + \frac{(1-e^\varepsilon)}{\beta}\log\left(\frac{1+e^{\alpha+\beta(x+t)}}{1+e^{\alpha+\beta x}}\right)$
Makeham–Beard (1932)	$\frac{e^\varepsilon + e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$	$te^\varepsilon + \frac{(e^{-\rho}-e^\varepsilon)}{\beta}\log\left(\frac{1+e^{\alpha+\rho+\beta(x+t)}}{1+e^{\alpha+\rho+\beta x}}\right)$

any real difficulties in fitting them. Here we have used derivatives-based methods for optimising the log-likelihood, where possible, with an explicit formulaic calculation of the information matrix for inversion to calculate the covariance matrix. Numerical approximations are used to verify the derivative calculations, or to substitute for derivatives when they could not be computed in a closed form. For converting into mortality rates, q_x, for use in older actuarial systems we use Eq. (2).

We orientate our descriptions of parameters around the best-known actuarial mortality law, that of Gompertz (1825). Since the hazard function is a straight line on a logarithmic scale, we will refer to the value of α as the `Intercept`, and deviations from this will be the main effect of a risk factor. The parameter β is the coefficient for age, and deviations from this for a categorical risk factor will be the interaction of that risk factor with age. The parameter, ε, will be denoted the Makeham parameter, while ρ will be denoted the Beard parameter. Both the Makeham and Beard parameters may interact with main effects, but not with age.

The survival models used outside actuarial work are typically different from those listed in Table 1. For example, R is a free statistical modelling package and survival models are available in the `survival` library. The distributions available in R include: Extreme value (Gompertz), Logistic, Normal (Gaussian), Weibull, Exponential, Rayleigh, Lognormal, Log-Logistic and *t*. The Rayleigh distribution is a special case of the Weibull distribution, so it need not be considered separately. By way of comparison, SAS is a proprietary statistical modelling package and survival models are available in the LIFEREG procedure. The distributions available in SAS include: Exponential, Logistic, Normal (Gaussian), Weibull, Lognormal, Log-Logistic and Generalised Gamma. We list the hazard and integrated hazard functions in Table 2.

The parameterisations in Table 2 are often different from what can be found elsewhere. We use four broad unifying principles in the parameterisations in this paper. First, where a parameter, θ say, must be positive, we use e^{θ} instead to allow θ to vary across the entire real line. This makes computation easier as one does not need extra programming to enforce the sign of the parameter. Second, we adopt a parameterisation such that an increase in a parameter's value means an increase in the mortality hazard. Third, for a given hazard function we adopt the simplest parameterisation possible. Fourth, for models which are a generalisation of one or more others, we adopt a parameterisation such that the general form simplifies into the more specific form when a parameter is set to zero or one, or tends to infinity.

As an example of the first principle, consider the hazard function, $h(t)$, for the Exponential distribution given by Collett (2003):

$$h(t) = \lambda \tag{7}$$

where $t \in [0, \infty)$ and $\lambda > 0$. In keeping with the first principle we therefore use the following equivalent alternative to Eq. (7):

$$h(t) = e^{\alpha} \tag{8}$$

which leaves α free to vary over the real line, thus saving the specification of a constraint.

Table 2. Some truncated distributions for future lifetime and their corresponding hazard and integrated hazard functions, $H_x(t)$, where $x > 0$.

Distribution	μ_x	$H_x(t)$
Exponential	e^α	te^α
Extreme value	See Gompertz hazard in Table 1	
Pareto	$\dfrac{e^\alpha}{x}$	$e^\alpha \log\left(\dfrac{x+t}{x}\right)$
Weibull	$e^\alpha x^{\sigma-1}$	$\begin{cases} e^\alpha \log\left(\dfrac{x+t}{x}\right), & \sigma=0; \\ \dfrac{e^\alpha}{\sigma}[(x+t)^\sigma - x^\sigma], & \text{otherwise.} \end{cases}$
Logistic	$\dfrac{1}{e^\sigma\left(1+\exp\left(-\dfrac{x+\alpha}{e^\sigma}\right)\right)}$	$\log\left(\dfrac{1+\exp\left(\dfrac{x+t+\alpha}{e^\sigma}\right)}{1+\exp\left(\dfrac{x+\alpha}{e^\sigma}\right)}\right)$
Log-Logistic	$\dfrac{e^{\alpha+\sigma}x^{e^\sigma-1}}{1+e^\alpha x^{e^\sigma}}$	$\log\left(\dfrac{1+e^\alpha(x+t)^{e^\sigma}}{1+e^\alpha x^{e^\sigma}}\right)$
Normal	$\dfrac{\dfrac{1}{e^\sigma\sqrt{2\pi}}\exp\left(-\dfrac{(x+\alpha)^2}{2e^{2\sigma}}\right)}{1-\Phi\left(\dfrac{x+\alpha}{e^\sigma}\right)}$	$\log\left(\dfrac{1-\Phi\left(\dfrac{x+\alpha}{e^\sigma}\right)}{1-\Phi\left(\dfrac{x+t+\alpha}{e^\sigma}\right)}\right)$
Lognormal	$\dfrac{\dfrac{1}{xe^\sigma\sqrt{2\pi}}\exp\left(-\dfrac{(\log x+\alpha)^2}{2e^{2\sigma}}\right)}{1-\Phi\left(\dfrac{\log x+\alpha}{e^\sigma}\right)}$	$\log\left(\dfrac{1-\Phi\left(\dfrac{\log x+\alpha}{e^\sigma}\right)}{1-\Phi\left(\dfrac{\log(x+t)+\alpha}{e^\sigma}\right)}\right)$
Inverse Gaussian	$\dfrac{\left(\dfrac{e^\sigma}{2\pi x^3}\right)^{\frac{1}{2}}\exp\left(-\dfrac{e^\sigma(x-e^{-\alpha})^2}{2e^{-2\alpha}x}\right)}{\text{IGS}(x)}$	$\log\left(\dfrac{\text{IGS}(x)}{\text{IGS}(x+t)}\right)$
Gamma	$\dfrac{e^\alpha x^{e^\lambda-1}\exp(-xe^{\alpha e^{-\lambda}})}{\Gamma(e^\lambda)-\gamma(e^\lambda, xe^{\alpha e^{-\lambda}})}$	$\log\left(\dfrac{\Gamma(e^\lambda)-\gamma(e^\lambda, xe^{\alpha e^{-\lambda}})}{\Gamma(e^\lambda)-\gamma(e^\lambda, (x+t)e^{\alpha e^{-\lambda}})}\right)$
Generalised Gamma	$\dfrac{e^\alpha x^{e^\lambda\sigma-1}\exp\left(-x^\sigma\left(\dfrac{e^\alpha}{\sigma}\right)^{e^{-\lambda}}\right)}{\Gamma(e^\lambda)-\gamma\left(e^\lambda, x^\sigma\left(\dfrac{e^\alpha}{\sigma}\right)^{e^{-\lambda}}\right)}$	$\log\left(\dfrac{\Gamma(e^\lambda)-\gamma\left(e^\lambda, x^\sigma\left(\dfrac{e^\alpha}{\sigma}\right)^{e^{-\lambda}}\right)}{\Gamma(e^\lambda)-\gamma\left(e^\lambda, (x+t)^\sigma\left(\dfrac{e^\alpha}{\sigma}\right)^{e^{-\lambda}}\right)}\right)$

As an example of the second principle, consider the definitions of the Lognormal and Inverse Gaussian distributions in Table 2, which both use $+\alpha$ where Lindgren (1976) and Collett (2003) use the more typical $-\mu$. The definitions in Table 2 mean that a higher value of α means an increase in risk.

To illustrate a combination of the first and third principles, consider the hazard function for the Weibull distribution given by Collett (2003):

$$h(t) = \lambda\gamma t^{\gamma-1} \tag{9}$$

where $\lambda > 0$ and $\gamma > 0$. Collett (2003) refers to λ as the scale parameter, and γ as the shape parameter, two terms we will return to in Section 4. We can merge the first two parameters in Eq. (9) into a single replacement parameter. If we make this new parameter exponentiated, we can also drop the first positivity constraint. Doing this also eliminates the need for the positivity constraint on the second parameter, so we can simplify the Weibull hazard thus:

$$h(t) = e^{\alpha} t^{\sigma-1} \tag{10}$$

where both α and σ are free to vary along the real line. Equation 10 is also an example of the fourth principle: when $\sigma = 1$, this definition of the Weibull model becomes the same as Eq. (8), since the Exponential distribution is a special case of the Weibull distribution.

In Table 2, IGS() is the survivor function for the Inverse Gaussian lifetime and is defined as follows:

$$\mathrm{IGS}(x) = \Phi\left((1 - xe^{\alpha})\sqrt{\frac{e^{\sigma}}{x}}\right) - \exp(2e^{\sigma+\alpha})\Phi\left(-(1 + xe^{\alpha})\sqrt{\frac{e^{\sigma}}{x}}\right) \tag{11}$$

where $\Phi()$ denotes the cumulative distribution function for a $N(0,1)$ variable. $\gamma()$ denotes the incomplete gamma function, defined as:

$$\gamma(e^{\lambda}, x) = \int_0^x e^{-s} s^{e^{\lambda}-1} ds \tag{12}$$

for any real-valued parameter λ with $x > 0$, and $\Gamma(e^{\lambda}) = \lambda(e^{\lambda}, \infty)$. This is not the usual definition of the incomplete gamma function, but Eq. (2) is consistent with the desire to avoid constraints on λ.

Note that the extreme-value distribution is the Gompertz model and the Logistic distribution is a special case of the Beard model. Richards (2008) gives worked equivalences for these.

4. Parameter naming convention

A scale parameter, σ, is normally defined as one which satisfies the following:

$$F(t; \sigma, \theta) = F(t/\sigma; 1, \theta) \tag{13}$$

where F is the cumulative distribution function for the probability distribution and where θ denotes one or more other parameters. Since the survivor function, ${}_tp_x = 1 - F_x(t)$, from Eq. (4) this is the same thing as saying:

$$H_x(t; \sigma, \theta) = H_x(t/\sigma; 1, \theta) \tag{14}$$

However, this definition is not adhered to universally. For example, Collett (2003) refers to λ in Eq. (9) as a scale parameter, but it does not have the property of σ in Eqs. 13 or 14. Similarly, the documentation for the `survreg` function in the `survival` library in

R Development Core Team (2004) shows that the terms intercept, scale and shape parameter are used very loosely within the same software system:

> There are mulitiple ways to parametrize a Weibull distribution. The `survreg` function inbeds (sic) it in a general location-scale family, which is a different parameterization than the `rweibull` function, and often leads to confusion.
>
> `survreg`'s scale = 1/(`rweibull` shape)
>
> `survreg`'s intercept = log(`rweibull` scale)
>
> R documentation, v2.10.0

With this inconsistency elsewhere, it is forgivable that we restructure the parameterisations in Table 2 according to the four principles in Section 3, and that we refer to occurrences of σ in Table 2 as being the scale parameter and occurrences of λ as the shape parameter. Our naming convention is therefore defined in Table 3.

There are other ways to parameterise these models. For example, Vanfleteren *et al.* (1998) use a different parameterisation for the Log-Logistic and Beard models because, as biologists, they are interested in a real-world biological interpretation for the parameters. Indeed, biologists use models for the hazard function because lifetimes are often measured in hours or days: in describing their experiments on *Caenorhabditis elegans*, Vanfleteren *et al.* (1998) called it a 'small worm [. . .] with a life span of 2–4 weeks, depending on culture conditions'. Using one-year mortality rates like q_x is too anthropocentric for many biological models. This point does still have some relevance to actuaries, even though they are usually only concerned with human lives. Some classes of business have such short life expectancies as to question whether one-year q_x models are sensible – for example, care annuities typically only have an average duration of 2 or 3 years.

Table 3. Naming convention for parameters used in Tables 1 and 2.

Parameter	Name
α	`Intercept`
β	`Age`
ε	`Makeham`
ρ	`Beard`
σ	`Scale`
λ	`Shape`

5. Actuarial mortality laws

Four of the actuarial mortality laws from Table 1 are plotted in Figure 4. The Gompertz (1825) law is the simplest mortality law allowing for age-related increases in mortality. It specifies an exponentially increasing hazard with age, i.e. a straight line on a logarithmic scale. The Gompertz law works well in the age range 60–90, but at higher ages it usually overstates mortality rates, while at lower ages it typically under-states mortality. The

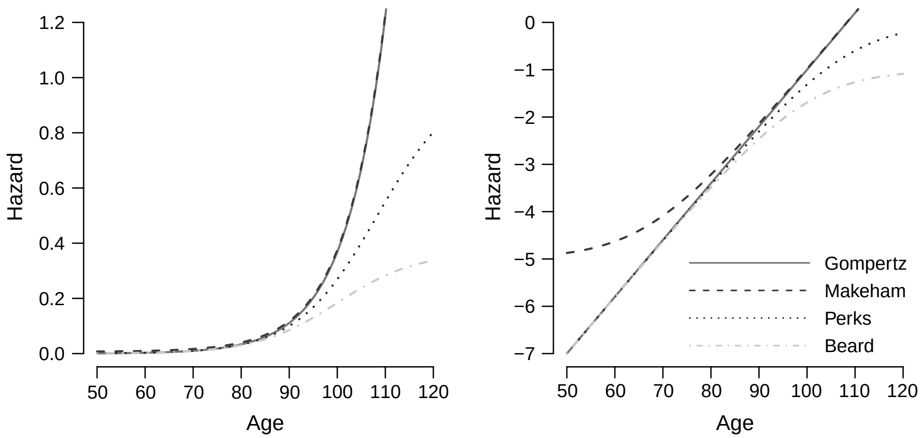

Figure 4. Hazard functions for Gompertz, Makeham, Perks and Beard mortality laws in Table 1 with $\alpha = -13$, $\beta = 0.12$, $\rho = 1$ and $\varepsilon = -5$. Natural scale (left) and Logarithmic scale (right).

Makeham (1859) law is similar, but with a constant, non-age-related element to mortality. This typically finds application below age 60 or so where the exponential pattern of the Gompertz law usually fails to hold. The Perks (1932) law has the same number of parameters as the Gompertz law, but the logistic form of the hazard curve allows for a slower-than-exponential increase in mortality at advanced ages. The Beard (1959) law is similar to the Perks law, but the extra ρ parameter allows for greater variation in the rate of change at advanced ages.

When fitted to actual mortality data, these laws will typically fail to fit well outside the range 60–90 for one reason or another. Above age 90 we normally see a slowdown in the rate of increase, called *late-life mortality deceleration* (Gavrilov & Gavrilova (2001)), which militates against the Gompertz and Makeham laws. Equally, pensioner mortality below age 60 typically does not decrease exponentially with reducing age either, thus invalidating the Gompertz, Perks and Beard laws. When working with a wide age range, say 50–110, we need a law which encompasses the behaviour of the Makeham law below age 60 and logistic behaviour above age 90. It is for this reason that the Makeham–Perks and Makeham–Beard laws typically work best of all the mortality laws.

Note that an alternative to the Makeham–Perks definition in Table 1 would be to use the following:

$$\mu_x = e^{\varepsilon} + \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \tag{15}$$

In practice, we find that the fits for the definition in Eq. (15) are identical to those using the definition in Table 1. However, since the definition in Eq. (15) usually takes more iterations to converge, we prefer the definition in Table 1. A similar comment applies to the definition of the Makeham–Beard law in Table 1, with the added benefit that this definition can arise both through heterogeneity arguments for the Makeham law

(Horiuchi & Coale (1990)) and also from viewing mortality as a cascade process (Richards (2008)).

6. Comparison of models

In this section we consider two measures of how well a mortality law fits compared to others. The first measure is the Akaike's Information Criterion (AIC) (Akaike (1987)), which balances the value of the log-likelihood function with the number of parameters used in the model. A lower AIC value is typically a better model. However, this is not quite the same thing as goodness of fit, so we make use of a result from Cox & Miller (1987), namely that the number of deaths observed in a group has a Poisson distribution with the Poisson parameter set to the sum of the integrated hazard functions. We therefore consider the goodness of fit for the most important risk factor, age, by comparing the total number of deaths at each integer age x with the sum of the integrated hazard functions over the range $[x, x+1)$. We can either use this data to calculate a Poisson deviance residual for visual inspection, or (as here) we can calculate a χ^2 test statistic. Note that a formal test will fail all of these models because they do not use all the risk factors available. However, we will use the χ^2 test statistic as a means of broadly comparing how well (or otherwise) the model fits the pattern of mortality by age.

Due to their fundamentally different structures, it is not possible to fit the same model for each mortality law or lifetime distribution. For example, the Exponential and Pareto distributions do not have flexibility in how mortality changes by age. Similarly, the accelerated failure-time distributions have a scale parameter, σ, in place of the coefficient of ageing, β, in the actuarial mortality laws. In Table 4 we therefore fit models which are as similar as can be: a single parameter for age-related variation (where possible) and a single constant parameter for gender differentials. In practical actuarial work, of course, we would use many more risk factors, as described in Richards (2008).

As shown in Table 4, the Makeham–Beard model fits best, regardless of whether this is measured by the AIC or the χ^2 statistic. Using the same data set with further risk factors for pension size and postcode-driven lifestyle group, and with age interactions, Richards (2008) also found the Makeham–Beard law to fit best. The laws with logistic-shaped hazards (Perks, Beard, Makeham–Perks and Makeham–Beard) are all materially better fits than the simpler Gompertz and Makeham laws.

The Exponential model with constant hazard fits very badly, as would be expected, while the Pareto model fits even worse due to the reducing hazard. Of the accelerated failure-time models, only the Weibull and Generalised Gamma models can be viewed as being useful, with a slightly better fit than the Gompertz and Makeham laws.

The Normal, Lognormal and Inverse Gaussian models are poor fits, in part due to the inconsistencies in hazard functions by age shown in Figures 12–14.

We have not shown any residual plots or tests for Table 4 as every model will fail due to patterns in the residuals. This is a result of the data set being large and powerful and there

Table 4. Comparison of fits for a basic model of mortality between 2000 and 2006 for annuitants aged 60–95.

Model specification and mortality law	Parameters	AIC	Improvement (worsening) over Gompertz	χ^2
Age + Gender:				
Gompertz	3	385,530	n/a	115
Perks	3	385,414	116	60
Age + Gender + Makeham:				
Makeham	4	385,532	(2)	115
Makeham–Perks	4	385,416	114	60
Age + Gender + Beard:				
Beard	4	385,375	155	72
Age + Gender + Makeham + Beard:				
Makeham–Beard	5	385,372	158	57
Gender:				
Exponential	2	427,492	(41,962)	40,421
Pareto	2	437,040	(51,510)	49,957
Scale + Gender:				
Weibull	3	385,525	5	108
Logistic	3	386,163	(633)	964
Normal	3	386,475	(945)	1,257
Log-Logistic	3	387,129	(1,599)	1,906
Lognormal	3	387,971	(2,441)	2,714
Inverse Gaussian	3	387,986	(2,456)	2,729
Shape + Gender:				
Gamma	3	387,384	(1,854)	2,143
Scale + Shape + Gender:				
Generalised Gamma	4	385,527	3	102

Note: For the accelerated failure-time models with a scale parameter, it is the age variable which has been transformed rather than duration since retirement. For each model the `Intercept` is implied and is not listed in the model specification.

being several significant risk factors which have not been included in the model. Interested readers should see Richards (2008) for handling of risk factors such as lifestyle, pension size and select period.

7. Variation by age

In Section 2 we discussed how left truncation was a problem for many implementations of the survival models in Table 2 and how actuaries got around this by working directly with the log-likelihood function in Eq. (6). Another feature of the scale-transformed survival models in Table 2 is that the scale parameter, σ, may need to vary by sub-group. The same applies to the age parameter, β, in Table 1. For example, Table 4 features models where the same value of β or σ applies equally to all lives in the portfolio, including both males and females. Picking the Gompertz `Age + Gender` model as an example, this assumes that the gender difference is constant on a logarithmic scale, i.e. that the ratio of male to female mortality rates is constant. However, Figure 5 shows how the ratio between male and female mortality in an annuity portfolio is not constant, i.e. the same value of σ (or β) cannot apply

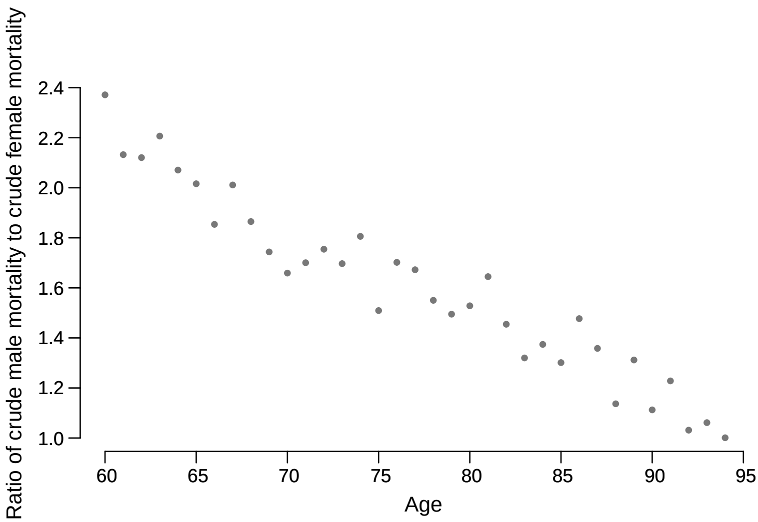

Figure 5. Ratio of crude mortality hazard for males to crude hazard for females in a large annuity portfolio. The excess of male mortality clearly diminishes with increasing age, i.e. the rates converge with age and an assumption of a constant proportion is invalid. Any survival model must therefore permit interactions, either with the β parameters in Table 1 or with the σ parameters in Table 2.

to both males and females. This is a common feature of most risk factors in actuarial work, so any model forcing the same value of β or σ across different groups will be sub-optimal.

Nevertheless, many software implementations of the survival models in Table 2 do indeed force the same scale parameter across all lives. One solution to this is to fit separate models for the various sub-groups in a portfolio. A better solution is to work directly with the log-likelihood in Eq. (6), which permits different values of β or σ for different sub-groups in the same unitary model. Working directly with the log-likelihood achieves three major advantages for actuaries: first, it greatly expands the range of survival models which can be fitted; second, it handles the left truncation which is a feature of all insured data; and third, it permits age-varying risk factors. The latter point is of great importance to actuaries because of the time value of money.

Table 5 shows the parameter estimates and standard errors for a Weibull model for the data behind Figure 5. It shows statistically significant excess male mortality in the positive value of the Gender.M parameter (females are the baseline). It also shows a statistically significant difference in the scale parameter for males in the Gender.M:Scale parameter: in effect, the value of σ for males is 1.39285 lower than the value for the female baseline of $\sigma = 10.7298$.

Table 6 shows the parameter estimates for an equivalent Perks model. The Perks model allows for variation by age in a different way to the Weibull model, i.e. through varying the age coefficient, β, by sub-group instead of varying the scale parameter, σ. The Perks model fits better, with an AIC 77 units lower than for the Weibull model in Table 5, and a χ^2 statistic which is 69 units lowers.

The model in Table 5 contains all interactions, so the same result could have been reached by splitting the portfolio into males and females and fitting a separate model to

Table 5. Parameters for Weibull model with `Scale*Gender`, AIC = 385,335 and $\chi^2 = 157$.

Parameter	Estimate	Standard error	Z-value	p-value
Intercept	−45.8629	0.3717	−123.38	0
Scale	10.7298	0.08519	125.95	0
Gender.M	6.491	0.4403	14.74	0
Gender.M:Scale	−1.3928	0.1010	−13.80	0

each. However, actuarial models are often richer than this and contain many more risk factors, such as pension size, lifestyle, birth cohort and select period. Often the best-fitting model is not one containing all the interactions, so it is preferable to fit models without sub-dividing the data set.

One irony in Tables 5 and 6 is that while the `Gender.M:Age` interaction is a significant parameter, and while the AIC has improved materially compared to the equivalent models in Table 4, the χ^2 statistics have actually worsened. This is not a major source of concern for such a simple model, as adding further risk factors to the models will reduce both the AIC and the χ^2 statistics compared to Table 4.

Table 6. Parameters for Perks model with `Age*Gender`, AIC = 385,258 and $\chi^2 = 88$.

Parameter	Estimate	Standard Error	Z-value	p-value
Intercept	−13.7851	0.09147	−150.71	0
Age	0.1320	0.001165	113.38	0
Gender.M	1.8162	0.1088	16.70	0
Gender.M:Age	−0.01736	0.001388	−12.51	0

8. Simulation

Modern portfolio management demands simulation of future assets and liabilities. Another advantage of survival models over q-type models is that it is often easy to simulate the future lifetime of an individual, thus making whole-portfolio simulations very fast. For example, by inverting Eq. (4) it is often possible to find a closed-form expression for the simulated future lifetime, t, of a life currently aged x. Table 7 lists closed-form expressions according to some of the mortality laws in Table 1 or distributions in Table 2.

For the Makeham, Makeham–Perks and Makeham–Beard laws it is possible to solve Eq. (2) in a few iterations using a Newton–Raphson algorithm. A useful choice of starting value is the formula in Table 7 for the equivalent law lacking the ε term; thus, the Gompertz formula in Table 7 provides a good initial value for iterating the Makeham law.

One other benefit of implementing such simulations lies in checking the model-fitting algorithms. Simulated data can be used to ensure that the simulation code and the model-fitting code are at least consistent.

Table 7. Formulae for simulating future lifetime, t, given current age x.

Law or distribution	Formula
Gompertz	$\dfrac{\log\left(1-\frac{\beta}{e^{\alpha+\beta x}}\log U\right)}{\beta}$
Perks	$\dfrac{\log(\exp(-\beta\log U+\log(1+e^{\alpha+\beta x}))-1)-\alpha}{\beta}-x$
Beard	$\dfrac{\log(\exp(-\beta e^{\rho}\log U+\log(1+e^{\alpha+\rho+\beta x}))-1)-\alpha-\rho}{\beta}-x$
Exponential	$\dfrac{\log U}{e^{\alpha}}$
Pareto	$\begin{cases}\exp\left(\frac{-\log U}{e^{\alpha}}\right), & x=0;\\ x\exp\left(\frac{-\log U}{e^{\alpha}}\right)-x, & \text{otherwise}\end{cases}$
Weibull	$\exp\left(\dfrac{\log\left(x^{\sigma}-\frac{\sigma}{e^{\alpha}}\log U\right)}{\sigma}\right)-x$
Logistic	$e^{\sigma}\left(\log\left(1+\exp\left(\frac{x+\alpha}{e^{\sigma}}\right)-U\right)-\log U\right)-x-\alpha$
Normal	$e^{\sigma}\Phi^{-1}\left(1-U(1-\Phi\left(\frac{x+\alpha}{e^{\sigma}}\right))\right)-x-\alpha$
Log-Logistic	$\exp\left(\dfrac{\log\left(\frac{1+e^{\alpha}x^{e^{\sigma}}}{U}-1\right)-\alpha}{e^{\sigma}}\right)-x$
Lognormal	$\begin{cases}\exp(e^{\sigma}\Phi^{-1}(U)-\alpha), & x=0;\\ \exp\left(e^{\sigma}\Phi^{-1}\left(1-U\left(1-\Phi\left(\frac{\log x+\alpha}{e^{\sigma}}\right)\right)\right)-\alpha\right), & \text{otherwise.}\end{cases}$

Note: U is a random number distributed evenly over (0, 1).

9. Non-parametric survival analysis

All the survival models fitted in this paper are parametric, which yields automatically smooth curves and obviates the need for a separate stage of graduation (smoothing). An alternative is non-parametric survival analysis, which can be used as a check on the reasonableness of the fitted parametric curves. An example of this was introduced by Kaplan & Meier (1958). One wrinkle for actuaries is that the standard Kaplan–Meier approach is typically defined with reference to the time since a medical study commenced. In actuarial work is makes more sense to define the non-parametric survival curve with respect to age attained rather than duration observed. The following definition will work for any portfolio whether it is closed or open to new business:

$$_{t_j}p_x = \prod_{i=1}^{j \le n} \left(1 - \frac{d_{x+t_i}}{l_{x+t_i^-}}\right) \tag{16}$$

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 $x+t_i$ and d_{x+t_i} is the number of deaths dying at age $x+t_i$. An example of this is given in Figure 6.

Note that the Kaplan–Meier curve definitely falls within the framework of a survival model, but conceptually it straddles the concepts of q_x and μ_x. The definition in Eq. 16 is clearly based around q_x, but the discretisation is decided by the data *a posteriori*, rather than by the analyst *a priori*. As the number of events in life-office portfolios is typically large, the discretisation steps can be quite small and the results quickly look like μ_x due to the relationship in Eq. (1). For example, the median age gap between deaths for the male lives in Figure 6 is one day, which is the smallest interval possible when using dates to measure survival times (the largest gap is 11 days). For this reason a purist might argue that the models fitted in this paper are actually for q_x with a daily interval, i.e. $_{\frac{1}{365}}q_x$, rather than for μ_x.

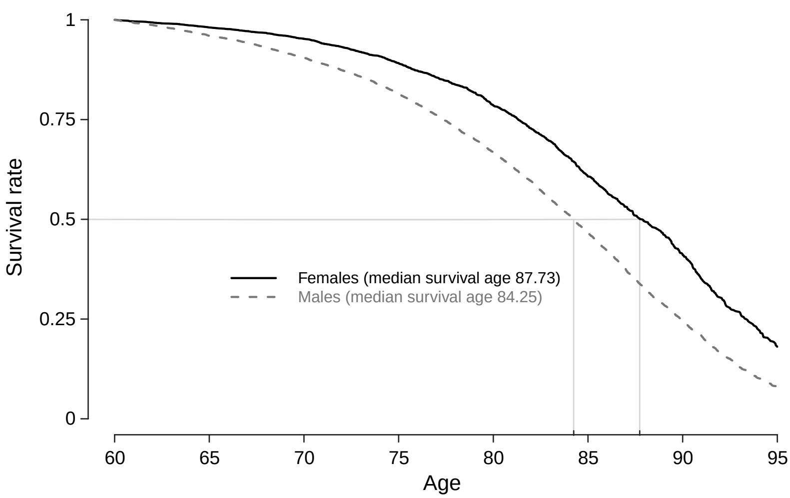

Figure 6. Kaplan–Meier survival curves for annuitants in a large UK portfolio, as per Eq. (16).

10. Conclusions

There is a wide choice of survival models available for modelling pensioner mortality. A particular feature which actuaries require is the ability to handle left-truncated data, since holders of life-assurance contracts typically enter observation well into adult life. Most models commonly available in standard software packages do not cater for left-truncated data, so actuaries tend to work directly with the log-likelihood function to fit their models. A further benefit of this is the ability to have parameters for age-varying risk to be set seperately for each sub-group. However, even after restructuring the fitting algorithms for left truncation and age-varying scale parameters, the commonly available survival models still fit less well than the mortality 'laws' documented by actuaries and demographers over 50 years ago. Survival models also make run-off simulations of portfolios both straightforward and fast.

Acknowledgements

The author thanks Dr. Iain Currie, Professor Angus Macdonald and Andrew Howe for helpful comments. Any errors or omissions remain the sole responsibility of the author. Data validation and preparation were done using Longevitas, which was also used to fit all the models listed. Graphs were done in R.

References

Aitken, M., Anderson, D., Francis, B. & Hinde, J. (1989). *Statistical modelling in GLIM* (pp. 283–285). Oxford: Oxford University Press.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrica* **52**, 317–333.

Beard, R. E. (1959). Note on some mathematical mortality models. In G. E. W. Wolstenholme & M. O'Connor (Eds.), *The lifespan of animals*, pp. 302–311. Boston, MA: Little Brown.

Chhikara, R. S. & Folks, L. (1989). *The inverse Gaussian distribution: theory, methodology, and applications*. New York: Marcel Dekker.

Collett, D. (2003). *Modelling survival data in medical research* (pp. 196–198). Boca Raton, FL: Chapman and Hall.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **24**, 187–220 (with discussion).

Cox, D. R. & Miller, H. D. (1987). *The theory of stochastic processes* pp. 153–154. London: Chapman and Hall.

Daw, R. H. (1951). Duplicate policies in mortality data. *Journal of the Institute of Actuaries* **77**, 261–267.

Gavrilov, L. A. & Gavrilova, N. S. (2001). The reliability theory of aging and longevity. *Journal of Theoretical Biology* **213**, 527–545.

Gompertz, B. (1825). The nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society* **115**, 513–585.

Horiuchi, S. & Coale, A. J. (1990). Age patterns of mortality for older women: an analysis using the age-specific rate of mortality change with age. *Mathematical Population Studies* **2** (4), 245–267.

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

Lindgren, B. W. (1976). *Statistical theory* (3rd ed, p. 190). London: Collier Macmillan.

Makeham, W. M. (1859). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries* **8**, 301–310.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries* **63**, 12–40.

R Development Core Team. (2004). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available online at: http://www.r-project.org

Richards, S. J. (2008). Applying survival models to pensioner mortality data. *British Actuarial Journal* **14**.

Richards, S. J. & Currie, I. D. (2009). Longevity risk and annuity pricing with the Lee-Carter model. *British Actuarial Journal* **14** (2), 257–326 .

Vanfleteren, J. R., De Vreese, A. & Braeckman, B. P. (1998). Two-parameter Logistic and Weibull equations provide better fits to survival data from isogenic populations of *Caenorhabiditis elegans* in axenic culture than does the Gompertz model. *Journal of Gerontology Series A* **53**, B393–B403.

Appendix A. Relationships between mortality laws and lifetime distributions

To demonstrate that the Logistic model is a special case of the Beard law, consider first the Beard hazard function in Table 1:

$$\mu_x = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\rho+\beta x}} \tag{17}$$

If we set $\alpha' = \dfrac{\alpha+\rho}{\beta}$, $\beta = e^{-\rho}$ and $\sigma = \rho$, then Eq. (17) can be rewritten as:

$$\mu_x = \frac{1}{e^{\sigma}\left(1 + \exp\left(-\dfrac{x+\alpha'}{e^{\sigma}}\right)\right)} \tag{18}$$

which we recognise as the Logistic hazard from Table 2. The Beard model appears in a different guise as the 'three-parameter logistic model' used by Vanfleteren *et al.* (1998), which gives the hazard function at age x (in days) as:

$$\mu_x = \frac{ab}{a + (b-a)e^{-kx}} \tag{19}$$

where a and b are positive and k is real-valued. Rearranging Eq. (19) we get:

$$\mu_x = \frac{\exp\left(\log\left(\dfrac{ab}{b-a}\right) + kx\right)}{1 + \exp\left(\log\left(\dfrac{a}{b-a}\right) + kx\right)} \tag{20}$$

and setting $\alpha = \log\left(\frac{ab}{b-a}\right)$, $\beta = k$ and $\rho = -\log b$ we get the Beard law again. The relationships between the 16 models in this paper are depicted in Figure 7.

The Exponential distribution involves a constant hazard, which is only useful for relatively short age ranges in actuarial work. Since mortality rates vary widely by age, we will not consider the Exponential distribution in great detail, other than to note that it is the Weibull distribution with $\sigma = 1$.

The Pareto distribution is the Weibull distribution with $\sigma = 1$. It involves a decreasing hazard, however, as shown in Figure 8. This is unlikely to be suitable for most types of mortality work, as shown by the increasing hazard in Figure 3, although it may find limited application in specialist business areas. One example might be annuities written on impaired lives, where one might expect very high initial rates of mortality decreasing after the contract commences. In such instances, the variable used in the hazard function would not be x, the increasing annuitant age, but r, the increasing duration from contract outset. Figure 8 therefore has a horizontal axis labelled with both age and time since outset, depending on how the Pareto distribution is defined. This choice of using age or duration is an option with other distributions.

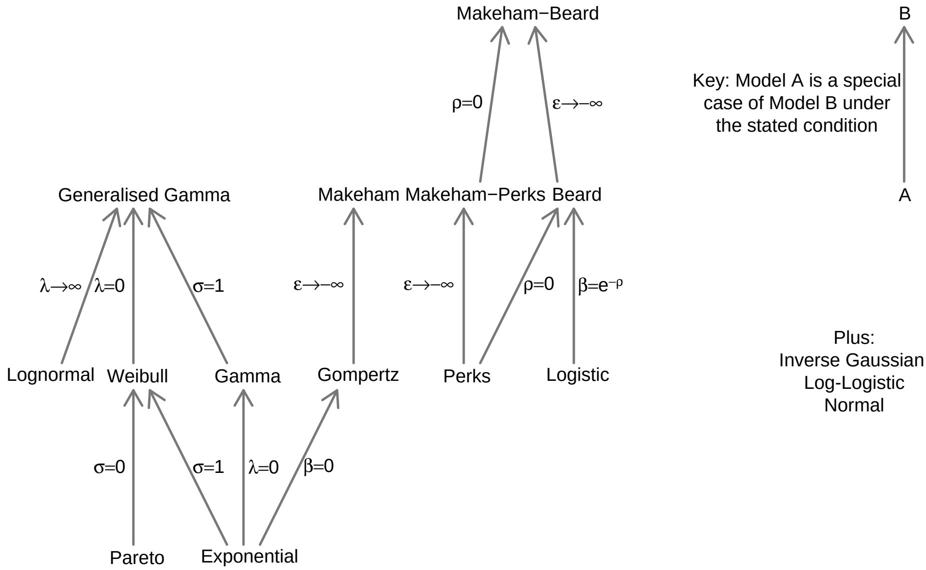

Figure 7. Relationships between the models defined in Tables 1 and 2. Models on the same horizontal level have the same number of basic parameters.

The Exponential and Pareto distributions are included here for completeness, but we expect them to perform very badly in comparison with models which allow for increasing mortality with age.

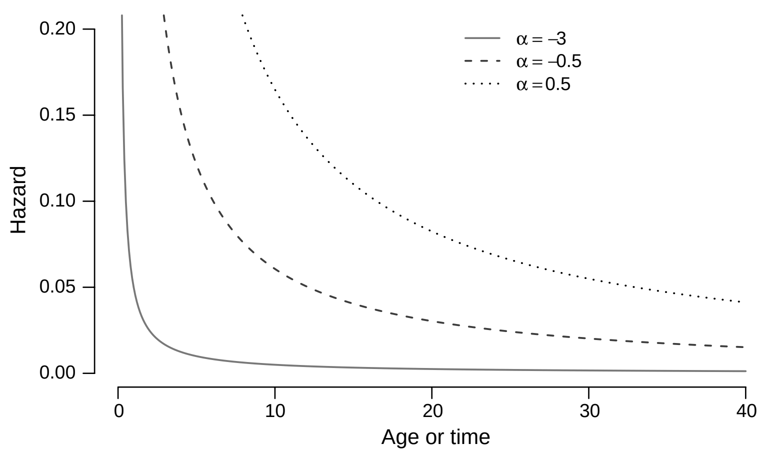

Figure 8. Hazard functions for a Pareto distribution defined in Table 2 with $\alpha = -3,\ -0.5$ and 0.5.

The Weibull distribution arises as a power transformation of the Exponential distribution. Along with the Lognormal, Log-Logistic, Gamma and Inverse Gaussian models it is known as an *accelerated failure-time distribution* (Collett (2003)). These distributions handle age-related changes in mortality in a different way from the actuarial mortality laws. Accelerated failure-time distributions have a *scale parameter*, σ, which does a similar job to the β parameter in the actuarial laws: both allow mortality rates to change with age.

If $\sigma = 0$ in the Weibull model then we have the special case of the Pareto model, while if $\sigma = 1$ in the Weibull model then we have the special case of the Exponential model. Figure 9 shows that the hazard function can replicate the exponentially increasing mortality typically seen at pensioner ages in Figure 3.

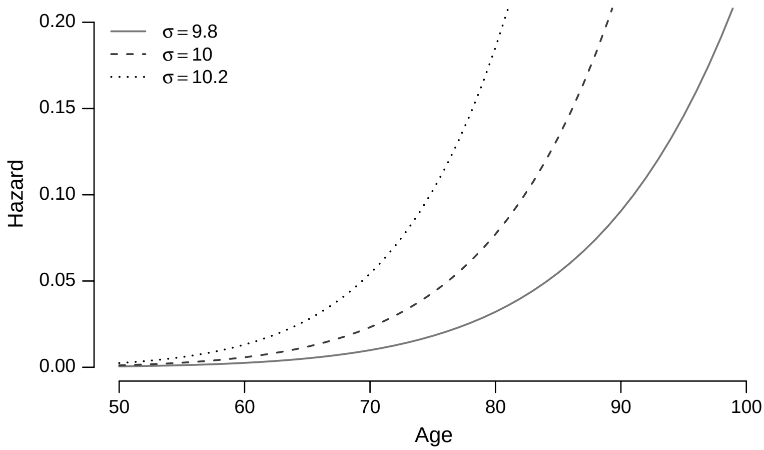

Figure 9. Hazard functions for a Weibull distribution defined in Table 2 with $\alpha = -42$ and $\sigma = 9.8$, 10 and 10.2. Varying α will simply scale the curves and will not change their basic shape or relationship to each other.

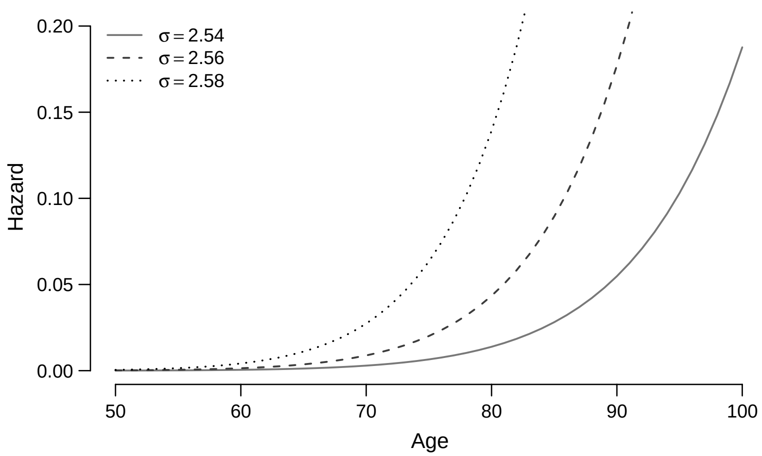

Figure 10. Hazard functions for a Logistic distribution defined in Table 2 with $\alpha = -59$ and $\sigma = 2.54$, 2.56 and 2.58.

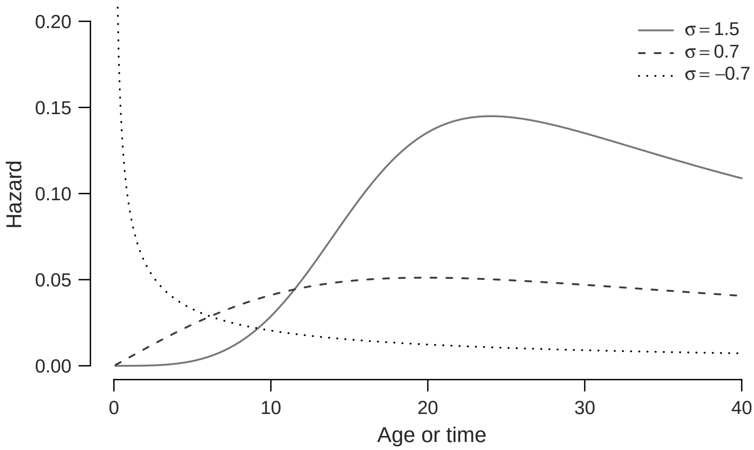

Figure 11. Hazard functions for a Log-Logistic distribution defined in Table 2 with a constant median and $\sigma =$ 16, 0.7 and -0.7. Styled after a similar graph in Collett (2003).

The Logistic distribution can produce exponentially increasing hazard rates, depending on the value of σ, as shown in Figure 10. Note that the Logistic distribution in Table 2 is a special case of the Beard law of mortality in Table 1, so the curves for the Beard law in Figure 4 also apply.

The Log-Logistic distribution yields a wide variety of hazard shapes, as shown in Figure 11. These shapes tend not to be appropriate for ordinary pensioner mortality, but they might be suitable for certain types of impaired-life annuities or care annuities.

The 'two-parameter' logistic distribution used by Vanfleteren *et al*. (1998) is the same as the Log-Logistic distribution. Vanfleteren *et al*. (1998) give the hazard function at age x in days, μ_x, as:

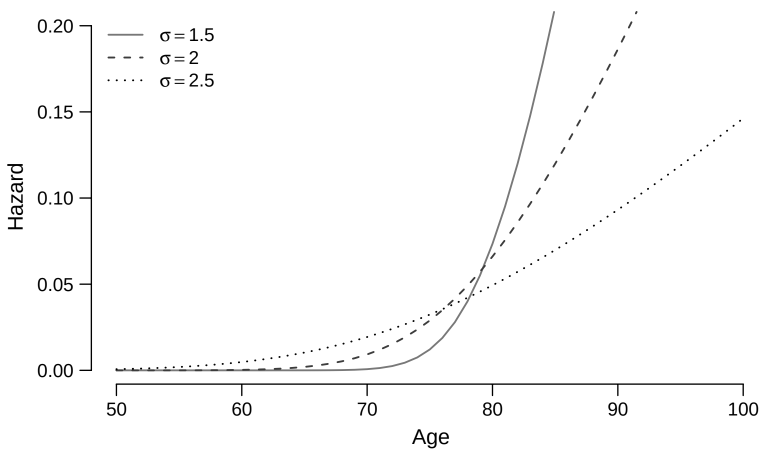

Figure 12. Hazard functions for a Normal distribution defined in Table 2 with $\alpha = -84$ and $\sigma = 1.5$, 2 and 2.5.

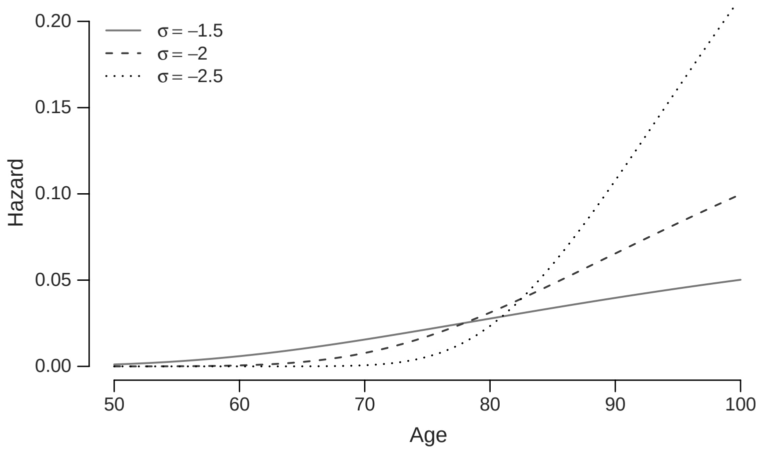

Figure 13. Hazard functions for a Lognormal distribution defined in Table 2 with $\alpha = -4.5$ and $\sigma = -1.5$, -2 and -2.5.

$$\mu_x = \frac{bx^{b-1}}{c^b + x^b} \tag{21}$$

for $c > 0$ and b real-valued. We can rearrange Eq. (21) as follows:

$$\mu_x = \frac{\exp(\log b - b\log c)x^{b-1}}{1 + \exp(-b\log c)x^b} \tag{22}$$

If we take Eq. (21) and set $\alpha = -b \log c$ and $e^{\sigma} = b$, we get the Log-Logistic hazard in Table 2.

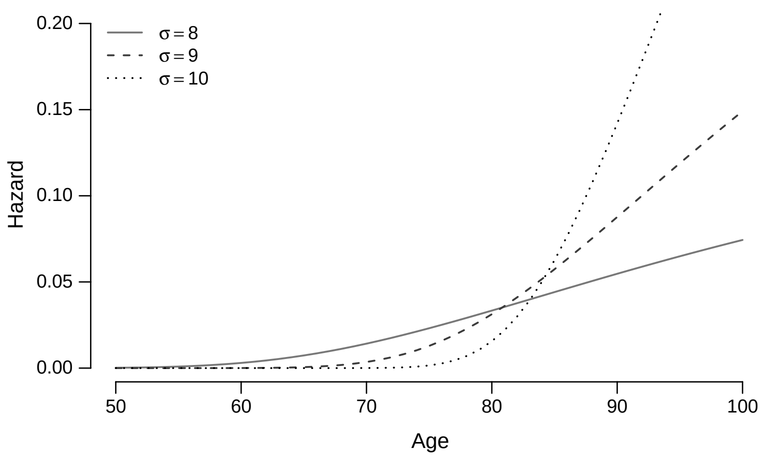

Figure 14. Hazard functions for an Inverse Gaussian distribution defined in Table 2 with $\alpha = -4.5$ and $\sigma = 8$, 9 and 10.

The Normal distribution can yield different rates of increase in risk, depending on the value of σ. However, it does not have the property of consistency that is exhibited by other models. For example, in Figure 12 the relationship between the three hazard curves at age 70 is completely reversed by age 80.

A Lognormal distribution for the lifetime of an individual, T, arises from the assumption that $\log T$ has a Normal or Gaussian distribution. Although this assumption is simply explained, the formula for the hazard in Table 2 is not particularly simple. However, as with the Normal model, the Lognormal also does not have the property of consistency that is exhibited elsewhere. For example, in Figure 13 the relationship between the three hazard curves at age 70 is completed reversed by age 90.

The Inverse Gaussian distribution has a number of important properties – see Chhikara & Folks (1989). In practice, however, it offers similar hazard shapes to the Lognormal distribution – compare Figure 14 with Figure 13. However, as with the Lognormal model, it also does not have the property of consistency that is exhibited elsewhere. For example, in Figure 14 the relationship between the three hazard curves at age 70 is completely reversed by age 90. In terms of implementation there are few software packages which offer the Inverse Gaussian model, not least because it has the most complicated hazard function of all the models in Table 2.

In contrast to most of the other models, neither of the two parameters in the Gamma distribution obviously sets the general level of mortality. The hazard function in Table 2 contains both the power of x (as per the Weibull model) and the scaled exponent of x (as per the Logistic model). As Figure 15 shows, varying either of the parameters will have roughly the same effect on the rate at which the hazard increases with age. The choice of which parameter is to be labelled α and which λ is therefore somewhat arbitrary. However, here we have chosen a definition such that when $\lambda = 0$ the parameter values for α will be the same as for the Exponential distribution, which is a special case of the Gamma

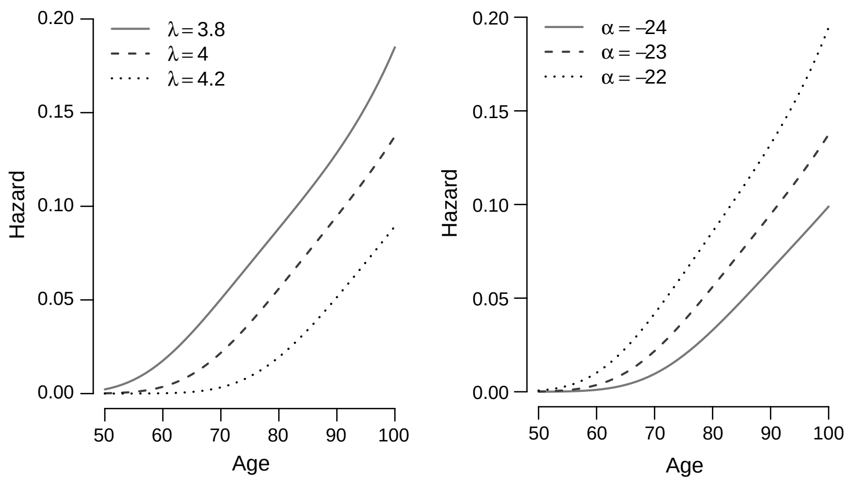

Figure 15. Hazard functions for a Gamma distribution defined in Table 2 with (left) $\alpha = -23$ and $\lambda = 3.8$, 4.0 and 4.2 and (right) $\alpha = -27$, -23 and -19 and $\lambda = 4$.

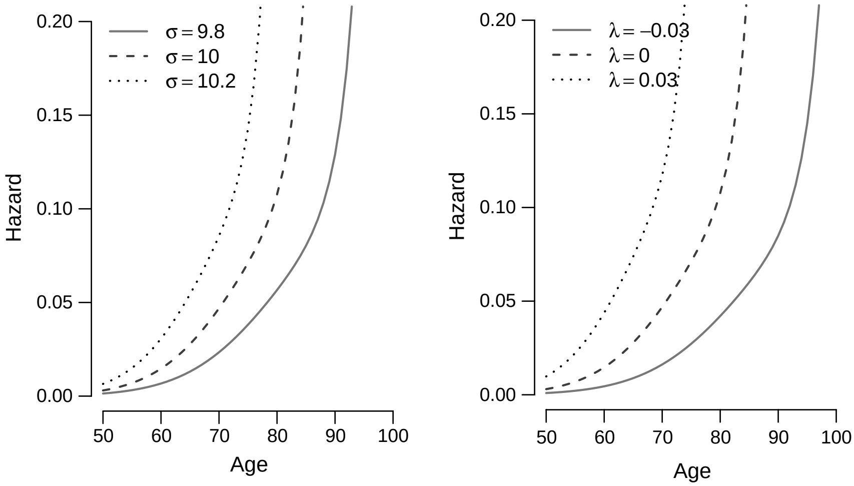

Figure 16. Hazard functions for a generalised Gamma distribution defined in Table 2 with $\alpha = -4.5$ and $\sigma = 8$, 9 and 10.

distribution. Note that if $\lambda > 0$ the Gamma hazard increases monotonically, while if λ is less than zero the hazard decreases monotonically.

As with the Gamma distribution, there is no simple parameter which sets the overall level of the hazard function of the Generalised Gamma. Figure 16 shows that varying σ and λ can have very similar effects on the rate at which the hazard increases with age. We note that the Generalised Gamma model is sometimes used as a means of choosing between alternative distributions, since three of the distributions listed in Table 2 are special cases of the Generalised Gamma distribution. We therefore choose a parameterisation consistent with them.

Using the parameterisation of the Generalised Gamma in Table 2, if $\lambda = 0$ we get the same definition as the Weibull distribution. Thus, if λ is not significantly different from zero, then a Weibull distribution might be more appropriate. The parameterisation of the Generalised Gamma in Table 2 has been chosen such that when $\lambda = 0$ the parameterisation is identical to that of the Weibull distribution.

Similarly, as $\lambda \to \infty$ then the Generalised Gamma distribution becomes the Lognormal distribution. Thus, if λ is large and yet not significant, then a Lognormal distribution might be more appropriate. Finally, if $\sigma = 1$ the Generalised Gamma simplifies to the ordinary Gamma distribution. Thus, if σ is not significantly different from 1 a Gamma distribution would be more appropriate.