ON CONTEMPORARY MORTALITY MODELS
FOR ACTUARIAL USE

By S. J. Richards & A.S. Macdonald

LONGEVITAS™

On Contemporary Mortality Models for Actuarial Use I: Practice

Richards, Stephen J.*and Macdonald, Angus S.†

July 30, 2024

Abstract

Actuaries must model mortality to understand, manage and price risk. Continuous-time methods offer considerable practical benefits to actuaries analysing portfolio mortality experience. This paper discusses six categories of advantage: (i) reflecting the reality of data produced by everyday business practices, (ii) modelling rapid changes in risk, (iii) modelling time- and duration-varying risk, (iv) competing risks , (v) data-quality checking and (vi) management information. Specific examples are given where continuous-time models are more useful in practice than discrete-time models.

Keywords: survival models, mortality shocks, selection effects, late-reported deaths.

1 Introduction

Actuaries must understand the mortality characteristics of portfolios they work on, and a key goal of any model is that it must reflect reality. This has two facets. First, a model must not make assumptions that contradict the data. In particular, the analyst should not have to discard data because it is inconvenient for the model; a model should match the data, not the other way around. Second, a model should be capable of reproducing real-world features of the risk. As we will see, continuous-time methods have specific advantages over discrete-time models in both respects.

In their landmark paper, Forfar et al. [1988] considered three different metrics for modelling mortality: (i) the discrete-time probability of death, q_x, (ii) the central mortality rate, m_x, and (iii) the continuous-time hazard rate, μ_x. Of the three, m_x has fallen somewhat into disuse, in part because it does not arise as a parameter in a statistical model. The probability q_x retains an important pedagogical role for students — there is nothing simpler than the idea of mortality viewed as a binomial trial over a fixed period like a year. However, this simplicity often collides with the messier reality of business processes. This paper explores the practical reasons why μ_x is usurping q_x in actuarial work. The companion paper, Macdonald and Richards [2024], provides the theoretical underpinning.

The plan of the rest of this paper is as follows. In Section 2 we state assumptions and define some terms. Section 3 compares the ability of discrete- and continuous-time models to represent the data produced by various business processes. Section 4 looks at a variety of mortality features that require greater granularity than a one-year, discrete-time view of risk. Section 5 illustrates the greater flexibility with investigation period offered by continuous-time methods. Section 6 explores the analysis of competing decrements. Section 7 considers the uses of some non- and semi-parametric techniques for data-quality checking, while Section 8 discusses their application for real-time management information. Section 9 discusses the results and Section 10 concludes. Appendix A contains numerous

*stephen@longevitas.co.uk, Longevitas Ltd, 24a Ainslie Place, Edinburgh, EH3 6AJ, UK. http://orcid.org/0000-0001-6859-6347

†A.S.Macdonald@hw.ac.uk, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK, and the Maxwell Institute for Mathematical Sciences, UK. http://orcid.org/0000-0001-6742-1158

case studies based on actual experience. Appendix B details the difficulties that discrete-time models have in dealing with fractional periods of exposure. Appendix C considers the conditions necessary to avoid bias when discarding data. Appendix D provides a primer in non- and semi-parametric methods for actuaries, while Appendix E outlines some features of grouped-count models.

2 Assumptions and terminology

We assume that we are primarily interested in the analysis of mortality, although many of the comments apply to other decrements. Mortality is most naturally described in a continuous-time setting[1]. That said, the most granular measure of time recorded in an administration system is typically a calendar date. This means that "continuous time" is for practical purposes synonymous with "daily". However, while an individual can die on any date, this information may not always be available for analysis — Case Study A.10 describes an example where only the month of death was available, while in Case Study A.17 only the year of death was available (see Appendix A). Where the date of death is not known exactly, deaths are said to be *interval censored*; the event is only known to have occurred between two points. In Case Study A.10, the interval is a calendar month, whereas in Case Study A.17 it is a year.

An event is an occurrence at a point in time that is of financial interest, such as a death or policy lapse. Other events are deemed irrelevant[2]. Events can occur in any order consistent with physical or procedural possibility, e.g. a pension-scheme member cannot retire after dying. Simultaneous events are ruled out by assumption. Note that there is a distinction between an event occurring and when (or if) that event is reported to the insurer or pension administrator; see the discussion of late-reported deaths in Section 4.3.

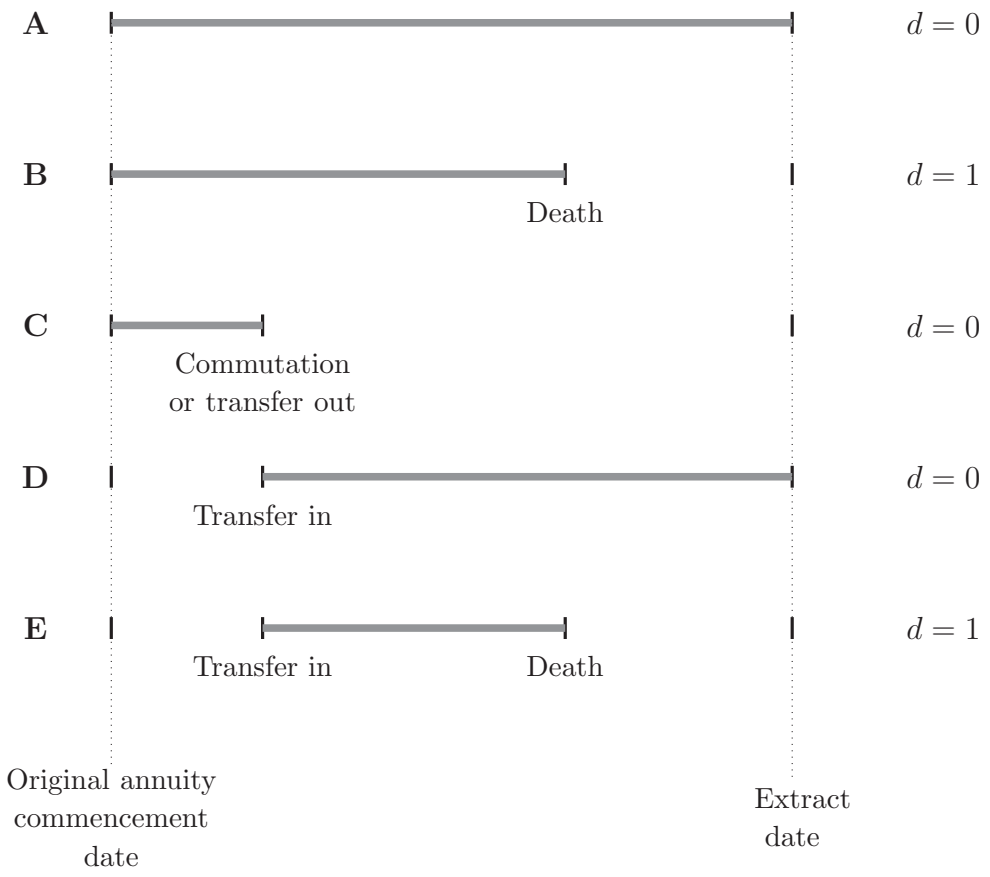
A discrete-time model for mortality assumes that the life in question exists at the start of a fixed period, such as one year, and is either alive ($d = 0$) or dead ($d = 1$) at the end of this period. No other event is permitted, such as an earlier exit. No information is available on the time of the event during the fixed period; we only know whether the life died, not when. Discrete-time models are typically parameterized by q or q_x. We will call such models 'q-type models'.

A continuous-time model for mortality assumes that the life in question is aged $x \geq 0$ at the start of observation and is observed for $t > 0$ years. At age $x + t$ the life is either dead ($d = 1$) or was alive when last observed ($d = 0$); see Cases A and B in Figure 1. If other events are accommodated by ceasing observation at age $x + t$, then the observation is regarded as censored with respect to the decrement of interest, i.e. death. A crucial distinction from discrete-time models is that we know when the event of interest took place. Continuous-time models typically denote the risk by μ or μ_x. μ_x was known to actuaries as the force of mortality, and to engineers as the failure rate. However, we prefer the term hazard rate, as used by statisticians [Collett, 2003, Section 1.3]. We will refer to such models as 'μ-type models'.

[1]Mortality may be represented by an unbroken chain of infinitesimal Bernoulli trials from one moment to the next. Macdonald and Richards [2024] describe this from an actuarial point of view

[2]Though we may deem other events to be irrelevant, meaning irrelevant for the application we have in mind, that does not mean they are ignorable. We must assume that they, like mortality, are generated by some mechanism(s); for example lapses might be governed by their own hazard rate. Then our aim can be formulated as follows: to estimate the parameters of interest (the mortality hazard) with minimal interference from other mechanisms in which we have no interest. For example, we would hope to be able to estimate the parameters of the mortality hazard rate without having to estimate anything to do with the lapse hazard rate. Strictly, we ought to construct a joint model of all events and justify such a procedure. Sometimes we are able to do so, as in a Markov model of state transition intensities [Macdonald et al., 2018, equation (15.21)], but often this is too difficult and we have to proceed without the comfort of any formal justification. See also Arjas and Haara [1984].

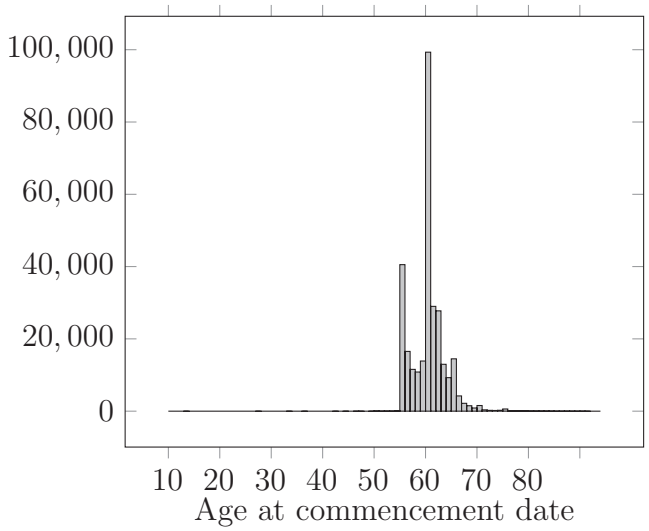
Figure 1: Timelines for various annuitant cases with usable exposure time, t, marked in grey. Case A is a survivor administered on the system from the annuity commencement date to the extract date. Case B is an annuity administered on the system from outset, but where death has occurred prior to the extract date. Case C represents an annuity administered on the system until either commutation, transfer to another system or transfer to another insurer. Cases D and E are annuities that were set up on another system or insurer at outset and subsequently transferred onto the current system. Cases with $d = 0$ are all right-censored with respect to mortality.



The most common type of censoring is where all that can be said is that death had not been reported at the last point of observation, known as *right-censoring*. Examples of right-censoring points include the date of extract, the date of transfer out or the date of trivial commutation; in each case, the life under observation was not reported to have died at the point of last observation. Cases A, C and D in Figure 1 are examples of right-censored observations. There are two sub-types of right-censoring: (i) 'obligate' right-censoring, where observation ceases at a point in time known in advance, and (ii) random right-censoring, where observation ceases at a point in time not known in advance. The distinction becomes important when the analyst decides to discard data; see Appendix C.

The start of an observation period for a life is qualitatively different from the end of the observation period. In actuarial work, most policyholders only become known to the portfolio well into their adult life; Figure 2 shows the distribution of age at commencement for annuities set up with a French insurer, where almost nobody enters observation before age 55 (an exception is shown in Figure 18). This is *left-truncation*, where the mortality of those not entering observation is by definition unknown. Left-truncation is a central feature of actuarial survival analysis, in contrast to, say, medical analysis [Collett, 2003]. Left-truncation arises naturally, such as when a policy first commences or retirement commences (Section 3.2). However, left-truncation can also be introduced by business processes (such as the system migrations discussed in Section 3.3) and deliberate analyst decisions (such as described in Section 5).

Figure 2: Histogram of age last birthday at annuity commencement date. Source: 298,906 annuities for a French insurer.



We assume that we have data on individuals. If we have data on benefits or policies, we will first need to create records for individual lives through a process of deduplication; see Macdonald et al. [2018, Section 2.5]. Deduplication is not merely a prerequisite for the independence assumption underlying statistical modelling, but it also results in a better understanding of risk and can even save money; see Case Study A.1 in Appendix A for an example. Note that having data on individual lives does not presuppose that we must model their risk individually, although this is desirable. However, extracting individual records does enable detailed validation and checking [Macdonald et al., 2018, Sections 2.3 and 2.4].

The above descriptions of q- and μ-type models assume the modelling of individuals for simplicity. However, the analyst also has the option to model grouped counts by allocating individuals to groups

and computing appropriate aggregate death counts and exposure measures; see Macdonald and Richards [2024, Appendix 1], especially Figure 4, for the details of the splitting of the individual observations and assigning them to rate intervals. The advantages and disadvantages of grouped-counts modelling are discussed in Macdonald et al. [2018, Section 1.7].

Time periods are often expressed in years. A practical issue is that the difference between two dates is a number of days, but calendar years do not always have the same number of days. Macdonald et al. [2018, p.38] standardise by dividing the number of days between two dates by 365.242 to account for leap years (Richards [2012, Table 15.1] gives a value of 365.24219 days for the astronomical cycle for the year 2000).

One question is how to handle dates at the start and end of a period of time. We adopt a policy of assuming that an interval starts at mid-day on the start date and ends at mid-day on the end date. For example, this means that the interval from 1st-31st January 2024 is taken to be 30 days long ($30 = 31 - 1$) and not 31 days. This allows us to subdivide intervals without double-counting, e.g. the interval from 1st-16th January 2024 is 15 days long and the interval from 16th-31st January 2024 is 15 days long. Intervals that start and end on the same date have zero length and are excluded from analysis. A death at the end date is assumed to happen at mid-day. Thus, if the exposure period is 1st-31st January 2024 and $d = 1$, then we have 30 days of exposure terminated with a death.

3 Reflecting reality: business processes

A fundamental assumption of a q-type model is that we should be able to observe each individual for a complete year, with the sole exception of their death occurring before the year-end. Otherwise, the complete year may not be interrupted by any right-censoring events, such as lapsing or the end of the investigation.

Pensions and annuities in payment are nominally a single-decrement risk, i.e. the only way they are supposed to cease is through the death of the recipient. On the face of it, this matches the q-type model in theory. However, in the following examples we will see that q-type models often do not match the reality of business practices, and that μ-type modelling is a more natural choice.

3.1 Bulk transfers out of a portfolio

Portfolios can experience mid-year transfers out *en masse*. Examples include a transfer of annuities to another insurer, the purchase of buy-out annuities in a pension scheme or transfer to another administration system. Figure 3 shows an example for a UK annuity portfolio, where a large block of annuity contracts was moved to another insurer in a Part VII transfer [FSM, 2000, PartVII].

Mid-year exits like in Figure 3(a) are a problem for q-style models. They violate the assumption that there is no other possible event besides death or survival to the end of the year. In contrast, mid-year exits pose no problem for modelling in continuous time — the transfer out date is the date of last observation, and the indicator variable is set $d = 0$ to signify that the life was alive at that date. See Case C in Figure 1, and Case Studies A.1 and A.4 in Appendix A.

3.2 New business

In addition to mid-year exits, most pension schemes and annuity portfolios also add new liabilities. These can be due to new retirements or spouse's annuities set up on death of the first life. It is also possible to have a mass influx of new pensions, such as a pension scheme seeing a wave of early retirements due to corporate restructuring. Figure 4 shows an annuity portfolio where new cases are added at the start of each month and where there was a surge in new business in December 2014.

If an analyst is modelling q-type rates aligned on calendar years, then any new records added

Figure 3: (a) Number of in-force annuities and (b) average age of in-force annuities at each date for a UK insurer. The discontinuities in late 2013 are caused by a transfer of liabilities to another insurer. New annuities are set up on any day, hence the seemingly continuous nature of the in-force annuity count at other times. Source: own calculations using experience data for all ages, January 2012–December 2014.

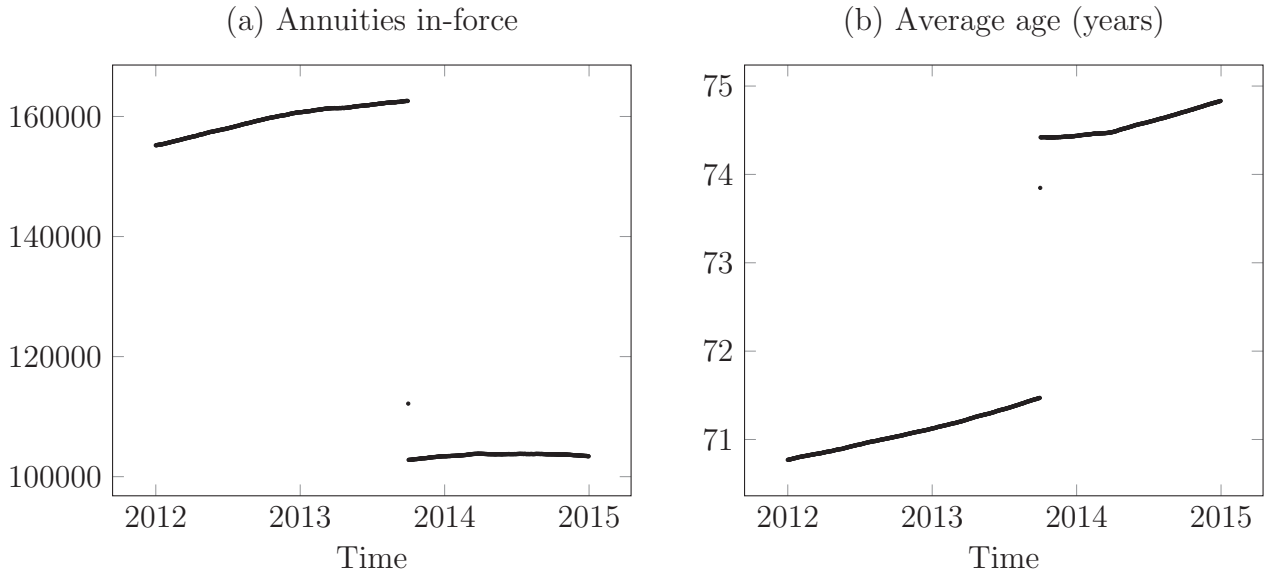
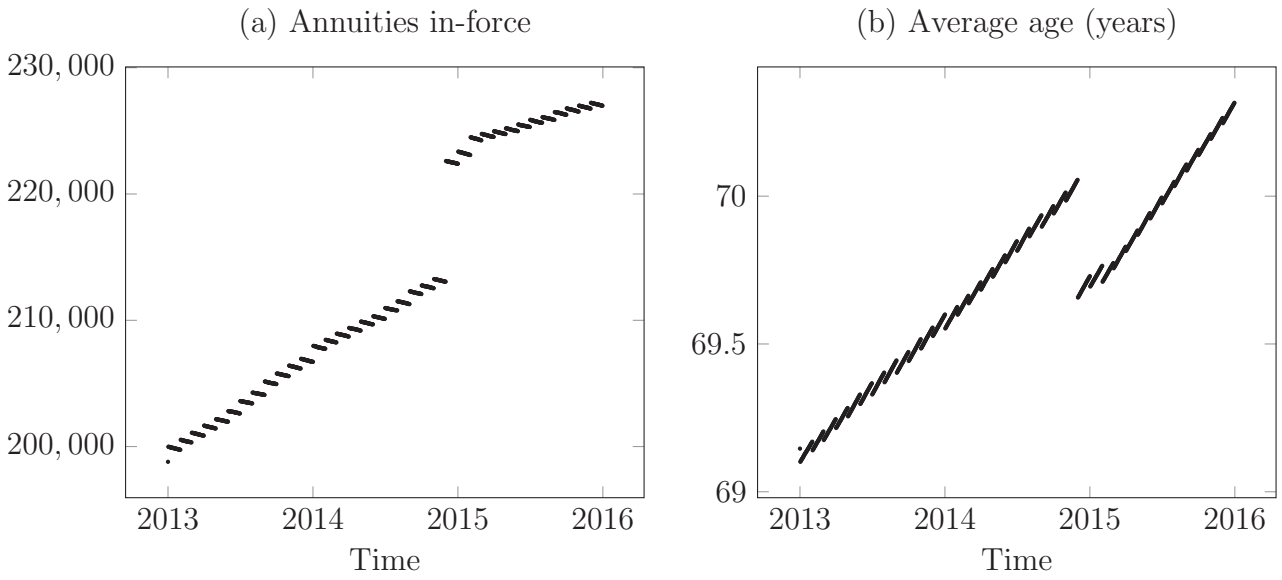


Figure 4: (a) Number of in-force annuities and (b) average age of in-force annuities at each date for a French insurer. The discontinuity in December 2014 was caused by a surge in new business. New annuities are set up on the first of the month, hence the stepped pattern of the annuity in-force count. Source: experience data for all ages, January 2013–December 2015.



during the year must be excluded from the model data set because a complete year of exposure is not possible (see Appendix B for the detailed reasoning underpinning this statement). In contrast, all new-business records can be used in a continuous-time model; see Case Study A.8 in Appendix A.

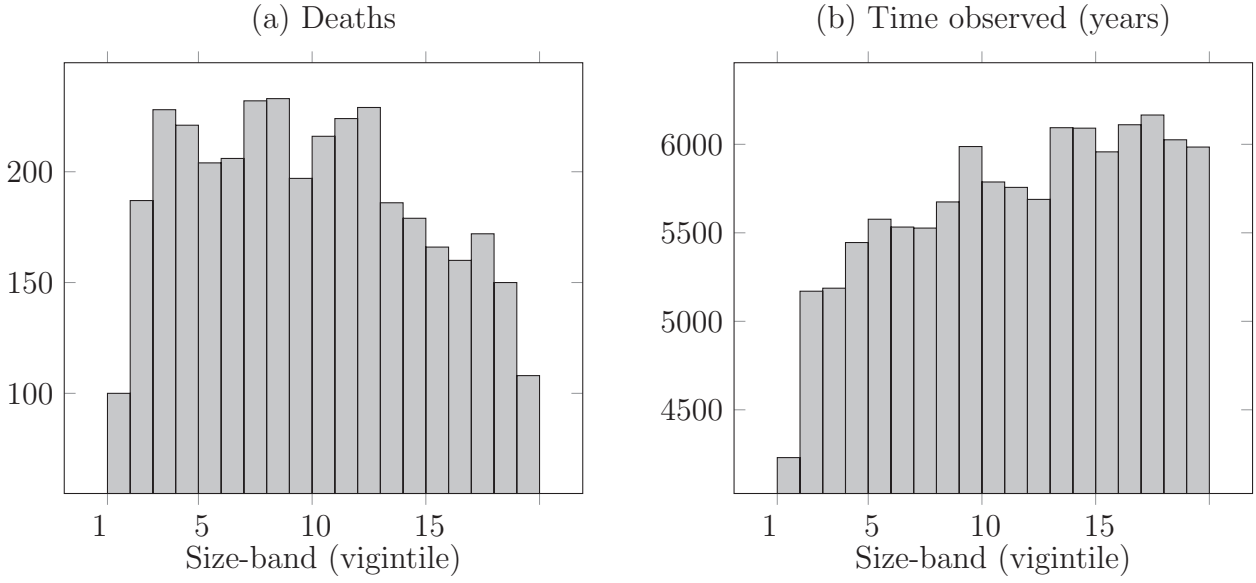
3.3 Bulk transfers into a portfolio

For every bulk transfer out of one portfolio in Section 3.1 there will be a corresponding bulk transfer in to another portfolio. Examples include an insurer receiving a Part VII transfer, an insurer writing a large bulk annuity, or an administration system receiving migrated records from another administration system. This is like the new business in Section 3.2, but with the added complication that the receiving system will be missing deaths between annuity commencement and the transfer-in date. This is of course a challenge for q-type models in the same way as for new business, but it is not a problem for continuous-time methods — see Cases D and E in Figure 1 and Case Study A.13 in Appendix A.

3.4 Commutations and temporary pensions

Even if a portfolio experiences no mass transfer out, it is still possible for individual liabilities to cease for a reason other than death. In the UK, pension benefits below a certain size can be commuted into a taxable lump sum. Administration expenses are largely fixed in nature, so the facility to commute small pensions saves disproportionate costs. However, this again means that the pension ceases for a reason other than the death of the recipient. Figure 5 shows the impact of this — although the number of lives in each size-band is the same, the time lived is much reduced for the smallest size-band due to non-death early exits. The number of observed deaths is correspondingly much reduced, as post-commutation deaths are of course unknown to the insurer.

Figure 5: (a) Deaths and (b) time observed in each pension size-band, showing the reduced exposure for the smallest pensions. Source: own calculations using experience data for Scottish pension scheme, ages 50–105, years 2000–2009. Pensioner records were sorted by pension size and grouped into twenty bands of equal numbers of lives (vigintiles). Size-band 1 represents the 5% of pensioners receiving the lowest pensions, while size-band 20 represents the 5% of pensioners receiving the largest amounts.



There are other ways for pensions to cease for a reason other than death. In Richards et al. [2013] a multi-employer pension administrator in Germany had (i) temporary widow(er) pensions,

(ii) widow(er) pensions ceasing on remarriage and (iii) disability pensions ceasing due to a return to work. In the Chilean national pension system, spouses' pensions also cease upon remarriage. Some UK pension schemes grant pensions to 'surviving adults' that can be terminated on remarriage or cohabitation with another person[TPS, 2010, 94(6)].

Commutations and temporary pensions pose a problem for q-style models of mortality. One option is to exclude such cases, but this depends on them being clearly identified as such; see Case Study A.9 in Appendix A. Another issue is the need to exclude any corresponding deaths — trivial commutations are not always processed immediately, which means it is possible for a commutation-eligible pensioner to die first. Appendix C shows how discarding lapsed cases leads to a bias in mortality estimation. Discarding such data is therefore to be avoided.

In contrast, commutations and temporary pensions do not pose any kind of problem for models in continuous time — the commutation date is the date of last observation, and the indicator variable is set $d = 0$ to signify that the life was alive at that date. See Case C in Figure 1 and Case Studies A.1 and A.4 in Appendix A.

3.5 Representation of data

In all the foregoing examples, non-death exits and mid-year entries are handled more easily by modelling in continuous time. This allows for simpler data preparation and corresponds more closely to the realities of the business processes than q-type models. To illustrate, we assume that we have a data extract from a pension or annuity administration system. For each record we calculate the data pair $\{d_i, t_i\}$ as follows. t_i is the real-valued time observed in years between the on-risk date and the off-risk date. The on-risk date is the latest of: (i) the annuity commencement date, and (ii) the transfer-in date (if any). If the transfer-in date is not known exactly, a suitable proxy date can be used, such as the date of first payment made on the system; see Case Study A.13 in Appendix A. The off-risk date is the earliest of: (i) the date of extract, (ii) the date of death (if dead), (iii) the date of commutation (if commuted), and (iv) the date of transfer out (if any). If the transfer-out date is not known exactly, a suitable proxy can be used, such as the date of last payment made on the system; see Case Study A.4 in Appendix A. The indicator variable is set $d_i = 1$ if the off-risk date is the date of death and $d_i = 0$ otherwise. This is depicted graphically in Figure 1.

Macdonald et al. [2018, Section 2.9.1] give an expanded consideration to include analyst-specified limits, such as minimum and maximum ages, modelling start and end dates and rules for excluding suspicious records. All of these are more easily handled in μ-type models than with q-type models.

4 Reflecting reality: risk features

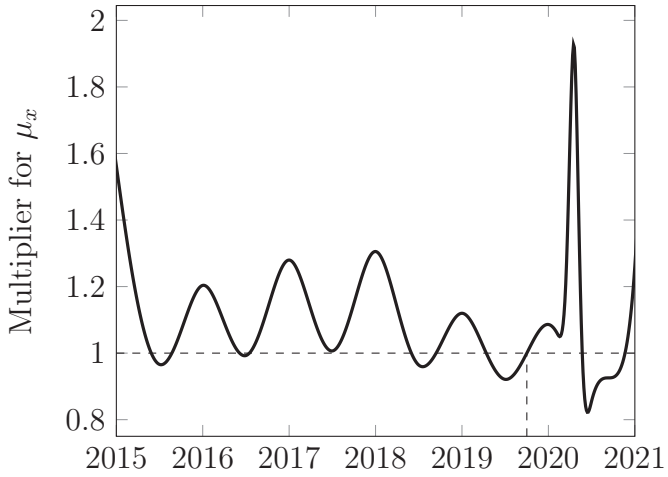
q-type methods are by definition unable to model mortality phenomena that take place over a shorter time frame than the fixed period they span. In contrast, continuous-time methods can better reflect rapidly changing phenomena, as demonstrated in the following sub-sections.

4.1 Mortality shocks

The covid-19 pandemic [The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020] produced intense spikes in mortality in a number of countries. This is reflected in the portfolio data that actuaries analyse; see the examples for annuity portfolios in France, the UK and the USA in Richards [2022b]. The first covid-19 shock in the UK was in April & May 2020, and the intensity of this two-month mortality shock cannot be captured with annual or even quarterly q-type rates. In contrast, continuous-time models can reveal rich detail. Figure 6 shows an example, where the first covid-19 shock is detected in April & May 2020 with seasonal variation in the years

beforehand. Underlying Figure 6 is a B-spline basis with knots at half-year intervals to pick up the substantial seasonal variation. In addition, *a priori* knowledge of population mortality shocks led to the addition of further B-spline knots in the first half of 2020, thus giving a basis of varying flexibility over time (there is no requirement for a B-spline basis to always have equally spaced knots; see Kaishev et al. [2016]). The result is a model for μ_x that allows mortality to vary smoothly over time, with periods of faster and slower change. Such detail is impossible with an annual q-type model.

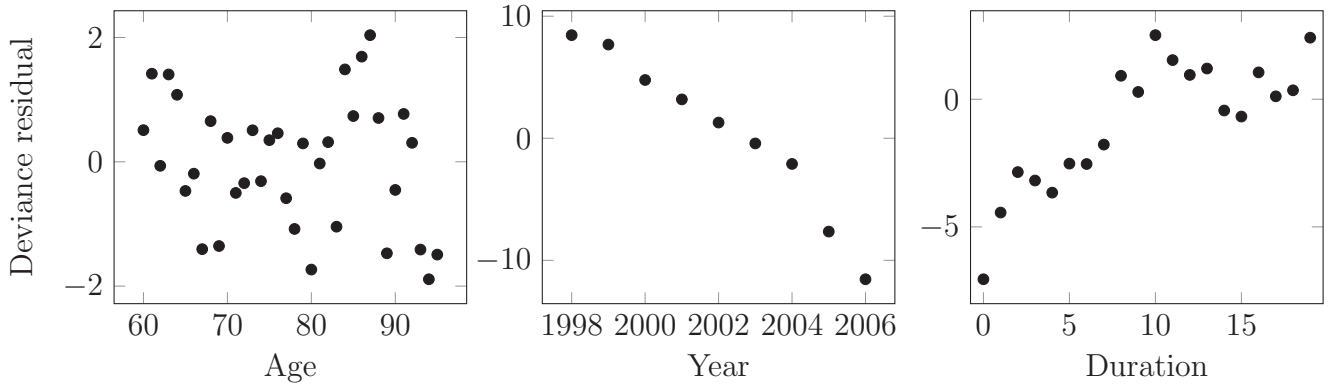
Figure 6: Mortality level over time for a UK annuity portfolio, where modelling in continuous time allows rich detail to be identified. The mortality level is standardised at 1 in October 2019. Source: Richards [2022c].



4.2 Period effects and selection

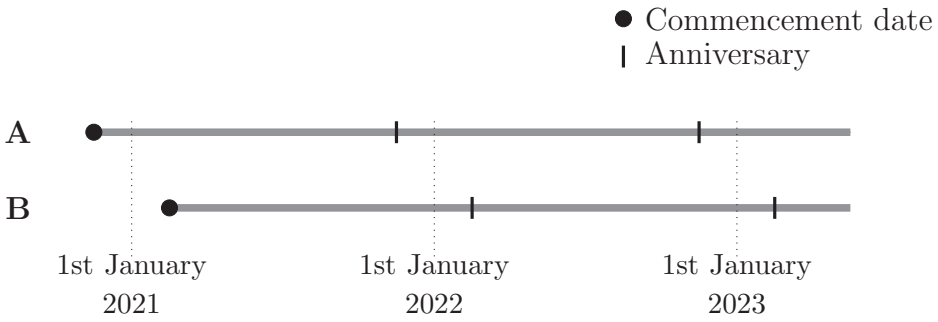
Figure 7 shows the deviance residuals [McCullagh and Nelder, 1989, Section 2.4.3] for a second portfolio of UK annuities. The model includes age and sex, and the random pattern of residuals by age suggests that the model is an acceptable fit for age-related mortality. However, the non-random patterns of residuals in time and duration suggest that it is also important to allow for both a period effects and selection effects.

Figure 7: Deviance residuals by age, year and duration since commencement. The systematic pattern of residuals in the centre and right panels show that both time and duration are important risk factors for mortality, and should both be included in a model. Source: own calculations for model for age and sex for second UK annuity portfolio, ages 60–95, years 1998–2006.



A drawback of a one-year q-type model is that it is limited in how it can handle time-varying risk, especially when there are two or more time dimensions. This is due to the requirement to align the one-year intervals, as shown in Figure 8 for period effects and duration effects. This limitation of q-type models improves a little if one of those risk factors is handled continuously (usually age). To see this, assume for simplicity that the basic mortality model is parametric, like $q_x = e^{\alpha+\beta x}/(1+e^{\alpha+\beta x})$ or $\mu_x = e^{\alpha+\beta x}$. In each case we have a function defined for any real age, x, not just integer ages. We can also align the period effects on the calendar years, say by assuming that α is a function of calendar time, for example piecewise constant on calendar years. Since we have a function for mortality at any real age, x, the fact that lives are at different ages during each year is not an issue for either the q_x or the μ_x model.

Figure 8: When policies can start at any point during the year, q_x models can either include annual period effects or selection effects, but not both. For example, a period effect for 2021 requires a full year of exposure for survivors, so only Case A can contribute. However, the splitting of the exposure at annual boundaries means that different curtate durations are exposed, thus excluding the modelling of selection effects as well. The converse also applies.



However, how can the q_x model accommodate selection effects for annuities commencing in the middle of the year? A one-year interval can either be aligned for annual period effects, or it can be aligned for selection effects, but it cannot be aligned for both. This is shown in Figure 8. In contrast, the μ_x model can simultaneously model period and duration effects without such restriction. Of course, one could reduce the period covered by the q_x model to, say quarter-years. But if the answer is to reduce the fixed period covered by a $q-$type model, why not go all the way and reduce the interval to daily modelling?

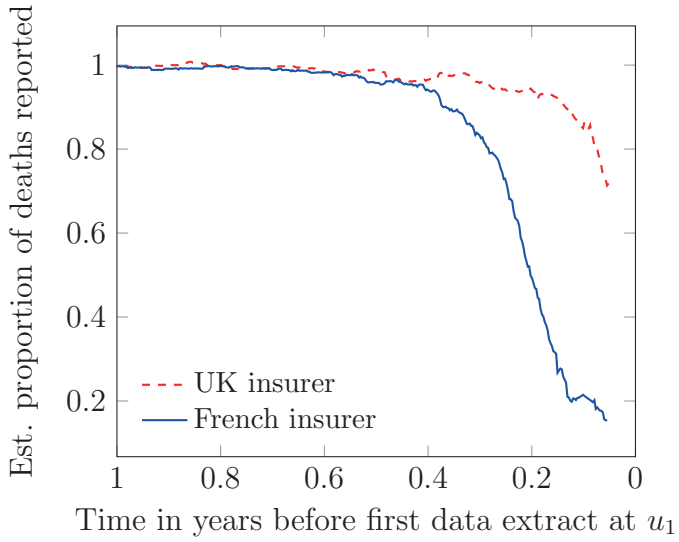
The problem of multivariate time scales is covered in detail in Andersen et al. [1993, Chapter X]. See also Case Study A.8 in Appendix A.

4.3 Occurred-but-not-reported (OBNR) deaths

One other useful aspect of continuous-time analysis is the ability to examine delays in the reporting of deaths. Lawless [1994] describes this as the problem of occurred-but-not-reported (OBNR) events, and Kalbfleisch and Lawless [1991] consider parametric models for estimating the distribution of delays. Section D.5 in Appendix D describes the calculation of a semi-parametric estimator of the portfolio-level hazard in continuous time. Assume that we have two estimates of the mortality hazard at the same point in time $y+t$. These two estimates are labelled $\hat{\mu}_y(t,u_1)$ and $\hat{\mu}_y(t,u_2)$, and they differ only in that they have been calculated using data extracts taken at two different times $u_1 < u_2$. Then $\hat{\mu}_y(t,u_1)$ will be more affected by delays in the reporting of deaths than $\hat{\mu}_y(t,u_2)$, with the effect decreasing the further $y+t$ lies before u_1. Figure 9 shows the ratio of $\hat{\mu}_y(t,u_1)/\hat{\mu}_y(t,u_2)$ as an

estimate of the proportion of deaths that have not been reported. The two portfolios have minimal OBNR cases half a year or more before time u_1, but have very different OBNR rates in the handful of months leading up to the extract. This level of insight is not possible with annual q-type analysis.

Figure 9: Estimated proportion of deaths reported for annuity portfolios in France and UK. The horizontal axis is reversed, as it measures the time before the first extract at calendar time u_1. Source: Richards [2022b, Section 4].



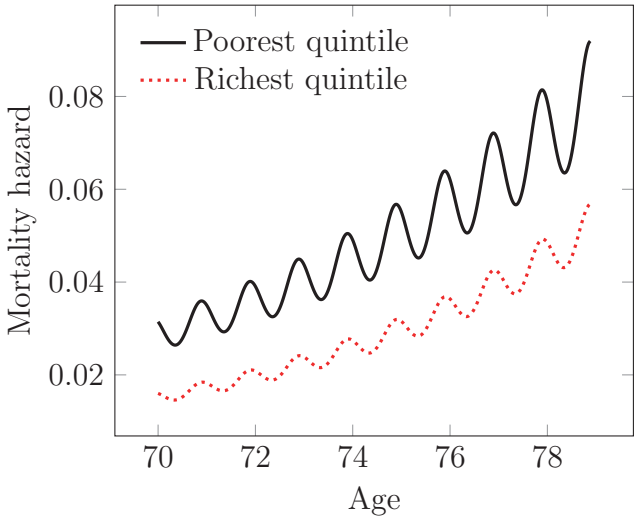
Continuous-time methods go further than the identification of reporting delays in Figure 9. Including a period affected by OBNR deaths will naturally tend to under-state mortality levels, which is one reason to discard the exposure period immediately prior to the extract date; Figure 9 suggests that six months should be discarded for the French annuity portfolio and three months for the UK one. However, continuous-time models make it possible to model these OBNR deaths and thus use the full experience data, even up to the date of extract itself. This OBNR allowance can be either parametric [Richards, 2022b, Section 5] or semi-parametric [Richards, 2022c, Section 10]. Richards [2022b, Section 7] showed how the use of daily data and continuous-time methods allowed the "nowcasting" of current mortality affected by late-reported deaths. These sorts of techniques are not possible with annual q-type approaches. See also Case Study A.15 in Appendix A.

4.4 Season

Annual q-type rates are by definition unable to model the seasonal variation shown in Figure 6. The mortality hazard rises overall as the lives age, but with fluctuations as the lives pass through the seasons of each year. The amplitude of these fluctuations grows larger with increasing age.

As a periodic factor, season has the unusual distinction of being both highly statistically significant in explaining mortality variation, and yet of low financial significance for long-term business [Richards, 2020, Section 9]. Nevertheless, accounting for seasonality can be useful in separating out the impact of pandemic shocks, and can help quantify the extent to which a heavy winter flu experience is unusual. Seasonal modelling has applications for public-policy work, such as the observation that poorer annuitants not only have higher mortality, but also experience stronger seasonal variation, as shown in Figure 10. Models of seasonal variation can also be useful in commercial situations, such as Case Studies A.5, A.7 and A.12 in Appendix A.

Figure 10: Modelled cohort mortality hazard for males initially aged 70 at 1st January by selected income quintile. Source: model calibrated to mortality experience of large UK pension scheme in Richards et al. [2020].



5 Investigation period

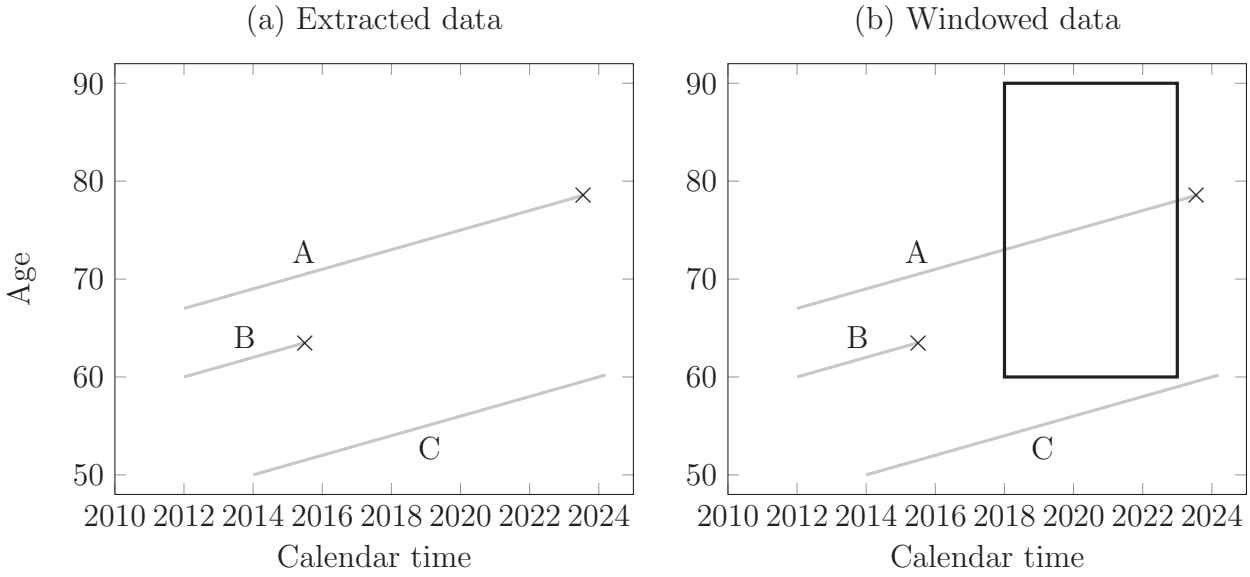
The investigation period will be set by the analyst, and there may be specific organizational policies for this. In setting the investigation period, the analyst needs to balance the benefits of statistical power (from using a longer period) against relevance (such as when mortality levels have been falling or the portfolio mix is rapidly changing); see Benjamin and Pollard [1986, page 49]. For example, ONS [2024] use population experience data over a three-year period to create the UK National Life Tables, while some life offices have a policy of using five-year period to set bases. Other investigations use longer periods of exposure still.

Specifying an investigation window causes additional left-truncation and right-censoring, as shown in Figure 11; in addition to the life only becoming known to the insurer at an adult age at policy outset (standard left-truncation) the analyst further discards the experience from policy outset to the start of the investigation window. Case Study A.18 in Appendix A concerns an annuity portfolio with uninterrupted experience data on the same administration system since outset. Despite the availability of decades of experience, the analyst would most likely only use the more recent data due to reasons of relevance. However, the length of the investigation period may also be limited by the data itself — Case Studies A.5 and A.7 in Appendix A are examples where the analyst has no scope for using a five-year investigation period.

Most investigation periods also span a whole number of years, in part to balance out the seasonal fluctuations in mortality; see Figure 10 and years 2016-2019 in Figure 6. However, if a model includes an explicit seasonal term [Richards et al., 2020], then there is no need to worry about any bias caused by unbalanced numbers of summers and winters. This permits μ-type models to have an investigation period of a non-integer number of years if a seasonal component is included; see Case Studies A.5 and A.7 in Appendix A.

The ability to use fractional numbers of years in an investigation extends particularly to transactions such as bulk annuities. Pension schemes in the UK sometimes change administrators, which invariably leads to the loss of mortality data prior to the switchover. If a pension scheme can only supply 2.5 years of mortality experience, then the insurer will not want throw away a fifth of the information just because a GLM for q_x can only use whole periods (although a q-type model is

Figure 11: An analyst wants to fit a Gompertz model of mortality to the experience between ages 60–90 over the period 1st January 2018-1st January 2023. An extract is taken of the data on 14th March 2024 and some specimen cases are shown in this Lexis diagram with exposure times in gray and deaths are marked with ×. (a) Raw experience data and (b) experience data within the analyst's window superimposed. Individual A died on 20th July 2023, but is right-censored at the end of the investigation period and so is regarded as alive in the analysis. Individual A's exposure time to the left of the window is also left-truncated. Individual B died before the investigation window opened and is excluded from the analysis. Individual C is also excluded from the analysis because they never reach the age range within the investigation period.



still possible if one does not use a GLM; see Appendix B). Any concerns about delays in reporting of deaths can be addressed in a μ-type model by including an explicit occurred-but-not-reported (OBNR) term; see Richards [2022b, Section 5].

At the time of writing, most portfolios will still have recent mortality experience affected by covid-19; see the spike in mortality in April and May 2020 in Figure 6. μ-type models can allow for such spikes in the experience [Richards, 2022c] in a way that q-type models cannot. However, another alternative to avoiding bias might be to exclude the affected period, although this presupposes that the covid-affected experience is an additive outlier with no relevance to later mortality. For a q_x GLM this involves discarding an entire year's worth of data, despite the fact that the shock in Figure 6 only lasts around two months (although, as before, a q-type model is still possible if one does not use a GLM). In contrast, μ-type models can more easily handle the *ad hoc* exclusion of short periods of experience, if the analyst feels that this is truly necessary. Macdonald and Richards [2024, Section 4.3] show how both left truncation and right censoring can be handled by setting an exposure indicator function, $Y(x)$, to zero to 'switch off' the contribution of exposure time and observed deaths to estimation. $Y(x)$ is a multiplier of μ_x, and to exclude the contribution of covid-affected experience, one could also set $Y(x) = 0$ for the period of any mortality shocks.

6 Competing decrements

A limitation of q-type models is that they assume that only two events are possible: survival to the end of the observation interval, or death during it. Even with notionally single-decrement business like pensions or annuities in payment, Section 3 shows that the business reality is that there are often other decrements like administrative transfers or trivial commutations.

However, the situation for q-type models becomes worse with classes of business that have multiple modes of exit by design. Consider the data in Case Study A.32, where there are two decrements of interest: death and entry into long-term care. $n = 42$ lives start exposure at age 88, with $d^{\text{mort}} = 5$ dying and $d^{\text{care}} = 3$ entering long-term care before reaching age 89. Due to these exits and other censoring points there are $E^c = 36.9039$ years of life observed in the age interval $[88, 89)$. With continuous-time modelling, the number of deaths and care events can be modelled as separate pseudo-Poisson variables [Macdonald and Richards, 2024], which lead to the same inference as assuming that:

$$d^{\text{mort}} \sim \text{Poisson}(\mu^{\text{mort}} E^c), \text{and} \tag{1}$$
$$d^{\text{care}} \sim \text{Poisson}(\mu^{\text{care}} E^c) \tag{2}$$

where μ^{mort} and μ^{care} are the hazards for mortality and long-term care, respectively. In contrast, q-type modelling has little theoretical support beyond binomial trials, which don't permit multiple types of event. The classical actuarial approach is to try and create an artificial single-decrement equivalence by reducing the number of lives:

$$d^{\text{mort}} \sim \text{Binomial}(q^{\text{mort}}, n^{\text{mort}}), \text{and} \tag{3}$$
$$d^{\text{care}} \sim \text{Binomial}(q^{\text{care}}, n^{\text{care}}) \tag{4}$$

where q^{mort} and q^{care} are the probabilities of mortality and long-term care, respectively. n^{mort} and n^{care} are pseudo-life-counts, both less than n. It is traditional to use an approximation like the following from Benjamin and Pollard [1986, equation (2.1)]:

$$n^{\text{mort}} \approx n - \frac{1}{2} d^{\text{care}}, \text{and} \tag{5}$$
$$n^{\text{care}} \approx n - \frac{1}{2} d^{\text{mort}} \tag{6}$$

The lack of rigour with the q-type approach to competing decrements can be seen when we apply it to our example:

$$n^{\text{mort}} \approx 40.5, \text{and} \tag{7}$$
$$n^{\text{care}} \approx 39.5 \tag{8}$$

But what is 0.5 of a binomial trial? Half a person? The binomial likelihood can still be evaluated and maximised for non-integral values for the number of trials, but the q-type models in equations (3) and (4) clearly do not correspond to reality.

See also Case Studies A.30, A.31 and A.32 in Appendix A.

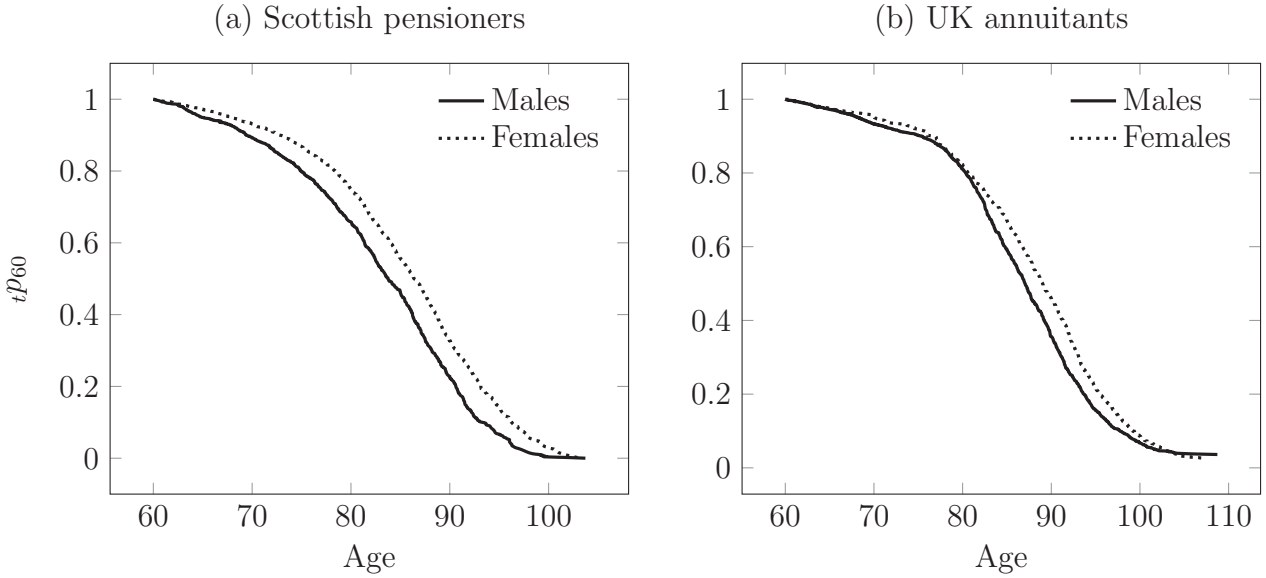
7 Data quality checks

Continuous-time methods also offer better checks of data quality. We look at how data issues can be revealed by a non-parametric estimate of survival by age, and a semi-parametric estimate of mortality hazard in time.

7.1 Non-parametric survival curve by age

The estimator from Kaplan and Meier [1958] estimates the survival curve at intervals determined by the time between deaths (the definition from Fleming and Harrington [1991] does something very similar). In the case of many actuarial portfolios, this interval is often as small as one day, making the Kaplan-Meier estimate effectively a continuous-time measure; see Section D.1 in Appendix D. An example is shown in Figure 12(a).

Figure 12: Kaplan-Meier survival curves from age 60 for (a) the Scottish pension scheme in Figure 5 and (b) a UK annuity portfolio. The odd shape of the survival curve for the UK annuitants, and the lack of a male-female differential between ages 60–80, suggests a data-corruption problem like the one described in Case Study A.6 in Appendix A. Source: Macdonald et al. [2018, Figure 2.8].

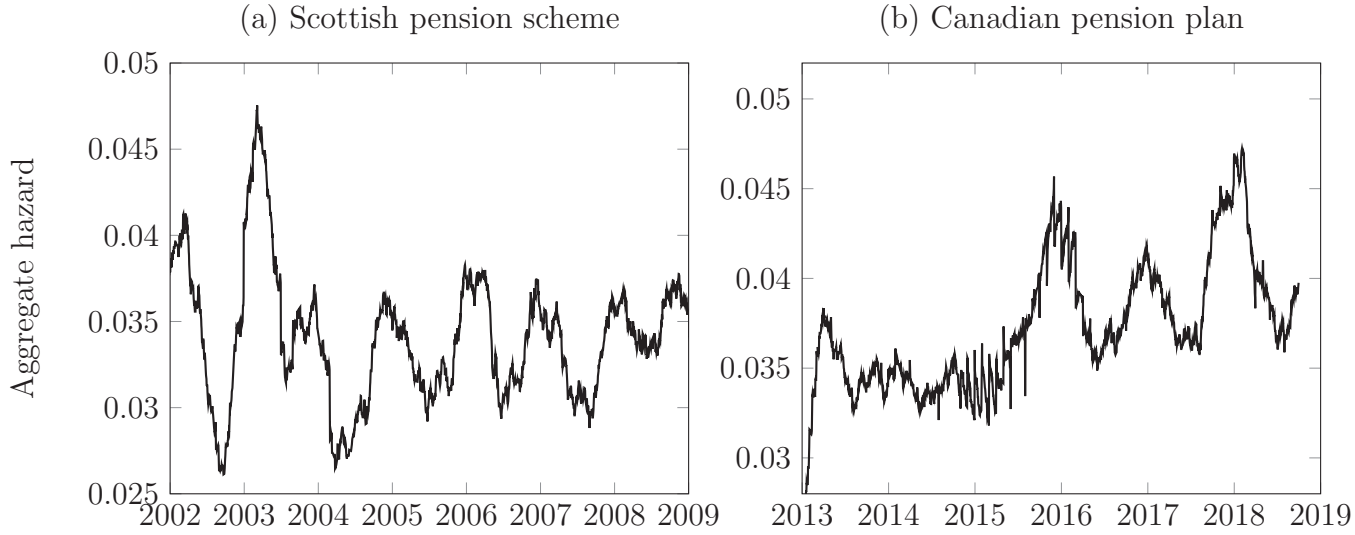


The value of the Kaplan-Meier estimate as a data-quality check is shown in Figure 12(b), where the odd shape and the lack of a male-female differential between ages 60 and 80 suggests a data-quality problem. Case Study A.6 in Appendix A describes a possible cause. Note that this sort of problem is trickier to detect with a traditional comparison of mortality against a standard table, and the visual nature of the Kaplan-Meier estimate is a major advantage when communicating with non-specialists. See also Case Studies A.2 and A.3 in Appendix A.

7.2 Semi-parametric hazard over time

Richards [2022b] proposed a semi-parametric estimate for the mortality hazard of a portfolio over time, and the details of its calculation are given in Section D.5 in Appendix D. Figure 13(a) shows the seasonal variation in a Scottish pension scheme, with particularly high mortality in the winter of 2002-2003. In contrast, the much larger Canadian pension plan in Figure 13(b) shows seasonal variation only from 2016 onwards, raising questions about the data quality in 2014 and 2015. Closer inspection of the Canadian mortality reveals that deaths are recorded continuously throughout the year, but that the first of each month has disproportionately more deaths than other times during the same month. This suggests an inconsistent approach to the recording of the date of death; see also Case Study A.15 in Appendix A. Detailed inspection showed that this data anomaly is present throughout the entire period shown in Figure 13(b), but it is more dominant in 2014 and 2015 and thus more visually apparent.

Figure 13: Estimated aggregate hazard in time for (a) the Scottish pension scheme in Figure 5 and (b) a Canadian pension plan with more lives and deaths. The Canadian pension plan is missing the expected seasonal pattern in 2014 and 2015, raising questions over the accuracy of the experience data during that period. Source: adapted from Richards [2022b, Figure A.2].



The mortality spike in the winter of 2002–2003 in Figure 13(a) raises a question: why would one exclude covid mortality spikes, when such seasonal spikes have typically been left in?

8 Management information

A major advantage of continuous-time methods is their ability to give timely and ongoing answers to management questions. For example, consider the surge in new business shown in December 2014 in Figure 4(a). The average age at commencement of the 9,570 new annuitants in December 2014 is 60.75 years. This is close to the average age at commencement of 61.34 years for the 9,033 new annuitants in the six months before and after December 2014. Although the commencement ages are very close, a natural management question is whether the mortality of the December 2014 annuitants is different from other recent new business. Figure 14 shows that this is indeed the case, with the December 2014 annuitants experiencing lighter mortality than their contemporaries. The statistic used here is the Nelson-Aalen estimate of equation (18) (see Section D.2), which can be updated continuously without waiting for a complete year of exposure as with a a q-type analysis; Figure 14 could be updated weekly, for example. This is a critical advantage when entering a new market; see Case Study A.8 in Appendix A.

Of course, the phenomenon of OBNR deaths discussed in Section 4.3 is a challenge for any methods using current data. However, as shown in Figure 9, continuous-time semi-parametric methods can also be used to estimate the timing and incidence of OBNR. If need be, graphs such as Figure 14 can be adjusted to allow for OBNR. This is analogous to the econometric techniques called "nowcasting" [Bańbura et al., 2013], where the most recently available information is used to estimate the current value of a statistic that won't be fully known until sufficient time has elapsed. This is illustrated in Figure 15(a), where the June 2020 data extract shows no sign of the pandemic mortality expected in April & May 2020. An allowance for OBNR allowed a "nowcast" of mortality, as shown in Figure 15. A later data extract used in Figure 15(c) shows that the OBNR-adjusted "nowcast" in panel (b) was accurate enough for business planning at the time it was made.

Figure 14: Nelson-Aalen estimate, $\hat{\Lambda}(t)$, of cumulative mortality rate for new business written (i) in December 2014, and (ii) in the six months on each side of December 2014 . Source: own calculations using new annuities written by the French insurer in Figure 4.

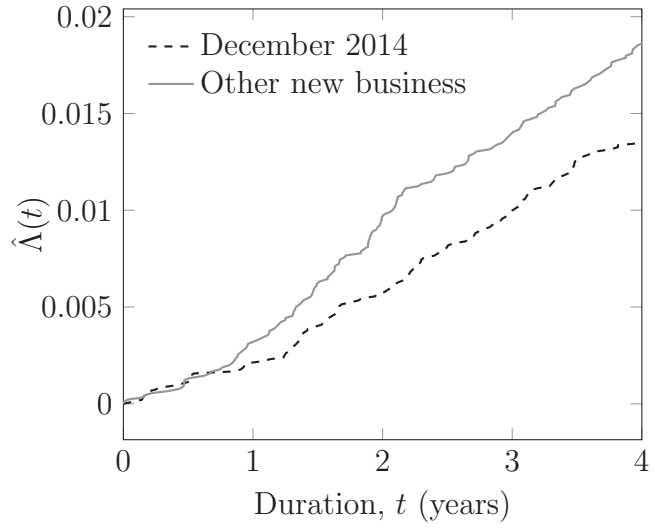
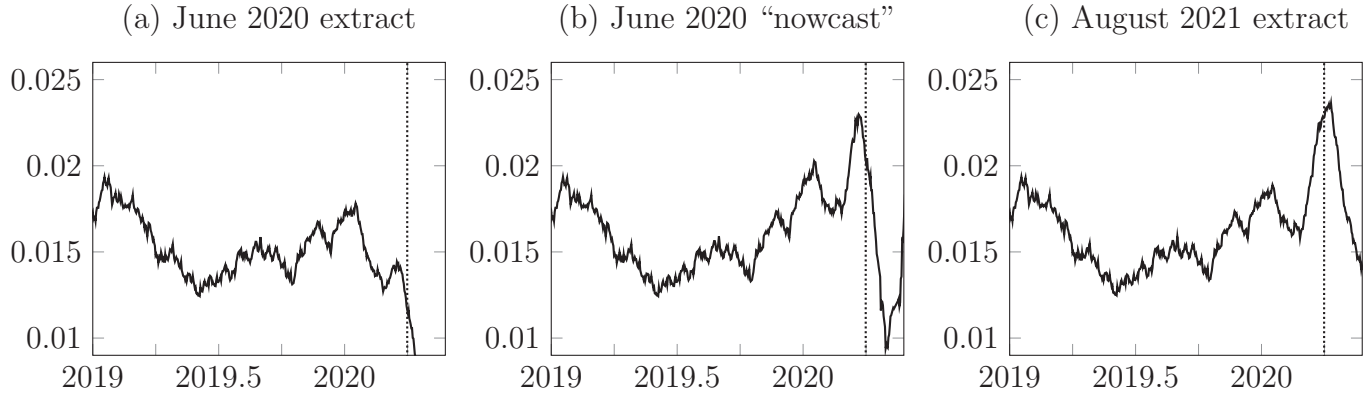


Figure 15: $\hat{\mu}_{2019}(t)$ for French annuity portfolio with $c = 0.1$: (a) calculated using June 2020 extract; (b) nowcast using Gaussian OBNR function estimated from June 2020 extract; (c) calculated using August 2021 extract. The vertical dotted line in each panel is at 1st April 2020.



9 Discussion

The problems of single-decrement q-type models in Section 3 can be partially solved by discarding some records and editing others. However, throwing away data is a luxury that not every analyst can afford; see Case Studies A.5 and A.7 in Appendix A. Furthermore, modifying the data to suit the needs of the model is surely the wrong direction of causality — a model should be picked that suits the data, not the other way around. As shown in Section 3, continuous-time methods are a better match to the reality of business processes on most administration systems.

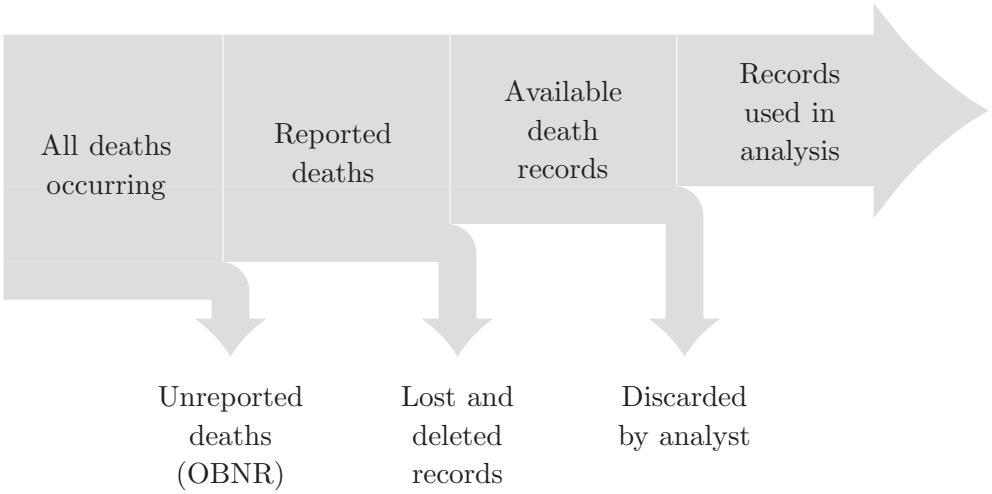
The challenges of q-type models in Section 4 can be addressed by reducing the fixed observation period to a quarter-year, or even a month. However, if the answer is to increase granularity, why not use the full detail afforded by the data, namely daily intervals? Administration systems hold lots of dates, which are a natural fit for continuous-time methods.

Alternatively, q-type models can be fitted using bespoke likelihoods for fractional years of exposure, as covered in Appendix B. However, if the analyst is able to do this, then a μ-type model will be more flexible and require less data preparation.

Individual records and exact dates allow the use of insightful non- and semi-parametric methods, as shown in Section 7. The Kaplan-Meier curve is a particularly useful tool for detecting subtle data-quality problems. And if one is anyway extracting individual records for data-checking, it makes sense to continue by modelling individual lifetimes if one can. The continuous-time representation attempts to describe these left-truncated, right-censored lifetimes as they really are (and presumably good business data processes share that aim to some extent). In contrast, the discrete-time representation for q-type models works the other way, and attempts to force nature (and the data) to look like a life table.

Figure 16 depicts the main categories of data loss between reality on the left and what the analyst has to work with on the right. There is not much that can be done about the middle sources of data loss: if records have been deleted or destroyed, they are unavailable to any type of model. However, μ-type models can deal with OBNR deaths better than q-type models, and less data is discarded with continuous-time methods than discrete-time ones.

Figure 16: Cumulative loss of mortality data in a portfolio.



10 Conclusions

Methods operating in continuous time, or in daily intervals, offer substantial advantages for actuaries. Chief among these are the ability to reflect the reality of everyday business processes, and the ability to handle rapid changes in risk level. In contrast, q-type models, especially those operating over a one-year interval, make assumptions about the data that are often not true in practice, and often require data to be discarded. In addition, annual q-type models are unable to reflect rapid intra-year movements in risk.

Continuous-time methods are also the more-practical solution to certain modelling problems. Examples include risks varying in more than one time dimension and modelling risks with competing decrements. In contrast, annual q-type models within a GLM framework are not capable of simultaneously modelling time- and duration-varying risk. For competing risks, q-type models force the pretence that the number of lives is smaller than it really is.

Continuous-time methods also offer useful data-quality checks. The visual nature of non- and semi-parametric methods make them particularly useful for communication with non-specialists.

There are occasional circumstances when daily information is unavailable, or even deliberately withheld, but actuaries should strive wherever possible to collect full dates and use continuous-time methods of analysis.

11 Acknowledgements

The authors thank Gavin Ritchie, Stefan Ramonat, Patrick Kelliher, Kai Kaufhold and Ross Ainslie for comments on earlier drafts. Any errors in the paper are the sole responsibility of the authors. The authors also thank the owner of the home-reversion portfolio in Case Study A.32, who wishes to remain anonymous. This document was typeset in LaTeX. Graphs were created with tikz [Tantau, 2024] and pgfplots [Feuersänger, 2015].

References

Financial Services and Markets Act 2000. UK Public General Acts 2000 c.8, 2000.

The Teachers' Pension Regulations 2010. UK Statutory Instruments 2010 No. 990, 2010.

Personal Information Protection and Electronic Documents Act. S.C. 2000, c. 5, 2018.

2021 Pension Insurance Data Tables (Revised 02/12/2024), 2024. URL `https://www.pbgc.gov/prac/data-books`.

R. R. Ainslie. Annuity and insurance products for inpaired lives. Staple Inn Actuarial Society, 2000.

P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer, 1993. ISBN 0-387-94519-9.

E. Arjas and P. Haara. A marked point process approach to censored failure data with complicated covariates. *Scandinavian Journal of Statistics*, 11:193–209, 1984.

BACS Direct Credit — an introduction to the service. Bacs Payment Schemes Ltd, 2019. URL `https://www.bacs.co.uk/media/cmuccdew/bdc-introduction.pdf`.

M. Bańbura, D. Giannone, M. Modugno, and L. Reichlin. Chapter 4 — Now-Casting and the Real-Time Data Flow. In Graham Elliott and Allan Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, pages 195–237. Elsevier, 2013. doi: 10.1016/B978-0-444-53683-9.00004-9.

Y. Ben-Shlomo, G. D. Smith, M. Shipley, and M. G. Marmot. Magnitude and causes of mortality differences between married and unmarried men. *Journal of Epidemiology and Community Health*, 47(3):200–205, 1993. doi: 10.1136/jech.47.3.200.

B. Benjamin and J. H. Pollard. *The Analysis of Mortality and Other Actuarial Statistics*. Heinemann, London, 1986. ISBN 0 434 90137 7.

J. D. Broffitt. Maximum likelihood alternatives to actuarial estimators of mortality rates. *Transactions of the Society of Actuaries*, 36:77–142, 1984.

D. Collett. *Modelling Survival Data in Medical Research*. Chapman and Hall, Boca Raton, second edition, 2003. ISBN 1-58488-325-1.

V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969. doi: 10.1137/1114019.

C. Feuersänger. *Manual for Package PGFPLOTS, Version 1.12.1*, 2015. URL `http://sourceforge.net/projects/pgfplots`.

T. R. Fleming and D. P. Harrington. *Counting processes and survival analysis*. Wiley, New York, 1991.

D. O. Forfar, J. J. McCutcheon, and A. D. Wilkie. On graduation by mathematical formula. *Journal of the Institute of Actuaries*, 115:1–149, 1988.

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

J. M. Hoem. A flaw in actuarial exposed-to-risk theory. *Scandinavian Actuarial Journal*, 1984(3): 187–194, 1984. doi: 10.1080/03461238.1984.10413766.

V. K. Kaishev, D. S. Dimitrova, S. Haberman, and R. J. Verrall. Geometrically designed, variable knot regression splines. *Computational Statistics*, 31(3):1079–1105, 2016. doi: 10.1007/s00180-015-0621-7.

J. D. Kalbfleisch and J. F. Lawless. Regression models for right truncated data with applications to AIDS incubation times and reporting lags. *Statistica Sinica*, 1(1):19–32, 1991. ISSN 10170405, 19968507. URL `http://www.jstor.org/stable/24303991`.

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

J. F. Lawless. Adjustments for reporting delays and the prediction of occurred but not reported events. *Canadian Journal of Statistics*, 22(1):15–31, 1994. doi: 10.2307/3315826.n1.

A. S. Macdonald. An actuarial survey of statistical models for decrement and transition data, II: competing risks, non-parametric and regression models. *British Actuarial Journal*, 2(2):429–448, 1996.

A. S. Macdonald and S. J. Richards. On Contemporary Mortality Models for Actuarial Use II: Principles. *Working paper*, 2024.

A. S. Macdonald, S. J. Richards, and I. D. Currie. *Modelling Mortality with Actuarial Applications*. Cambridge University Press, Cambridge, 2018. ISBN 978-1-107-04541-5.

A. Madrigal, F. Matthews, D. Patel, A. Gaches, and S. Baxter. What longevity predictors should be allowed for when valuing pension scheme liabilities? *British Actuarial Journal*, 16(I):1–62 (with discussion), 2011. doi: 10.1017/S1357321711000018.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, second edition, 1989. ISBN 0-412-31760-5.

W. Nelson. Theory and applications of hazard plotting for censored failure times. *Technometrics*, 14:945–965, 1958.

ONS. National life tables QMI: Quality and methodology information for the national life tables. Technical report, Office for National Statistics, 2024. URL `https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/nationallifetablesqmi`. Accessed on 2024-03-11.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

E. G. Richards. Calendars. In S. E. Urban and P. K. Seidelman, editors, *Explanatory Supplement to the Astronomical Almanac*, pages 585–624. University Science Books, Mill Valley, California, 2012.

S. J. Richards. Applying survival models to pensioner mortality data. *British Actuarial Journal*, 14(II):257–326 (with discussion), 2008. doi: 10.1017/S1357321700001720.

S. J. Richards. Mis-estimation risk: measurement and impact. *British Actuarial Journal*, 21 (3): 429–457, 2016. doi: 10.1017/S1357321716000040.

S. J. Richards. A Hermite-spline model of post-retirement mortality. *Scandinavian Actuarial Journal*, 2020:2:110–127, 2020. doi: 10.1080/03461238.2019.1642239.

S. J. Richards. Modelling mortality by continuous benefit amount. *Scandinavian Actuarial Journal*, pages 695–717, 2022a. doi: 10.1080/03461238.2022.2025891.

S. J. Richards. Real-time measurement of portfolio mortality levels in the presence of shocks and reporting delays. *Annals of Actuarial Science*, 16(3):430–452, 2022b. doi: 10.1017/S1748499522000021.

S. J. Richards. Allowing for shocks in portfolio mortality models. *British Actuarial Journal*, 27:1–22 (with discussion), 2022c. doi: 10.1017/S1357321721000180.

S. J. Richards and D. Robinson. Anti-selection and annuity pricing. *The Actuary magazine*, May 2005.

S. J. Richards, K. Kaufhold, and S. Rosenbusch. Creating portfolio-specific mortality tables: a case study. *European Actuarial Journal*, 3 (2):295–319, 2013. doi: 10.1007/s13385-013-0076-6.

S. J. Richards, S. J. Ramonat, G. Vesper, and T. Kleinow. Modelling seasonal mortality with individual data. *Scandinavian Actuarial Journal*, pages 1–15, 2020. doi: 10.1080/03461238.2020.1777194.

SAS Institute Inc. *SAS/STAT 14.3 User's Guide*, 2017.

T. Tantau. *TikZ and PGF manual for Version 3.1.10*, 2024. URL `https://pgf-tikz.github.io/pgf/pgfmanual.pdf`.

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. *China CDC Weekly*, 2(8):113–122, 2020. ISSN 2096-7071. doi: 10.46234/ccdcw2020.032. URL `http://weekly.chinacdc.cn//article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51`.

A. Tims. Retired teacher's pension stopped as provider refuses to believe she is not dead. *The Guardian*, 20th January, 2024.

R. C. Willets. Mortality in the next Millennium. *Staple Inn Actuarial Society, London*, 1999.

Appendices

A Case Studies

The following are case studies are based on actual experience. In most case studies, continuous-time methods provided important benefits over q-type models.

Case Study A.1

The Pension Benefit Guaranty Corporation (PBGC) in the United States charges a flat annual fee of \$37 per participant in a pension plan in 2024, rising to \$52 in 2031 [PBG, 2024, Table M-16]. To minimise payments to the PBGC, a pension plan therefore buys annuities to close out the liabilities for the individuals with the smallest pensions. The pensions that are bought out are treated like commutations, with the observation end date being the date of annuity purchase.

A continuous-time model can include the bought-out annuities, as they will be like Case C in Figure 1. Note that the PBGC fee is per participant, so there is a financial benefit from identifying multiple records relating to the same person among retained cases; see the discussion of deduplication in Macdonald et al. [2018, Section 2.5].

Case Study A.2

A consulting actuary is asked by a reinsurer to set a pricing basis for a longevity swap. However, the Kaplan-Meier curves for males and females exhibits a pattern like that of Figure 12(b), suggesting that the data are corrupted. The consultant informs the reinsurer, which asks the cedant for better-quality data. The cedant refuses, and the consultant recommends that the reinsurer walks away from the transaction, which it does.

Case Study A.3

A consulting actuary is asked by the sponsoring employer of a pension scheme to validate the mortality basis used for funding calculations. The consulting actuary calculates the Kaplan-Meier curves for males and females and, like Case Study A.2, finds a pattern like that of Figure 12(b), suggesting that the data are corrupted. The third-party pension-scheme administrator is ordered to implement a data clean-up programme, which produces better data and thus allows a reliable basis to be derived.

Case Study A.4

The insurer of the annuities in Figure 3 needs to set a mortality basis for the year-end valuation in 2015. A simple multi-year comparison against a standard table is inappropriate, as Figure 3(b) shows that the transferred annuities are not a random sample of the overall portfolio. A q-type analysis is complicated by the exit from observation of 60,000 annuities during 2013. A better solution is a μ-type analysis, using the date of the last annuity payment on the system as a proxy for the date of right-censoring. This allows both transferred and remaining annuities to receive the same treatment as Cases A and B in Figure 1. Such a solution works best for annuities paid relatively frequently, e.g. monthly.

Case Study A.5

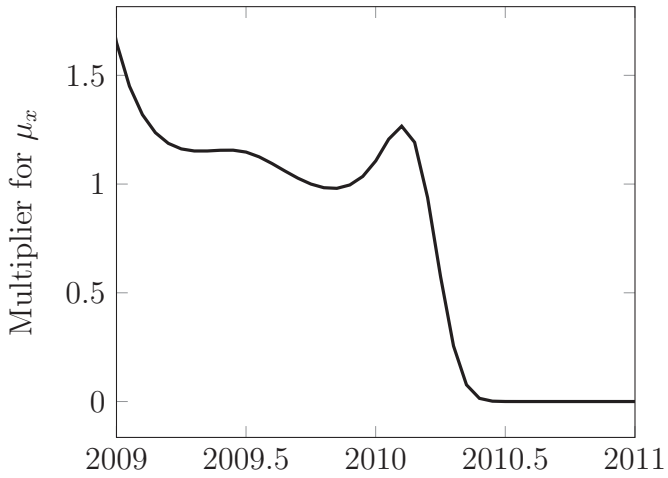
An actuary is tasked with setting a mortality basis for a large annuity portfolio. Eighteen months before the actuary's appointment, annuities in payment were migrated to a new administration platform and records of prior deaths were lost. The actuary only has one-and-a-half years of experience

data. Annuities set up before the migration are therefore like Cases D and E in Figure 1, whereas newer annuities set up post-migration would be like Cases A and B.

The portfolio has scale in terms of the number of annuities, but it has no depth in terms of historic data. Using an annual q-type GLM makes this depth problem even worse: the requirement for a complete year of potential exposure means discarding a third of what little historical data is available. The most recent half-year of experience would be discarded in order not to have mortality levels under-stated due to delays in reporting of deaths. Much of the new business written would therefore also be excluded from q-type analysis, robbing the actuary of information on possible selection effects.

In contrast, a continuous-time solution is to model μ_x and use all eighteen months of data, including the new business. Delays in reporting of recent deaths could be handled either by using a parametric OBNR term [Richards, 2022b] or using the approach of Richards [2022c], as illustrated in Figure 17. In the latter case, the level of mortality to use for a long-term purpose like valuation would have to be decided by judgement, say by picking the modelled (and thus smoothed) mortality level mid-way between a winter peak and a summer trough.

Figure 17: Mortality multiplier over time for French annuity portfolio, with multipliers standardised at 1 for October 2009. The data extract was taken in min-2010, and the impact of reporting delays (OBNR) is very pronounced.



Case Study A.6

A pricing analyst is looking for risk factors to use for annuitant mortality. One of the data fields available is whether an annuity has an attaching spouse's benefit, i.e. a contingent annuity that commences on death of the first life. Statistical modelling suggests that the presence or absence of a spouse's benefit is highly significant, and has a mortality effect on a par with the sex of the annuitant. The direction of the effect is plausible, with single-life annuitants exhibiting higher mortality [Ben-Shlomo et al., 1993].

However, the analyst is aware of the potential for record corruption from Case Study A.2. Specifically, if the main annuitant dies and the spouse is then discovered to have pre-deceased them, the record of the contingent benefit might be removed during death processing. This would make the annuity look like a single-life annuity from outset, and thus distort the apparent mortality differential. The analyst emails the IT department with his concerns and is told that they are unfounded. Skeptical, he makes a further telephone enquiry and is again assured that the feared corruption is not present.

Believing that he may have found an important new risk factor for pricing, the analyst writes a note to the chief actuary. The analyst then remembers that these quality assurances come from the same IT department that lost all the mortality data in Case Study A.5. The analyst performs a third investigation, and discovers that his original suspicions were correct. The excess mortality for single-life annuitants has been grossly exaggerated by the deletion of pre-deceased spouse records.

Case Study A.7

An insurer with a large annuity portfolio has an administration system that automatically deletes records of annuitants who died more than two years in the past. Annuities are set up like Cases A and B in Figure 1, but after two years a Case A annuity becomes like Case D, while Case B annuities simply disappear from the administration system entirely. An urgent request is made to stop this rolling deletion of valuable data.

Similar to Case Study A.5, an annual q-type GLM only permits the use of a single year of data — the most recent six months are ignored due to late-reported deaths, leaving 1.5 years of data, but a further half-year of unaffected data also has to be discarded to meet the requirement of only using complete years.

As in Case Study A.5, a continuous-time solution is to model μ_x and use all two years of data. Delays in reporting of recent deaths could be handled either by using a parametric OBNR term [Richards, 2022b] or using the approach of Richards [2022c], as illustrated in Figures 6 and 17. As in Case Study A.5, the level of mortality would be chosen by judgement, such as a modelled value from spring or autumn.

Case Study A.8

A UK insurer has historically written annuities based solely on the experience of its own maturing money-purchase pension policies. The insurer decides to enter the open market for non-underwritten annuities. However, there is a parallel open market for underwritten annuities, where better terms are offered to lives with demonstrable health conditions [Ainslie, 2000]. The insurer is concerned that the existence of the underwritten market creates selection effects in the new open-market standard annuities that it is writing. The experience data of annuitants from its own maturing pension policies are not believed to have such selection effects, so the new mortality pricing basis contains some guesswork [Richards and Robinson, 2005].

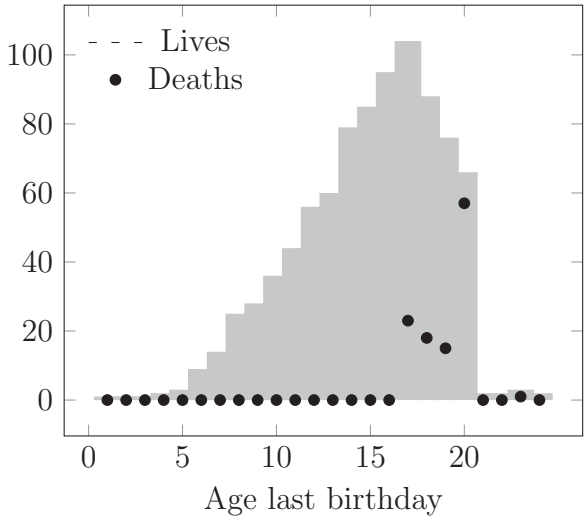
At the end of the first year of the new strategy the insurer wants to urgently (i) validate its pricing assumptions with respect to selection and (ii) set a valuation mortality basis for the new open-market annuities. This cannot be done particularly cleanly using q_x models, as a full year of exposure has not yet been achieved for the new business. In contrast, a continuous-time analysis based on μ_x can easily handle the fractional exposures of the annuities written during the year, with no special modifications required; see Richards [2020, Section 6] for an example.

See also the related discussion around Figure 14.

Case Study A.9

A medium-sized UK pension scheme provides data for analysis. Inspection of the data suggests that temporary and commuted pensions have been erroneously marked as deaths; see Figure 18, where almost all 20 year-olds are allegedly die. While this mistake is irrelevant for modelling mortality at ages 50 and over, it provides an indication that the data have not been properly processed. In particular, the mortality of recipients of small pensions may be considerably over-stated due to commuted cases being misclassified as deaths.

Figure 18: Lives and "deaths" at young ages for medium-sized UK pension scheme. Temporary pensions to children cease once they leave full-time education, but such cessations have been erroneously labelled as deaths in the data extract. Source: own calculations.



Case Study A.10

A Canadian pension plan will not provide full dates of death, as Canadian privacy laws apply to those who have died in the past twenty years [PIP, 2018]. Instead, only the year and month of death are provided.

Since we only know the month in which death occurs, this is *interval censoring* [Macdonald, 1996, Section 2.3]. A continuous-time model can be built assuming that deaths take place on the 15$^{\text{th}}$ of the month. The distortion will be small, with the actual date of death being on average only a week away from this assumption (and with average deviations being close to zero). Note, however, that the tracking statistic of Richards [2022b] will be affected by this loss of information, and the Kaplan-Meier survival curve will have a more ridged appearance than would be the case with exact dates of death.

Case Study A.11

An insurer is thinly capitalised with a small solvency margin. In the hope of lowering its capital requirements, the insurer asks for reinsurance quotations for its large annuity portfolio . However, even the cheapest reinsurance quotation has a mortality basis that is stronger than the insurer's existing reserving basis. This would require the insurer to pay an up-front premium, which it cannot afford.

While the insurer can of course decline to reinsure, it cannot now leave its annuitant mortality basis unchanged. It has just discovered the market price for its annuitant mortality, and must now strengthen its mortality basis.

Case Study A.12

An insurer with a large annuity portfolio needs to plan staffing levels. In particular, the work processing death notifications is often upsetting and stressful, and limits are in place for the amount of time an individual member of staff can be assigned to this task. The insurer uses a mortality model with a season term (see Section 4.4) to make short-term forecasts of likely deaths, thus informing staffing requirements and rotas.

Case Study A.13

An insurer has assumed the annuity liabilities of another insurer. In the UK this is a court-sanctioned process known as a Part VII transfer [FSM, 2000, PartVII]. The IT department sets up the transferred annuities on the administration system with the original commencement date. This makes it difficult to separately identify the transferred annuities, which is problematic because there are periods of survival between commencement and transfer without corresponding deaths.

A continuous-time solution is to use the date of the first annuity payment on the new system as a proxy for the transfer-in date. This allows the new liabilities to be treated like Cases D and E in Figure 1. As in Case Study A.4, such a solution works best for annuities paid relatively frequently, e.g. monthly.

Case Study A.14

In contrast to Case Study A.13, a system migration takes place where the commencement date on the new system is the date of migration, not the original date of retirement (which is then lost).

The migrated cases look like Cases A and B in Figure 1, but in reality they are like Cases D and E. The exposure period can be calculated, but no modelling of duration or selection effects can take place due to not having the actual retirement date.

Case Study A.15

An insurer runs a programme to identify unreported annuitant deaths. Instead of using the actual date of death, the insurer processes the discovered cases *en masse* using the date of processing, resulting in a large spike in deaths recorded for that particular date.

Here the survival times are unreliable for the mass-processed cases. Superficially, it might seem preferable to use a model that specifically discards precise survival times, i.e. a q_x analysis. However, this is only likely to be true if such exercises are carried out sufficiently regularly that the inaccuracy of the date of death is modest. There is a strong financial incentive to run regular existence-verification exercises, say quarterly, as each identified case will release reserves. It is particularly beneficial to run such exercises prior to the year-end statutory valuation.

Case Study A.16

A UK insurer has written a lot of bulk-annuity business and thus has significant liabilities for both pensions in payment and deferred pensions. The actuary is tasked with setting a basis for both sub-types of business, which are administered on the same computer system.

The occurred-but-not-reported (OBNR) deaths for pensions in payment is a relatively contained problem. While there are reporting delays, they can be measured and allowed for, as described in Section 4.3. Part of the reason for this is that the insurer has a constant feedback mechanism — most UK pension payments are made via automated bank credit [Bac, 2019], and there is a specific rejection code for an attempt to pay money into the account of a decedent (ARUCS2 — Beneficiary Deceased).

However, the situation for deferred pensions is different — with no premiums to collect or payments to make, the reporting of deaths of deferred pensioners is typically sparse with very long delays. It is not uncommon for the insurer to only find out about a past death when it tries to make the first pension payment at retirement age. It is therefore common for the mortality data for deferred liabilities to be unreliable, and for a basis to be derived without reference to experience data.

The situation for deaths of deferred pensions can sometimes be improved by the regular use of third-party services, which provide matches to known deaths of varying degrees of confidence based

on name, date of birth and last known address. However, it is important to use such external data thoughtfully, as clumsy application can lead to bad publicity [Tims, 2024].

Case Study A.17

The German multi-employer pension scheme in Richards et al. [2013] stores only the year of death. The data collected by the pension scheme have been designed for q-style analysis only. The exact date of death could be obtained from another source, but the third-party administrator demands payment for bespoke work not covered by the service contract. An additional issue is that pensioners dying during the calendar year of retirement are not recorded at all, neither in the in-force at the start of the year nor in the deaths.

If this were an insurer concerned about anti-selection, as in Case Study A.8, then it might be worth paying for the exact dates of death for a continuous-time model. However, for a pension scheme with slowly changing membership, the cost was not felt to be justified. A q_x model could be built with the data, or else a μ_x model assuming that all deaths occur in the middle of the year. Two further consequences of not knowing the date of death are that (i) the tracking statistic of Richards [2022b] cannot be used, and (ii) the Kaplan-Meier survival curves will take a pronounced step shape, as exact ages of death are not known.

Case Study A.18

An annuity portfolio has been administered continuously since outset on the same computer system without any record deletion, archiving of historic deaths or migration onto or off the system. All annuity records are like Cases A, B and C in Figure 1, and it would be possible to model mortality along these lines if (i) mortality levels were stable, (ii) mortality differentials were stable and (iii) selection effects were stable. However, in practice these things change over time, so most organizations restrict their modelling to the most recent experience, such as the past five years. This means that older annuities are deliberately turned from Cases A and B into Cases D and E, respectively (and very old Cases B and C are excluded entirely). See also Figure 11 for an illustration of the application of such an investigation "window".

Note that such continuity of experience can also be found with some very large, self-administered pension schemes. The outsourcing of pension-scheme administration is invariably bad news for the mortality experience data, especially changes of administrator.

Case Study A.19

An analyst at a UK insurer is directed to use the annual valuation extracts for analysis. The extracts are policy-orientated and have been validated for year-end statutory balance-sheet calculations. However, the valuation extracts contain neither names nor postcodes, so the analyst cannot perform the deduplication necessary for statistical modelling [Macdonald et al., 2018, Section 2.5]. The absence of postcodes further means that geodemographic mortality modelling cannot be done either [Macdonald et al., 2018, Section 2.8.1]. Finally, the annual 'snapshot' nature of the valuation extracts means that policies that commence and terminate during the calendar year are not included, thus hampering the modelling of anti-selection.

The analyst therefore arranges for a direct extract of policy records from the administration system. Also, since mortality models routinely use geodemographic profiles [Madrigal et al., 2011], a request is made to add postcodes to future valuation extracts.

Case Study A.20

Many valuation systems are heavily restricted in the mortality risk factors they can cope with. It is common to have the flexibility to upload a new mortality table of q_x rates, but this typically limits the calculations to using just age and sex as risk factors. Beyond this, such systems sometimes also allow the application of a multiplier to the q_x rates for portfolio-specific customization. This is a problem for modern mortality analysis, which typically uses many more risk factors than just age and sex — see the list in Section 1.

One solution is the equivalent-reserve method popularised by Willets [1999]. This involves calculating an annuity factor using the multi-factor mortality model, then finding the percentage of the given mortality table that produces the same value. This is done for various risk-factor combinations in Richards et al. [2013, Table 13]. However, this approach can be extended to the entire portfolio valuation, thus automatically weighting the mortality risk factors by their impact on the liabilities. The resulting table percentages can then be fed into the valuation system. An important point to note is that financial distortions occur if the resulting table percentage falls far outside the 85-115% range. If this is the case, then it is crucially important to change the reference table so that the percentages are closer to 100%.

Case Study A.21

A US insurer is bidding for a multi-billion dollar buy-out of a large pension plan. The proposed transaction is so large that on its own it will meet the insurer's new-business target for this type of risk. The pricing committee is both keen to win the business, but also worried about the financial impact of possibly getting the mortality basis wrong. A request is made for a statistically rigorous confidence interval around the best estimate of current mortality in the pension plan (the future-improvement assumption is set centrally at the corporate level).

The pricing basis was derived using an individual survival model calibrated to the pension plan's experience data. This allowed the analysts to use the methodology of Richards [2016] to measure the sensitivity of the liability value to the uncertainty surrounding the parameter estimates. A process of simulating alternative parameter estimates consistent with the covariance matrix produced a thousand alternative valuations of the liabilities. The relevant quantiles of these simulated liabilities were then back-solved using the equivalent-reserve method of Case Study A.20 to produce a confidence interval for the best-estimate mortality pricing assumption.

Case Study A.22

An underfunded UK pension scheme is approached by an advisor suggesting that health questionnaires should be sent to pensioners. Those pensioners qualifying for enhanced annuities due to poor health will then have annuities bought from a specialist insurer. Since the enhanced annuities will typically cost less than the notional reserve held in respect of these pensioners, the advisor argues that this will improve the scheme's funding position.

However, if pensioners in poorer health are selected out in this way, the remaining lives are by definition the healthier, longer-lived ones. Thus, the scheme's mortality basis therefore needs to be strengthened for the remaining members after the enhanced-annuity exercise, and there should be no net improvement in the funding position. Furthermore, carrying out this exercise limits future buy-out options for the scheme — some insurers in the UK specifically ask about past ill-health buy-outs, and refuse to quote if such an exercise has been carried out in the preceding five years.

Case Study A.23

An insurer decides that it no longer wants to write annuities for its portfolio of vesting pension policies. It decides instead to operate a small panel of annuity providers and take commission as an introducer. This way the insurer can make money from the annuities without having to carry the liabilities on its balance sheet. The insurer provides the mortality experience of its existing annuity portfolio to prospective annuity writers seeking to join the panel. The mortality-experience data contain details of the pension product type and whether a guaranteed annuity rate (GAR) is present.

The panel annuity writers need to calibrate a mortality model specific to the profile of this portfolio, which can be done using survival models; see Richards [2008]. It is important to create a differentiated mortality model that, where statistically justified, uses risk factors beyond age, sex, pension size and postcode; see Madrigal et al. [2011] and Richards et al. [2013]. Product type, distribution channel and annuity options selected by the policyholder are often signals for higher or lower mortality. The panel annuity writer with the least sophisticated mortality basis will be selected against by winning mainly mispriced business [Ainslie, 2000].

Case Study A.24

A long-established company has a large pension scheme. The company's market value has steadily shrunk over the years, while the pension scheme has become rather large. The company wants the pension scheme to buy backing annuities to reduce the volatility in its published financial statements. The nature of the company's business has changed radically over time, and the workforce and pensioners have changed along with it. The older pensioners are mainly drawn from a formerly blue-collar workforce with occupational exposure to dust and chemicals. In contrast, recent retirees are from a more white-collar workforce running a more services-orientated business.

Pension size is a less-reliable proxy for socio-economic differentials here. Former blue-collar workers tended to have longer periods of service with lower salaries. However, their pension amounts overlap substantially with former higher-salaried white-collar staff, who also had shorter periods of service. A crude comparison against a standard table would not pick up the shift in the nature of the socio-economic profile. This is particularly important here because the older, blue-collar pensioners dominate the death counts, while the younger, white-collar pensioners dominate the liabilities. The preferred solution is to fit a scheme-specific survival model, using the postcode-driven geodemographic profile as a means of identifying the socio-economic differentials; see Richards [2008], Madrigal et al. [2011] and Macdonald et al. [2018, pages 32–36] for details.

Case Study A.25

A UK insurer wants to reinsure a large block of annuity business exceeding £10 billion in reserves. To reduce exposure to a single counterparty, and to open the bidding to smaller and medium-sized reinsurers, the portfolio is split into four different blocks to be bid on separately. The blocks have not been selected randomly, but according to internal criteria of the insurer. The insurer provides the mortality experience of the entire annuity portfolio to prospective bidders.

Even if they are only planning on bidding for one or two blocks, bidders will use the entire data set to calibrate a mortality model for pricing. It is important that all relevant risk factors are built into the model, since bidders using simplistic pricing are at risk of winning only the blocks they misprice [Ainslie, 2000].

Case Study A.26

A large UK pension scheme wants to buy annuities to back the pensions in payment. There is an executive sub-scheme containing former executives, senior managers and the surviving spouses of both. However, the executive sub-scheme is too small for a stand-alone analysis, especially a crude comparison against a standard table. Instead, a mortality model is calibrated by the bidding insurer using age, sex, pension size, geodemographic profile and a binary flag denoting membership of the executive sub-scheme. Unsurprisingly, members of the executive sub-scheme have lower mortality than non-members, even after allowing for risk factors like pension size and geodemographic type. This turns out to be significant for pricing the buy-out, as the executive sub-scheme has a lower average age and thus accounts for a larger share of reserves than is implied by their share of annual pensions in payment.

Case Study A.27

A UK pension scheme has far heavier levels of mortality than predicted by a standard commercial model allowing for age, sex, pension size and geodemographic profile. Much of the workforce in question had past occupational exposure to asbestos, so there are grounds to believe that this is a real result. However, the office-based members of the workforce did not have this exposure. The question is whether the experience data is credible enough to justify a weaker mortality basis than the standard commercial model? And, if so, what is statistically justifiable?

A bespoke mortality basis is derived for the pension scheme using survival models; see Richards et al. [2013] and Madrigal et al. [2011] for examples of scheme-specific risk factors that can be used.

Case Study A.28

A UK insurer wishes to improve its mortality model by using postcodes and geodemographic profiles. However, upon extracting the mortality experience data it finds that almost all deaths have the same postcode, namely that of the insurer's head office. It transpires that servicing staff change the address of deceased annuitants to avoid accidentally sending post to the dead. From a strict data perspective this is a misuse of the address field, but the business requirement of not further upsetting surviving spouses takes precedence.

It is impossible to calibrate a geodemographic mortality model when only survivors have a residential postcode. The insurer needs to extract the history of address changes for the annuitants, and select the residential postcode for geodemographic profiling. However, it is also possible for an address to be changed to that of a solicitor for an annuitant entering care. Fortunately, good geodemographic profilers in the UK typically distinguish between residential and non-residential postcodes. If a history of addresses is available, the most recent residential profile should be used for modelling.

Case Study A.29

A UK insurer wishes to improve its mortality model by using postcodes and geodemographic profiles. Analysis reveals that UK-resident annuitants without a geodemographic profile have high mortality. Further investigation reveals that postcodes are recorded for these annuitants, but that these postcodes are not listed in the geodemographic profiler database. The geodemographic profiler is used for marketing purposes, and only contains currently valid UK postcodes, of which there are around 1.7 million. There are a further half-million retired postcodes that have been used by the Post Office in years past, but which are no longer in use. However, dead annuitants are more likely to have a retired postcode, and so a bias arises.

The servicing department is unwilling to incur the expense and disruption of updating address

details for deceased annuitants. The solution is to get the provider of the geodemographic profiler to back-fill their database with historic postcodes and provide the relevant geodemographic profiles for them. This reduces the bias by assigning geodemographic codes to dead annuitants with out-of-date postcodes.

Case Study A.30

A critical-illness (CI) portfolio has a total of 267 CI claims from almost 130,000 life-years of exposure. A reinsurer wants to assess the extra risk — if any — posed by smokers. However, there are two apparent problems. First, the smokers account for just 56 of the claims and 20,000 life-years of exposure. Second, lapses are vastly more common than CI claims.

The presence of a dominant competing decrement makes a q-type model a poor choice for modelling mortality; see Section 6. The actuary therefore treats lapses as censored observations and fits a μ-type model for age, sex and smoker status, among other covariates. Despite the small number of smoker CI claims, the model shows conclusively that smokers have a 57% higher CI claim rate than non-smokers, with a standard error of 15%. This gives a p-value of 0.02% for the effect of smoking, i.e. the result is highly significant even for an apparently small number of claims.

Case Study A.31

An investor owns a large portfolio of home-reversion plans. These are residential properties that have been purchased while granting the former owners lifetime tenancy. These are not equity-release mortgages — the investor has purchased the actual properties. Nor are these usufruct arrangements — only the tenants named on the lease can live in the property, and they cannot rent it out to other tenants.

The investor makes money from paying the former owners less than the market value of the property in exchange for the right to live there rent-free. The investor can only take unencumbered possession of the property when the last-named tenant leaves. There are two main decrements of interest to the investor: mortality and inception rates for long-term care (LTC). In the supplied experience data deaths occur nearly four times more frequently than LTC events. However, there is also a number of cases where the tenants bought back their property.

The presence of two competing risks and another decrement makes a q-type model a poor choice; see Section 6. The actuary therefore fits two μ-type models, one each for mortality and LTC inception. For the mortality model, both LTC and buyback events are treated as censored observations. For the LTC model, both deaths and buybacks are treated as censored observations.

Case Study A.32

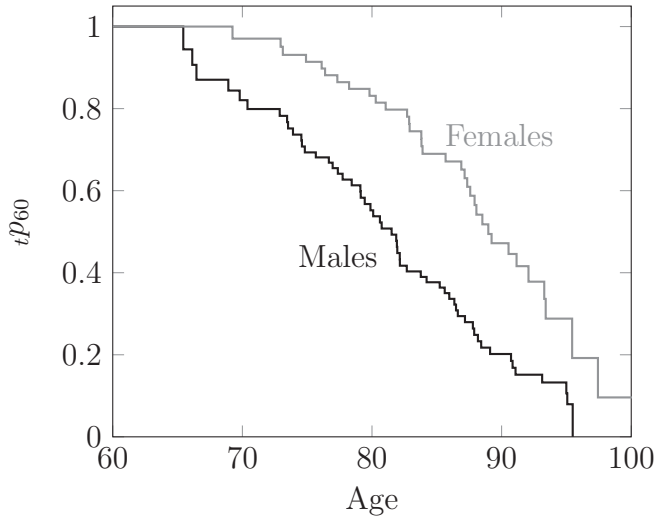
An actuary is asked to analyse a second, smaller portfolio of UK home-reversion plans (see Case Study A.31 for background details on this type of business). In addition to the expected decrements of mortality and entry into long-term care, the actuary finds some additional decrements, such as voluntary surrender of the lease and divorce of married tenants. Voluntary surrender of the lease was deemed to be a precursor stage of long-term care, e.g. when an elderly tenant moves in with one of their adult children. Divorces are trickier — they are obviously neither mortality or care-related, but divorce leads to a named tenant being taken off the lease, thus bringing forward the repossession of the property in the same manner as death or entry into long-term care.

The presence of multiple competing risks makes a q-type model a poor choice; see Section 6. The actuary includes voluntary surrenders as events in the modelling of the care decrement. Divorces are handled as censoring events for the lives taken off tenancy agreements, despite their financial impact

being similar to mortality and care events.

A plot of the mortality-only survival curves after allowing for competing risks is shown in Figure 19. This exhibits a gap of eight years in the median survival age of males and females, which is far larger than the equivalent gap in standard mortality tables. The gap in Figure 19 is, however, consistent with the experience data of Case Study A.31 (not shown), suggesting that portfolio-specific analysis is essential for this class of business.

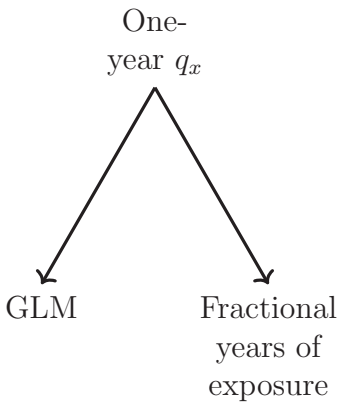
Figure 19: Mortality-only Kaplan-Meier survival curves for lives in a home-reversion portfolio with competing risks including long-term care inception, property buyback, voluntary surrender of lease and divorce.



B Fractional years of exposure for q-type models

A recurring theme in this paper is that annual q-type models typically require the analyst to discard fractional years of exposure. This is certainly true when using standard packaged models, such as GLMs; Figure 20 shows the trade-off available. Even if a likelihood cannot lead to a GLM it may still be perfectly useable as a model, see Macdonald and Richards [2024, Section 3.7].

Figure 20: Trade-off decision tree for one-year q-type models. One cannot have both a GLM and fractional years of exposure.



Note that the trade-offs with q_x models go far further than Figure 20. For example, an analyst may decide that she has sufficient data to not need fractional year of exposure, and so can use a GLM for a one-year q_x. However, this still means that she cannot simultaneously model period effects and selection effects, as described in Section 4.2.

In this appendix we explore some of the options available to a q_x modeller for handling fractional years of exposure. As we will see, these options are incompatible with the use of a GLM, but they are achievable with a bespoke likelihood optimized in another way. However, if one is optimizing a likelihood directly, one might as well use a survival model.

In the rest of this appendix we have a single life aged x that is observed for t years with $0 < t < 1$. d is an indicator variable taking the value 1 if the life is dead at age $x + t$ and zero otherwise. We consider various ways of incorporating ${}_tq_x$ into a likelihood, where q_x is the probability of death in a single year of age allowing for all risk factors.

B.1 Standard Bernoulli GLM likelihood

When a single full year of exposure is present, the standard individual contribution to a Bernoulli log-likelihood, ℓ^{GLM}, is given in Table 1:

Table 1: Log-likelihoods for q_x models, with and without allowance for fraction years of exposure.

Model	Log-likelihood	Formula
GLM	ℓ^{GLM}	$(1-d)\log(1-q_x) + d\log q_x$
Constant hazard	ℓ^{CH}	$t(1-d)\log(1-q_x) + d\log(1-(1-q_x)^t)$
UDD	ℓ^{UDD}	$(1-d)\log(1-tq_x) + d\log q_x$
Balducci	ℓ^{B}	$(1-d)\log(1-q_x) + d\log q_x - \log(1-(1-t)q_x)$
Madrigal et al. [2011]	ℓ^{M}	$(t-d)\log(1-q_x) + d\log q_x$

B.2 Constant hazard

Under the constant-hazard assumption [Broffitt, 1984, Table 1], $1 - {}_tq_x = (1-q_x)^t$. The individual contribution to the Bernoulli log-likelihood, ℓ^{CH}, is given in Table 1. We can see from comparing GLM and CH log-likelihoods that the assumption of a constant hazard to allow for fractional years of exposure cannot be accommodated within a GLM because of the quite different form of the multiplier for d on the right.

B.3 Uniform Distribution of Deaths (UDD)

Under the UDD assumption [Broffitt, 1984, Table 1], ${}_tq_x = tq_x$. The individual contribution to the Bernoulli log-likelihood, ℓ^{UDD}, is given in Table 1. Note that we have dropped an additive term t^d as it does not involve the parameter q_x, and so does not change inference. We can see from comparing the GLM and UDD log-likelihoods that the UDD allowance for fractional years of exposure cannot be accommodated within a GLM due to the presence of t in $\log(1 - tq_x)$.

B.4 Balducci assumption

Also known as the hyperbolic assumption, the Balducci assumption is based on flawed reasoning [Hoem, 1984] and implies a decreasing hazard with age (Broffitt [1984, page 85] described the Balducci assumption as leading to "some unreasonable results"). It is therefore deprecated in favour of more realistic assumptions like the constant hazard or UDD. We include it here merely for historical curiosity; although it often appears in exercises for actuarial students, the Balducci assumption should never be used in real work. Under the Balducci assumption [Broffitt, 1984, Table 1]:

$$ {}_tq_x = \frac{tq_x}{1 - (1 - t)q_x}. \tag{9}$$

The contribution to the Bernoulli log-likelihood, ℓ^{B}, is given in Table 1. There is an additive t^d term that does not affect inference for the parameter q_x and is therefore omitted. Once again, we can see from comparing the GLM and Balducci log-likelihoods that the Balducci assumption to allow for fractional years of exposure cannot be accommodated within a GLM because of the extra third term on the right of ℓ^{B}.

B.5 Weighting the log-likelihood

Madrigal et al. [2011, Section 4.3.1] claimed that fractional years of exposure could be included in a q-type model by "weight[ing] the contribution of each of the membership records according to its exposure to risk in a year". However, this does not allow for fractional years of exposure, as discussed in Madrigal et al. [2011, page 61]. Madrigal et al. [2011, page 62] claim that their individual contribution to the Bernoulli log-likelihood, ℓ^{M}, is as given in Table 1 (after some re-arrangement). However, ℓ^{M} cannot be part of a GLM because of the $(t - d)$ coefficient of $\log(1 - q_x)$. Furthermore, weighting observations by the interval length is not a valid way to allow for fractional exposures.

B.6 Conclusions

As per Figure 20, the only way to have a one-year q-type GLM is to discard fractional years of exposure. If the actuary insists on a q-type model, but relaxes the requirement to have a GLM, then it is possible to directly use a likelihood with either the constant-hazard or UDD assumptions. Apparent alternatives are either deprecated (such as the Balducci assumption [Hoem, 1984]) or simply invalid (such as weighting the exposures as in [Madrigal et al., 2011, Section 4.3.1]).

However, a one-year q-type model still involves compromises, such as the inability to simultaneously model period effects and selection effects demonstrated in Section 4.2. If an analyst is willing to directly maximise a likelihood, then a survival model for μ_x is both simpler and less restrictive.

B.7 Open-source software implementations

The R [R Core Team, 2021] `stats` package contains the `glm()` function for fitting ordinary GLMs for q-type and μ-type models. However, q-type GLMs can only be fitted where the exposure intervals are of identical length, i.e. no fractional varying years of exposure.

C Discarding Data

C.1 Background

Methods based on the estimation of q_x came naturally in early studies of mortality, since the aim was to construct a life table. Notably, the seminal attempt to give an *explanatory* model of mortality [Gompertz, 1825] was based on μ_x, but this lead was pursued only sporadically in the following two centuries. Even in 1988, the landmark paper of its time [Forfar et al., 1988] treated the estimation of μ_x and q_x equally. There are, of course, situations where the data make a binomial model most appropriate (see Case Study A.17 in Appendix A) but otherwise the q-type modeller faces a choice between discarding data that do not suit the hypothesis, or adapting the hypothesis to suit the actual data.

When starting out in mortality studies, it is often stated, or simply assumed, that all data must be included in the analysis — that discarding some selected data must bias the results — thus presenting the q-type modeller with a problem. However, this is rarely demonstrated, which is understandable, as doing so requires much of the apparatus which has not yet been seen. In this appendix, we consider the legitimacy of discarding parts of the data. Our approach is informal, drawing on the more formal arguments in Andersen et al. [1993], which are based on counting processes. We refer the reader there for a more rigorous treatment.

For simplicity we confine attention to an interval of age $[x, x+1]$, and suppress the initial age x, so we consider the interval of time or duration $t \in [0, 1]$. We assume a constant force of mortality μ on this interval, therefore also a binomial model with parameter $q = 1 - \exp(-\mu)$. This is enough for our purpose because:

(a) if $\hat{\mu}$ is the maximum likelihood estimate of μ then $1-\exp(-\hat{\mu})$ is the maximum likelihood estimate of q; and

(b) our analysis can be extended to non-constant parameters μ_{x+t} and q_{x+t} but at the cost of much more notation and the use of product-integrals (see Macdonald and Richards [2024]).

C.2 Data and Likelihood

We observe M individuals, and for the ith individual we have the following data:

(a) durations r_i of entry to observation and t_i of exit from observation, with $0 \leq r_i < t_i \leq 1$; and

(b) an integer-valued indicator d_i of the reason for observation ceasing; $d_i = 1$ if the reason was death, and $d_i = 0$ if the reason was right-censoring.

The ith individual is observed on the interval $[r_i, t_i]$, which we denote by Δ_i, and we denote the interval $[0, 1]$ by Δ. From Macdonald and Richards [2024] we borrow the process $Y^i(t)$, equal to 1 if the ith individual is under observation 'just before' time t, and equal to 0 otherwise (the 'just before' is a technicality we can ignore).

To extend the model to allow for voluntary lapsation, all that is needed is to (i) assume that a constant force of lapsing ν operates alongside μ, and (ii) to define another indicator w_i, taking the value 1 if lapsing is the reason for observation ceasing, and 0 otherwise.

The contribution of the ith individual to the likelihood is the factor L_i defined as:

$$L_i = \exp\left(-\int_{\Delta_i} Y^i(t)\,(\mu+\nu)\,dt\right)\,\mu^{d_i}\,\nu^{w_i}, \tag{10}$$

the total likelihood, denoted by L, is:

$$L = \prod_i L_i, \tag{11}$$

and the MLE, denoted by $\hat{\mu}$, is:

$$\hat{\mu} = \frac{\sum_i d_i}{\sum_i \int_{\Delta_i} Y^i(t)\, dt}. \tag{12}$$

C.3 Obligate versus Random Right-Censoring

Observation of the ith individual is left-truncated if $r_i > 0$. Observation is right-censored if $d_i = 0$, but we distinguish two kinds of right-censoring:

(a) *random* right-censoring means the policy lapsed, $w_i = 1$; and

(b) *obligate* right-censoring means the individual was observed to survive until some *predetermined* time (in a sense to be made precise below) at which observation necessarily ended. Examples of such times are:
   (1) the end of the interval $[0, 1]$ (or more generally, the rate interval);
   (2) the date of data extraction;
   (3) the date of transfer out of the portfolio; or
   (4) the end-date of the investigation (for example, in a CMI-style investigation).

The key point about obligate right-censoring is that the time of obligate right-censoring is known as soon as the individual enters observation[3]. Technically, the time t'_i when observation on the ith individual must end is a random variable with $r_i < t'_i \leq 1$, but its value becomes known when the ith individual enters observation at time r_i (that is, t'_i is $\mathcal{F}_{r_i}$-measurable). So we have three possible cases:

(a) Case 1: obligate right-censoring, $d_i = w_i = 0$, $t_i = t'_i$;

(b) Case 2: random right-censoring, $d_i = 0$, $w_i = 1$, $t_i < t'_i$; and

(c) Case 3: death, $d_i = 1$, $w_i = 0$, $t_i < t'_i$.

The main result of this Appendix is that the data on *selected* lives can be discarded, based on the time r_i of entry and the time t'_i of obligate right-censoring, without biasing the estimation of μ. However, randomly right-censored data cannot be so discarded.

Define $\Delta'_i = [r_i, t'_i]$, the greatest interval on which the ith individual *might* be observed, and define $\Delta = [0, 1]$. Then:

(a) observation of the ith individual is randomly right-censored if $w_i = 1$; and

(b) observation of the ith individual is left-truncated or obligate right-censored if $\Delta'_i \neq \Delta$.

[3]Censoring at a pre-determined time is generally known as 'Type I' censoring in survival analysis (Collett [2003], Andersen et al. [1993]). However, the observation of a subject in an actuarial investigation usually begins left-truncated at a random time (age) and the time of mandatory censoring is then determined dynamically by the age definition and the observational plan; see Figure 11(b). The name 'obligate censoring' seems more descriptive, and a suitable contrast to 'random censoring' caused by lapsing.

C.4 Discarding Left-truncated and Obligate Right-censored Data

If $\Delta_i' \neq \Delta$ then the individual cannot complete a full year's exposure. If the q-type analyst discards such data, the modified likelihood, denoted by L^{Obl}, is:

$$L^{\text{Obl}} = \prod_{\Delta_i'=\Delta} L_i. \tag{13}$$

Assuming that there is no selective reason for entry to be left-truncated or exit to be obligate right-censored, all such data can be discarded, deaths and lapses along with exposure data, without biasing any likelihood-based estimation. Note that this is an assumption that often seems reasonable, but may be hard to verify. Therefore:

$$\hat{\mu}^{\text{Obl}} = \frac{\sum_{\Delta_i' \neq \Delta} d_i}{\sum_{\Delta_i' \neq \Delta} \int_{\Delta_i} Y^i(t)\, dt} \longrightarrow \mu. \tag{14}$$

See Section III.2.2, also Example III.2.3, in Andersen et al. [1993] for a formal justification.

C.5 Discarding Randomly Right-censored Data

If randomly right-censored data are discarded the modified likelihood, denoted by L^{Ran}, is:

$$L^{\text{Ran}} = \prod_{w_i \neq 1} L_i = \prod_{w_i \neq 1} \exp\left(- \int_{\Delta_i} Y^i(t)\, (\mu + \nu)\, dt \right) \mu^{d_i}. \tag{15}$$

If $w_i = 1$ and the ith individual's data is excluded, the time $|\Delta_i|$ for which the individual was exposed to the risk of death and lapse, is removed from the exposed-to-risk, but no deaths are removed from the data. The MLE $\hat{\mu}^{\text{Ran}}$ is:

$$\hat{\mu}^{\text{Ran}} = \frac{\sum_{w_i \neq 1} d_i}{\sum_{w_i \neq 1} \int_{\Delta_i} Y^i(t)\, dt} = \frac{\sum_{i} d_i}{\sum_{w_i \neq 1} \int_{\Delta_i} Y^i(t)\, dt} \geq \frac{\sum_{i} d_i}{\sum_{i} \int_{\Delta_i} Y^i(t)\, dt} = \hat{\mu} \tag{16}$$

where the last inequality is strict if and only if at least one individual lapses. Therefore, discarding randomly right-censored data will bias the estimate of μ upwards, and also the estimate of q. Of course, the argument displayed in (16) is exactly the same as the normal argument for not excluding censored data in a survival analysis.

D Non- and semi-parametric methods

Between the worlds of discrete-time and continuous-time analysis lies a middle ground of non- and semi-parametric methods. These are based on occurrence-exposure ratios, but where the interval is neither fixed nor set by the analyst. Instead, the interval sizes are determined by the data, specifically the gaps between each death and the next. As the portfolio size grows, the gaps shrink to a single day, and thus non- and semi-parametric methods quickly become continuous-time methods for all practical purposes. Actuaries should not be confused by the d/l notation for ratios into thinking that these are q-type methodologies — the occurrence-exposure ratios are estimating mortality over short intervals, i.e. they are estimates of μ.

In this appendix we look at three different methodologies for three different time dimensions: age, duration and calendar time. These methodologies have much to offer practising actuaries, from the data-quality checking of Section 7 to the real-time reporting of Section 8.

D.1 Kaplan-Meier estimator

The methodology of Kaplan and Meier [1958] is non-parametric, and is a re-arrangement of the experience data into a survival curve. This survival curve can be by duration, such as is often the case with medical trials [Collett, 2003, Section 2.1.2]. However, the most useful definition of the Kaplan-Meier estimator for actuaries is the one by age in Macdonald et al. [2018, equation (8.3)]:

$$ {}_t\hat{p}_x = \prod_{t_i \le t} \left(1 - \frac{d_{x+t_i}}{l_{x+t_i^-}} \right), \tag{17} $$

where x is the outset age for the survival function, $\{x + t_i\}$ is the set of 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 occurring at age $x + t_i$. Informally, and following Section 2, we can think of $l_{x+t_i^-}$ as the number of lives alive just after midnight on a given day, with deaths d_{x+t_i} happening around mid-day. Figures 12(a) and 19 show examples of Kaplan-Meier curves.

The Kaplan-Meier estimator in equation (17) is the product-limit estimator of the survival curve discussed in Macdonald and Richards [2024, Section 4.2]. It is also the product-integral of the Nelson-Aalen estimator in equation (19) below.

D.2 Nelson-Aalen estimator

The methodology of Nelson [1958] is non-parametric, and is a re-arrangement of the experience data into a cumulative risk curve. This cumulative risk can be by duration, such as is often the case in medical work:

$$ \hat{\Lambda}(t) = \sum_{t_i \le t} \frac{d_{t_i}}{l_{t_i^-}}, \tag{18} $$

where $\{t_i\}$ is the set of distinct durations at death, $l_{t_i^-}$ is the number of lives alive immediately before duration t_i and d_{t_i} is the number of deaths occurring at duration t_i. Figure 14 shows an example of equation (18) being used to monitor differences in mortality for new annuity business.

However, another useful definition for actuaries is one by age:

$$ \hat{\Lambda}_x(t) = \sum_{t_i \le t} \frac{d_{x+t_i}}{l_{x+t_i^-}}. \tag{19} $$

The Nelson-Aalen estimator is the non-parametric estimator of the integrated hazard in Macdonald et al. [2018, equation (3.23)].

D.3 Fleming-Harrington estimator

The methodology of Fleming and Harrington [1991] is non-parametric, and uses the Nelson-Aalen estimator of the cumulative hazard to estimate the survival curve from age x:

$$ {}_t\hat{p}_x = \exp\left(-\hat{\Lambda}_x(t)\right), \tag{20} $$

where $\hat{\Lambda}_x(t)$ is the Nelson-Aalen estimator from equation (19). The Fleming-Harrington estimator is the non-parametric equivalent of the identity:

$$ {}_tp_x = \exp\left(-\int_0^t \mu_{x+s}ds\right) = \prod_{s\in(0,t]}(1-\mu_{x+s}ds). \tag{21} $$

The Fleming-Harrington estimator is slightly preferable to the Kaplan-Meier estimator for very small data sets, in part because approximate confidence intervals for the latter are not restricted to $[0,1]$ [Collett, 2003, Sections 2.2.1 and 2.2.3]. However, for the data sets of the size that actuaries use there is usually no meaningful difference between the two estimators.

D.4 Relationships

There are important relationships between the Kaplan-Meier, Nelson-Aalen and Fleming-Harrington estimators. The product-integral of a piecewise-constant function such as equation (19) is a discrete product such as equation (17). Therefore, the Kaplan-Meier estimator is the product integral of the Nelson-Aalen estimator; this is exact, not approximate. The Fleming-Harrington estimator in equation (20) is the exponential of (minus) the Nelson-Aalen estimator in equation (19), and is an approximation to the Kaplan-Meier estimator in equation (17) (and a good one for the datasets typically used by actuaries).

In some literature [SAS Institute Inc., 2017, page 5336] an alternative to equation (19) for small samples is given as:

$$ \hat{\Lambda}_x(t) = \sum_{t_i\le t}\sum_{j=0}^{d_{x+t_i}-1}\frac{1}{l_{x+t_i^-}-j}. \tag{22} $$

Equations (19) and (22) are identical if $d_{x+t_i}=1$. Even where some d_{x+t_i} are greater than 1, equation (22) only produces materially different answers with small sample sizes. For actuarial data sets there is little meaningful difference, so we prefer equation (19) for simplicity.

D.5 Semi-parametric methods

Richards [2022b] introduced a semi-parametric estimator in calendar time for investigating short-term fluctuations in mortality. The methodology has two stages, the first of which is to calculate the non-parametric Nelson-Aalen estimator of the cumulative hazard with respect to calendar time, rather than age or duration:

$$ \hat{\Lambda}_y(t) = \sum_{t_i\le t}\frac{d_{y+t_i}}{l_{y+t_i^-}}, \tag{23} $$

where y is the outset calendar time for the cumulative risk, $\{y+t_i\}$ is the set of distinct dates of death, $l_{y+t_i^-}$ is the number of lives alive immediately before date $x+t_i$ and d_{y+t_i} is the number of deaths occurring on date $y+t_i$. Informally, and following Section 2, we can think of $l_{y+t_i^-}$ as the number of lives alive just after midnight on a given day, with deaths d_{y+t_i} happening around mid-day. $\hat{\Lambda}_y(t)$ is the equivalent of equation (19), but with respect to calendar time instead of age. A practical point is that the real-valued calendar time is calculated using the actual number of days in the year, rather than the average of 365.24 given in Section 2. For example, 14th March 2023 would be represented as 2023.1973 (being 72 days into a 365-day year); in contrast, 14th March 2024 would be represented as 2024.1995 (being 73 days into a 366-day year).

The second step is to estimate the portfolio-level hazard in time, $\mu_y(t)$. This ignores the age of individual lives, and so is only useful for investigating fluctuations over relatively short periods of time, where the age profile of the portfolio does not change radically. To estimate $\mu_y(t)$ from equation (23) we need to apply some degree of smoothing over a suitable short interval, c. We thus obtain a semi-parametric estimate of the mortality hazard at a point in time, $\hat{\mu}_y(t,c)$, using c as a bandwidth parameter. This can be done in a number of ways, but a simple one is to use a central difference in the estimator in equation (23) around the time of interest:

$$\hat{\mu}_y(t,c) = \frac{1}{c}\left(\hat{\Lambda}_y(t+c/2) - \hat{\Lambda}_y(t-c/2)\right), \tag{24}$$

where c is selected by the analyst to smooth random variation. When $c \leq 0.5$, $\hat{\mu}_y(t,c)$ will reveal detail such as seasonal variation (Figure 13(a)), mortality shocks [Richards, 2022b, Figure 3] and data-quality issues (Figure 13(b) and [Richards, 2022b, Appendix A]). When $c \gg 0.5$ short-term variations are smoothed out. Over longer periods of time $\hat{\mu}_t(t,c)$ will reflect changes in the age distribution of the portfolio [Richards, 2022b, Figure 2].

Equation (24) is calculated using an extract at a point in time, u, so we could write $\mu_y(t,c,u)$ to make it explicit that the calculation is with respect to a particular data extract at time u. This can be useful when estimating delays in reporting deaths; see Section 4.3 and Richards [2022b, Section 4], where the ratio $\mu_y(t,c,u_1)/\mu_y(t,c,u_2)$ provides an estimate of the proportion of deaths reported up to time u_1, under the assumption that the estimate at time u_2 contains most of the unreported deaths up to time u_1. This presupposes that the gap $u_2 - u_1$ is large enough to allow most OBNR cases up to u_1 to have been subsequently reported. In reality, the estimate of $\mu_y(t,c,u_2)$ at times before u_1 will also be affected by OBNR, but this effect can be minimised by a suitably large gap between u_1 and u_2.

As an alternative to equation (24), $\hat{\mu}_y(t,c)$ could be calculated directly in a single step as:

$$\hat{\mu}_y(t,c) = \frac{2}{c}\sum_{t-c/2<t_i\leq t+c/2} K\left(\frac{2(t_i-t)}{c}\right)\frac{d_{y+t_i}}{l_{y+t_i^-}}, \tag{25}$$

where $K(u)$ is a kernel function for smoothing; see Andersen et al. [1993, Section IV.2.1]. Equation (24) results from using the uniform kernel:

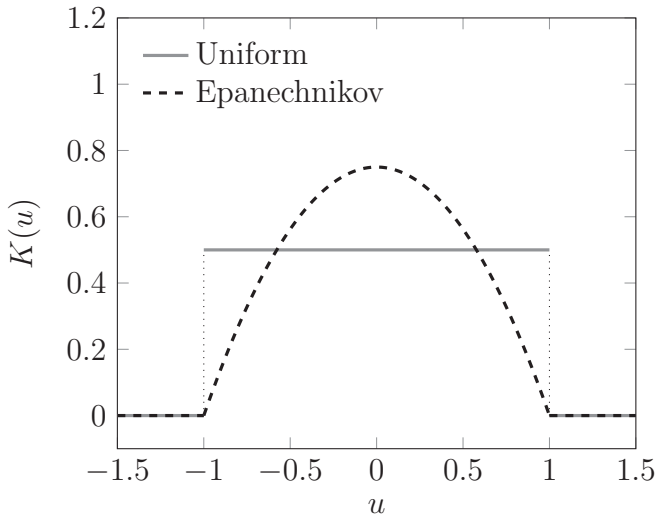
$$K(u) = \begin{cases} \frac{1}{2} & |u| \leq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{26}$$

but other kernel smoothers with greater efficiency are available, such as the Epanechnikov kernel [Epanechnikov, 1969]:

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2) & |u| \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

The kernel functions in equations (26) and (27) are illustrated in Figure 21. The Epanechnikov kernel is likely to be more suitable for applications where measuring short-term fluctuations is important.

Figure 21: Sample kernel-smoothing functions in equations (26) and (27).



D.6 Open-source software implementations

The R [R Core Team, 2021] `survival` package contains the `survfit()` function for computing the Kaplan-Meier and Fleming-Harrington estimators in equations (19) and (22). The Python module `scikit-survival` contains an implementation of the Kaplan-Meier estimator. Macdonald et al. [2018] also contains R source code. Richards [2022b, Appendix C] contains R source code for the semi-parametric estimator in equation (24).

E Individual lifetimes v. grouped counts

In the body of the paper we show why μ-type methods are superior to q-type methods. We also assume that individual records are extracted, as this enables better data-quality checking [Macdonald et al., 2018, Chapter 2]. However, the analyst still had two choices for a μ-type model: the modelling of individual lifetimes (a survival model) or the modelling of grouped counts.

E.1 Individual lifetimes

Suppose we have the following for life i: the entry age into observation, x_i; the time observed, t_i; and the indicator, d_i, which takes the value 1 if life i died at age $x_i + t_i$ and zero otherwise. The contribution, L_i, of life i to the likelihood is then:

$$\begin{aligned} L_i &= {}_{t_i}p_{x_i}\mu_{x_i+t_i}^{d_i} && (28)\\ &= \exp\left(-\int_0^{t_i}\mu_{x_i+s}ds\right)\mu_{x_i+t_i}^{d_i} && (29)\\ &= e^{-\Lambda_{x_i}(t_i)}\mu_{x_i+t_i}^{d_i}, && (30) \end{aligned}$$

where $\Lambda_x(t) = \int_0^t \mu_{x+s}ds$ is the *integrated hazard*; see Macdonald et al. [2018, equation (3.23)]. The corresponding contribution of life i to the log-likelihood, ℓ_i, is:

$$\begin{aligned} \ell_i &= \log L_i && (31)\\ &= -\Lambda_{x_i}(t_i) + d_i \log \mu_{x_i+t_i}. && (32) \end{aligned}$$

For a data set of n lives, the full likelihood is then $L = \prod_{i=1}^{n} L_i$ and the corresponding log-likelihood is $\ell = \sum_{i=1}^{n} \ell_i$. Modelling individual lifetimes is the most elemental mortality model; Macdonald and Richards [2024] call this a 'micro' model. One benefit of modelling individual lifetimes is that more information can be used on individual attributes and covariates. For example, Richards [2022a] showed how to include the exact pension amount as a continuous covariate in mortality modelling.

E.2 Grouped counts

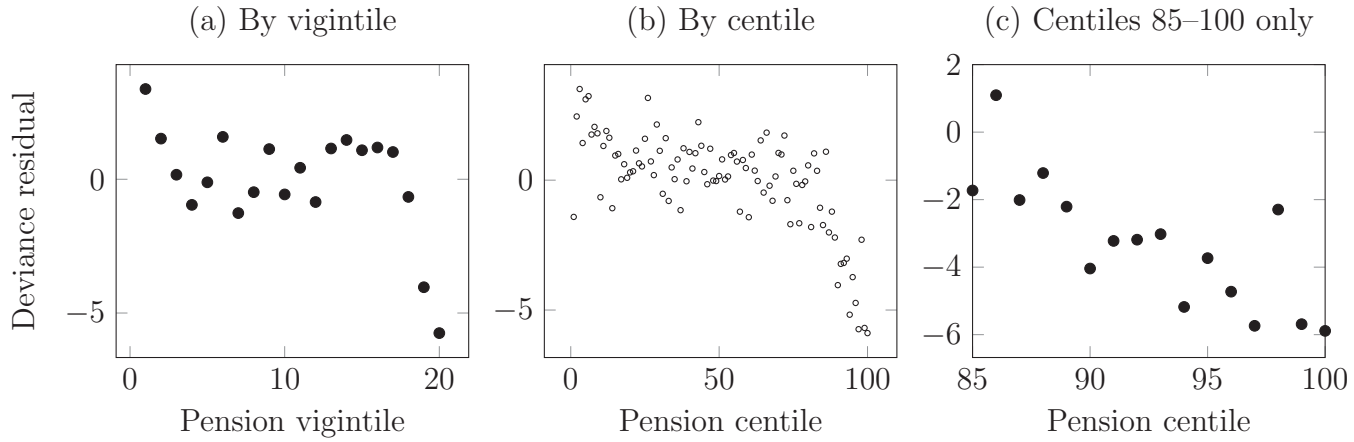
An alternative approach is to create homogeneous sub-groups of shared characteristics. Suppose for sub-group j we have n_j lives with total time observed $E_j^c = \sum_{i=1}^{n_j} t_i$ and total observed deaths $D_j = \sum_{i=1}^{n_j} d_i$. Assuming a constant mortality hazard rate of μ_j, the contribution to the likelihood, L_j, is:

$$L_j = e^{-E_j^c\mu_j}\mu_j^{D_j}. \quad (33)$$

Equation (33) looks like a Poisson likelihood, and this leads to the common misconception that D_j has a Poisson distribution with mean parameter $E_j^c \mu_j$. This is not true, as demonstrated in detail by Macdonald and Richards [2024, Section 3]. In fact, the random variable is the pair (D_j, E_j^c), which cannot have a Poisson distribution. However, the likelihoods for a Poisson model for D_j and the pair (D_j, E_j^c) are the same up to a multiplicative constant, so the inference will be the same; Macdonald and Richards [2024, Section 3] called the likelihood in equation (33) *pseudo-Poisson*. This means that a pseudo-Poisson grouped-count model like equation (33) can be fitted using the machinery of a Poisson GLM, even though the observed death count D_j is not Poisson.

Grouped-count models lose information due to grouping. Care is required with the assumption of a constant hazard rate, μ_j, which requires that lifetimes are split into small enough age ranges for this approximation to be justified; see Macdonald and Richards [2024, Appendix 1]. Even greater care is required with continuous variables like pension size, where the assumption of a constant effect within a discrete band can lead to systematic over-estimation of mortality of pensioners. Figure 22 shows the deviance residuals by pension size-band for a model with age and sex as covariates. The non-random variation suggests that the wealthiest pensioners have significantly lower mortality. If we split the lives into twenty equal-sized groups, Figure 22(a) suggests that the top 10% of amounts are different from the rest. However, Figure 22(c) shows that the assumption of a constant mortality effect within this group is false — there is continuous variation by pension amount. This is important for actuarial purposes, because the assumption of a constant effect risks financial bias — all other factors held equal, the financial impact of those on the right always outweighs the financial impact of those on the left.

Figure 22: Deviance residuals by pension size-band, with 1 representing the smallest pensions. Source: Mortality model for age and sex for English pension scheme from Richards [2022a].



One counter-argument sometimes advanced in favour of discretization is that a model could be fitted using the 100 size-bands in Figure 22(b), say with some penalty function to dampen the inevitable volatility in parameter estimates. However, this would be an additional complexity to deal with the problem caused by discretization. It is far simpler to instead treat variables like pension size continuously and not have the problem arise in the first place. Richards [2022a, Tables 7 and 8] showed that treating pension amount as a continuous risk factor for mortality produced a model with the fewest parameters and the lowest-equal information criterion. This is an extension of graduation by mathematical formula, where a smooth progression in mortality rates by age is guaranteed during the model fit [Forfar et al., 1988]. We go a step further and treat all continuous variables this way, not just age.

E.3 Stratification

Irrespective of whether mortality modelling is of individual lifetimes or groups, there is the question of how to handle sub-groups with shared characteristics. One approach is to sub-divide the experience data and fit separate models, i.e. *stratification* of the data set. In general stratification should be avoided wherever possible, as it increases the overall number of parameters required and weakens the statistical power of the data set [Macdonald et al., 2018, Section 7.3]. However, stratification can on rare occasions be justified where the mortality hazard rate has fundamentally different properties, such as where sub-groups have a different shape of mortality curve by age.

E.4 Open-source software for individual lifetimes

The R [R Core Team, 2021] `survival` package contains the `survreg()` function for fitting parametric survival models. However, most of the models available do not handle left-truncation. Macdonald et al. [2018, Chapter 5] provide R scripts for fitting models to left-truncated data, with discussion on use of the R's `nlm()` and `solve()` functions. Commercially supported services are also available.

ON CONTEMPORARY MORTALITY MODELS FOR ACTUARIAL USE II: PRINCIPLES

By Angus S. Macdonald†and Stephen J. Richards‡

ABSTRACT

We reprise some common statistical models for actuarial mortality analysis using grouped counts. We then discuss the benefits of building mortality models from the most elemental items. This has two facets. First, models are better based on the mortality of individuals, rather than groups. Second, models are better defined in continuous time, rather than over fixed intervals like a year. We show how survival probabilities at the 'macro' level arise at the 'micro' level from a series of Bernoulli trials over infinitesimally small time periods. Using a multiplicative representation of the mortality hazard rate, we show how counting processes naturally represent left-truncated and right-censored actuarial data, individual or age-grouped. Together these explain the 'pseudo-Poisson' behaviour of survival model likelihoods.

KEYWORDS

Bernoulli Trial, Counting Process, Likelihood, Product-integral, Pseudo-Poisson Model

CONTACT ADDRESSES

† Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK, and the Maxwell Institute for Mathematical Sciences, UK.
‡ Longevitas Ltd., 24a Ainslie Place, Edinburgh EH3 6AJ
Corresponding author: Angus Macdonald, A.S.Macdonald@hw.ac.uk.

> *If I dared, I would say we must have a theory — the word 'theory' is so much disliked by so many Englishmen, and is considered by them so 'unpractical', that I avoid it all I can; though I cannot see, myself, that it is very 'practical' to do things without knowing the theory of how to do them.*
>
> *Wintringham & Blashford-Snell (1973)*

1. Introduction

1.1 *In Search of a Continuous-time Model of Mortality*

Richards & Macdonald (2024) set out some practical benefits of using 'continuous-time' models of mortality. This expository paper asks what we mean by a 'continuous-time' model of mortality. As we seek an answer, in the theoretical basis of actuarial mortality modelling, we provide the language and notation to keep actuaries abreast of some fairly recent developments. The purpose of the paper is not to provide novel results, but to demonstrate how an actuary can see familiar objects in novel ways.

In fact, the idea of a 'continuous-time' model of mortality is not clear-cut or self-contained, and it leads us to consider two contrasts, which we may think of as modelling phenomena on a 'micro' scale versus phenomena on a 'macro' scale. These are:

(a) very informally, the choice of infinitesimal time unit dx versus a discrete time unit, which we take to be a year; and
(b) models based on individual lives versus models based on collectives of lives, including, inter alia, the collection of data based on individual lives versus collection of age-grouped data.

1.2 *Inspiration from the Life Table*

The life table is an obvious source of inspiration. Indeed, in the past some have viewed the whole subject as being the construction of life tables, see for example Batten (1978). A life table is a model of a cohort of identical and independent individuals, followed from some initial selection event at integer age $x_0 \geq 0$ (such as birth, with $x_0 = 0$) until mortality has extinguished the whole cohort, typically represented by the function l_x, interpreted as the expected number left alive at integer ages $x \geq x_0$. The two key features are:
(a) the focus on the collective rather than the individual; and
(b) the time unit of a year;

which are both 'macro' properties. If we model the number of deaths between integer ages x and $x+1$ as a random variable D_x, then this formulation of the life table immediately suggests the binomial distribution as a model for D_x, see Section 2.3.

A slightly different view of the life table inspires a different model. Allow l_x to range over all real $x \geq x_0$, not just integer ages, and interpret the ratio l_{x+n}/l_x as the probability that an individual alive at age x survives to age $x+n$ ($x \geq x_0, n \geq 0$). This leads to further ideas, namely:
(a) a model in which death is possible at any moment of time; and
(b) the force of mortality or hazard rate (our preferred term) μ_x at age x as a measure of the instantaneous risk of death;

which are both 'micro' properties. However, observation is still 'macro', of the collective rather than of the individual. This setup suggests a Poisson model for D_x, see Section 2.4.

1.3 *Models Based on Individuals: The Pseudo-Poisson Model*

More recent introductions to the subject begin with the definition of the future lifetime of a person age x as a non-negative random variable, denoted by T_x. For brevity and completeness, we compress the definitions of related quantities into Table 1, see Dickson et al. (2020) or Macdonald et al. (2018) for details. Of course, the actuarial symbols ${}_tq_x, {}_tp_x$ and μ_x would be defined, based on the life table, in the process of obtaining the binomial and Poisson models in Section 1.2, but the point is that they are now defined by their rôles in the distribution of T_x.

If we define T_0^i to be the random lifetime of the ith individual under observation, this model focusses attention on:
(a) the individual rather than the collective; and
(b) events happening instantaneously, meaning during a short time period h as we let $h \to 0^+$;

Table 1: Definitions of quantities based on T_x, the random future lifetime at age x. The consistency condition assumes that $x_0 = 0$, and ensures that calculations based on the distribution of T_x will never contradict calculations based on the distribution of T_y ($y \neq x$).

Quantity	Definition
Distribution function:	$\mathrm{P}[T_x \leq t] = {}_tq_x$
Survival function:	$\mathrm{P}[T_x > t] = {}_tp_x$
Hazard rate :	$\lim_{h \to 0^+} \frac{{}_hq_x}{h} = \mu_x$
Density function:	$\frac{d}{dx}\,{}_tq_x = {}_tp_x\,\mu_{x+t}$
Consistency condition:	$\mathrm{P}[T_x \leq t] = \mathrm{P}[T_0 \leq x + t \mid T_0 > x]$

which are both 'micro' properties. The most important idea is expressed in the heuristic:

$$\mathrm{P}[\text{Dead by age } x+h \mid \text{Alive at age } x] = {}_hq_x \approx h\,\mu_x \quad (\text{for small } h). \tag{1}$$

Knowing the density function of each T_x^i (Table 1), we can write down the probability of any observations, hence a likelihood, and that leads to the following explanation of why the Poisson model of Section 2.4 works so well. Suppose we assume a constant hazard rate at each age, we observe M individuals and there are D deaths. Then:

(a) the model based on individual random lifetimes gives us, in principle, an exact probability of observing D deaths; while

(b) the Poisson model of Section 2.4 gives us only an approximate probability of observing D deaths (it must do since $D \leq M$ but $\mathrm{P}[D > M] > 0$ under a Poisson model); however

(c) both models give us exactly the same likelihood of observing D deaths (up to irrelevant factors).

It follows that inference based on the likelihood will be identical under both models. This leads us to call the model based on individual lifetimes (and a constant hazard) the pseudo-Poisson model (Section 3.7). We continue to seek the proper foundations of a mortality model at the 'micro' level in the model of individual lifetimes.

1.4 *Dynamic Life History Models I: Truncation and Censoring*

The individual life-history model lets us write down exact probabilities of observed events, if we know the hazard rates. It also lets us deal with incomplete observation, in particular:

(a) left-truncation: an individual enters observation having already survived to some age $x_a > 0$; and

(b) right-censoring: the individual leaves observation while still alive, so we observe only that $T_{x_a} > x_b$ for some age $x_b > x_a$.

A neat device allows us to avoid the complication of keeping track of ages x_a and x_b when writing expressions such as likelihoods. Define a process:

$$Y^i(x) = I_{\{i\text{th individual alive and under observation at age } x^-\}} \tag{2}$$

(age x^- means 'just before age x' and is a technicality). The 'under observation' condition takes care of left-truncation and right-censoring. Then, for example, the integrated hazard rate over the time spent under observation by the ith individual (an important quantity in many calculations), can be written:

$$\int_{x_a}^{x_b} \mu_x \, dx = \int_0^\infty Y^i(x) \, \mu_x \, dx. \tag{3}$$

We see that the process $Y^i(x)\,\mu_x$ acts as a dynamic or stochastic hazard rate tailored to the ith individual, and greatly simplifies expressions involving integrals, since all integrals can now be now taken over $(0, \infty]$ (Section 4.6).

1.5 *Dynamic Life History Models II: Back to Bernoulli*

In a utilitarian sense the job was finished with Section 1.3, but the heuristic (1) suggests more to come. For, if ${}_hq_x \approx h\,\mu_x$ then ${}_hp_x \approx 1 - h\,\mu_x$, and if we let δ_x be an indicator, equal to 1 if death occurs at age x, and 0 otherwise, then what is observed 'during' time h is the outcome of a Bernoulli trial with parameter $h\,\mu_x$ and probability:

$$(1 - h\,\mu_x)^{1-\delta_x} \, (h\,\mu_x)^{\delta_x}. \tag{4}$$

We would like to take all such consecutive 'instantaneous' Bernoulli trials while the individual is alive and under observation, and multiply their probabilities (4) together. In all of probability theory, there is nothing simpler than a Bernoulli trial, so we really would have reduced a probability in a mortality model to its constituent 'atoms'; the ultimate 'micro' level. That is what we describe in Section 4. To do so we introduce two ideas, which give us the notation needed to write down a product of Bernoulli probabilities like (4) in a rigorous way.

(a) Counting processes: A counting process $N^i(x)$ starts at $N^i(0) = 0$ and jumps to 1 at time T_0^i if the ith individual is then under observation. Then its increment $dN^i(x)$ indicates an observed death, and is a rigorous version of the informal δ_x in (4). Between them, $N^i(x)$ and $Y^i(x)$ let us write the Bernoulli trial probability (4) formally as:

$$(1 - Y^i(x)\,\mu_x\,dx)^{1-dN^i(x)} \, (Y^i(x)\,\mu_x\,dx)^{dN^i(x)} \tag{5}$$

and this allows for left-truncation and right-censoring.

(b) Product-integral: The product-integral is the device that lets us multiply all the infinitesimal Bernoulli trial probabilities. We defer further description to Section 4.2 and Appendix 2 and just give the final form of the likelihood contributed by the ith individual, denoted by L_i:

$$L_i = \prod_{x \in (0,\infty]} (1 - Y^i(x)\,\mu_x\,dx)^{1-dN^i(x)} \, (Y^i(x)\,\mu_x\,dx)^{dN^i(x)}. \tag{6}$$

The product-integral is identified by a product over all values of an interval ($x \in (0, \infty]$ here) and the presence of the variable of integration (dx here) in the integrand.

Therefore, we have a mortality model with the following properties:

(a) It is irreducible, in the sense that it is composed of consecutive (infinitesimal) Bernoulli trials.
(b) It is based on behaviour at the 'micro' time scale.
(c) It is based on individual lives.
(d) Aggregated, over time and over individuals, it explains the Poisson-like nature of likelihoods, therefore estimation based on the collective at the 'macro' time scale.
(e) It allows for left-truncation and right-censoring.
(f) It is easily extended to multiple-decrement and multiple-state models.

We started out by trying to pin down what we meant by the vague term 'continuous-time mortality model'. Now we have an answer, although our endpoint is just the starting point for the modern statistical study of survival models (Section 4.10), see Andersen et al. (1993).

1.6 *Plan of this Paper*

We start in Section 2 with Forfar et al. (1988), a definitive account of graduation using binomial and Poisson models, which we call mortality models at the 'macro' scale. Then in Section 3 we turn to models at the 'micro' scale based on individual lifetimes, and find the origins of Poisson-like behaviour at the 'macro' scale. In Section 4 we bring together models of individual lifetimes and models based on behaviour over small intervals h as $h \to 0^+$, and find that all probabilities in a mortality model arise as a product of consecutive (infinitesimal) Bernoulli trials. Section 5 concludes.

2. Binomial and Poisson Models

2.1 *Forfar et al. (1988)*

In a landmark paper, Forfar et al. (1988) gave comprehensive accounts of two models for survival data, namely the binomial and Poisson models. These defined: (a) the random variable D_x, to be the number of deaths observed at age x; and (b) a suitable measure of exposure to risk at age x, assumed to be non-random, that we will call V_x. Then the occurrence-exposure rate D_x/V_x was shown to be an estimate of the model parameter, a mortality rate $\hat{q}$ in the binomial model, and a hazard rate $\hat{\mu}$ in the Poisson model[1].

Forfar et al. (1988) helpfully located the old subject of parametric graduation in a modern statistical setting, including model specification, likelihood, score function and information, covariance matrix, model selection and parametric bootstrapping. The treatment was heavily influenced by the authors' work for the Continuous Mortality Investigation Bureau (the CMIB, now CMI) particularly in respect of data collection. The advance it represented may be gauged by comparison with contemporary texts such as Batten (1978) and Benjamin & Pollard (1980).

[1] A third model was offered, identical to the Poisson model except that the occurrence-exposure rate was taken to be an estimate $\hat{m}$ of the life table quantity m_x, the central rate of mortality (see Neill (1977) for example). Since m_x does not arise naturally as the parameter in any well-defined statistical model we will not pursue this further.

Both binomial and Poisson models are rooted in simple thought-experiments, require no statistics beyond a first course in data analysis, and can, with qualifications, be implemented in standard statistical packages such as `R` (R Core Team 2021). This gives them considerable staying power.

2.2 *The Rate Interval and Δ_k Notation*

The rate interval is an interval of age (or calendar time) on which an individual is assigned a given age label. It is the means of assigning an age label to an individual exposed to the risk of death and at the time of death. Note that rate intervals are only needed with age-grouped data, not models based on individual lives. They are treated in detail in texts such as Benjamin & Pollard (1980). We assume that the rate interval, when we need one, is the year of age $(x, x+1]$ defined by 'age last birthday'. The CMI, for another example, use the year of age $(x - 1/2, x + 1/2]$ defined by 'age nearest birthday'.

We assume that the data are covered by K rate intervals and that in the abstract these may be denoted by Δ_k $(k = 1, 2, \ldots, K)$; and that a sum over all rate intervals may be denoted by $\sum_k$, and a product likewise by $\prod_k$.

2.3 *Binomial Models*

The binomial model is based on the following thought-experiment: take E_x lives at the start of a year, all alive at age x and assumed to be 'statistically independent' in respect of their mortality risk. Then E_x is the measure of exposure referred to as V_x in Section 2.1, usually here called the initial exposed-to-risk. Define D_x to be the number who are dead at the end of the year, and q_x to be the probability that a life alive at age x dies not later than age $x + 1$. Then the following are easily shown.

(a) D_x has a binomial(E_x, q_x) distribution, with first two moments $\mathrm{E}[D_x] = E_x\, q_x$ and $\mathrm{Var}[D_x] = E_x\, q_x\, (1 - q_x)$.

(b) As a function of parameter q_x, the data (D_x, E_x) has likelihood function:

$$\begin{aligned} L(q_x) &= \binom{E_x}{D_x} q_x^{D_x}\, (1 - q_x)^{E_x - D_x} \\ &\propto q_x^{D_x}\, (1 - q_x)^{E_x - D_x}, \end{aligned} \tag{7}$$

leading to the maximum likelihood estimate (MLE) $\hat{q}_x = D_x/E_x$ which is asymptotically unbiased $(\mathrm{E}[\hat{q}_x] = q_x)$ with variance $\mathrm{Var}[\hat{q}_x] = q_x\, (1 - q_x)/E_x$.

(c) The estimate $\hat{q}_x$ is an estimate of q_x, that is, the function value at start of the rate interval $(x, x + 1]$.

D_x can be viewed as the number of successes out of E_x independent Bernoulli trials, each with probability of success (death) equal to q_x. The idea of the Bernoulli trial as the fundamental 'atom' of mortality risk appears again in Section 4.5.

2.4 *Poisson Models*

The Poisson model depends on a different thought-experiment. An unspecified number of individuals is observed, alive during the relevant rate interval of age $(x, x+1]$, such that the total time alive and under observation is a non-random quantity E_x^c. Now E_x^c is

the measure of exposure referred to as V_x in Section 2.1, usually here called the central exposed-to-risk. A constant force of mortality μ is assumed at all ages in the rate interval $(x, x+1]$. Define D_x to be the number of observed deaths. Then the following can be shown.

(a) D_x has a Poisson$(\mu\, E^c_x)$ distribution with first two moments $\mathrm{E}[D_x] = \mathrm{Var}[D_x] = \mu\, E^c_x$.

(b) As a function of parameter μ, the data (D_x, E^c_x) has likelihood function:

$$\begin{aligned} L(\mu) &= (\mu\, E^c_x)^{D_x}\, \exp(-\mu\, E^c_x)/D_x! \\ &\propto \exp(-\mu\, E^c_x)\, \mu^{D_x}, \end{aligned} \tag{8}$$

leading to the MLE $\hat{\mu} = D_x/E^c_x$ which is asymptotically unbiased ($\mathrm{E}[\hat{\mu}] = \mu_x$) with variance $\mathrm{Var}[\hat{\mu}] = \mu/E^c_x$.

(c) Assuming a relatively even distribution of exposure over the rate interval $(x, x+1]$, the MLE $\hat{\mu}$ estimates $\mu_{x+1/2}$.

2.5 *Terminology*

Binomial and Poisson models may be described in different ways. The binomial model admits of no conceivable time other than its own time unit; it is unambiguously a discrete-time model. It may also be called a q-type model in honour of its conventional parameter. The Poisson model is a candidate for a continuous-time model, although it turns out to be an extreme representative of a whole class, see Sections 3 et seq.. It may also be called a μ-type model in honour of its conventional parameter. See Richards & Macdonald (2024) for both terminologies.

2.6 *Assessment of the Binomial and Poisson Models*

2.6.1 *Feasibility of the Thought-Experiment: Binomial Model*

To carry out the binomial thought-experiment we would need a homogeneous sample of E_x individuals age x, observed to be alive or dead at age $x+1$. This contrasts with observation of (say) members of a pension scheme or life office policyholders. Real data often includes exits for reasons other than death and not under the modeller's control, see (Richards & Macdonald 2024, Section 3 and Appendix) for examples. The requirements of the binomial experiment will not be met by: (a) individuals entering observation between ages x and $x+1$; and (b) individuals leaving observation between ages x and $x+1$, for reasons other than death.

Thus we are led to ask, what is the probability of surviving over any fraction of the rate interval? For example, an individual joining at age $x - 1/2$ and surviving to age x requires the calculation of ${}_{1/2}p_{x-1/2}$. The binomial model gives no satisfactory answer. Strictly, the question lies outside the bounds of the model. Even if we could implement the thought-experiment, the model posits only the number of lives observed at the start and end of the rate interval.

Nevertheless, an answer may be demanded, because individuals can and do join or leave an investigation in the middle of the rate interval, see (Richards & Macdonald 2024, Section 3) for numerous examples in practice. The analyst is obliged to make some

assumption about mortality between ages x and $x+1$, for which the binomial model gives no guidance. Three popular assumptions have been:
(a) a uniform distribution of deaths;
(b) the Balducci hypothesis;
(c) a constant hazard rate.

See Macdonald (1996) or Richards & Macdonald (2024) for a discussion of these. Here we just remark that (c), a constant hazard rate, is mathematically the simplest, fully consistent with the Poisson model, and also consistent with modelling individual lifetimes as in Section 3.

2.6.2 *Feasibility of the Thought-Experiment: Poisson Model*

The Poisson thought-experiment is not troubled by fractions of the rate interval. Since the hazard rate is assumed to be a constant μ during the rate interval $(x, x+1]$, the probability of dying during the sub-interval $(x+a, x+b]$ (given alive at age $x+a$, for $0 \le a < b \le 1$) is $1 - \exp(-\mu\,(b-a))$.

The Poisson thought-experiment is not met in practice, however, for different reasons. The distribution of D_x is Poisson only if the exposed-to-risk E_x^c is non-random, for example pre-determined. This is not the case if: (a) the population being sampled is finite, with known maximum size M individuals, say, because then $D_x \le M$, but $\mathrm{P}[D_x > M] > 0$ under any Poisson distribution; or (b) the exposure times of the individuals in the sampled population are not known in advance, because then E_x^c is random. Moreover, D_x is usually a component of the bivariate random variable (D_x, E_x^c). In such cases we call D_x pseudo-Poisson, see Section 3. For estimation purposes, however, it behaves as a true Poisson random variable would, see Section 3.7.

2.6.3 *Occurrence-exposure Rates, Age-grouped Data and Graduation*

The estimates $\hat{q}_x = D_x/E_x$ and $\hat{\mu}_x = D_x/E_x^c$ are examples of occurrence-exposure rates. Both they and their sample variances (Sections 2.3 and 2.4) require only the age-grouped totals D_x and E_x or E_x^c to be reported to the analyst, rather than data on each individual. Such totals may easily be extracted from ordinary data files used in the business; they greatly reduce the volume of data required (which used to matter a lot); and they reduce the risk of accidentally breaching data-protection rules (which matters now). On the other hand they do not allow the level of checking and cleaning of the data that is possible with individual data (Macdonald et al. 2018, Chapter 2).

If age-grouped data are prepared by someone other than the analyst, the modelling is wholly dependent on the thoroughness and diligence of the source provider. This is a material concern for risk-transfer transactions, such as reinsurance, bulk annuities and portfolio transfers. If a model is to be used to price a risk transfer, the analyst should always insist on individual records, regardless of whether the intent is to use models based on individuals or age-grouped counts.

Occurrence-exposure rates $\hat{q}_x$ or $\hat{\mu}_x$ are normally smoothed or graduated for practical use. For this purpose a likelihood may be calculated as the product of the likelihoods for each rate interval, using age-grouped data. Other, non-likelihood methods may also be used (Forfar et al. 1988).

2.7 *Generalized Linear Models (GLMs)*

To the list of properties in Sections 2.3 and 2.4 we could have added "(d) Leads to a simple Generalized Linear Model (GLM) for graduating age-grouped mortality data."

GLMs were introduced by Nelder & Wedderburn (1972), and contain three elements: (a) a random component, Y_x; (b) a systematic component, η_x; and (c) a link function, g. A GLM connects the expectation of Y_x to η_x via g as follows:

$$\eta_x = g(\mathrm{E}[Y_x]). \tag{9}$$

The component η_x is the linear predictor; in mortality work it is a linear function of age, x, and a corresponding covariate vector, $\boldsymbol{z}_x$. Let $\boldsymbol{\theta}$ be the vector of parameters to estimate, and let $\boldsymbol{X}$ be the corresponding model matrix. Each observation Y_x has a corresponding row in $\boldsymbol{X}$. For a binomial GLM we have:

$$Y_x = \frac{D_x}{E_x}, \qquad \eta_x = \boldsymbol{X}\boldsymbol{\theta}[x,\,]. \tag{10}$$

For a Poisson GLM with the link function $g(x) = \log(x)$ we have:

$$Y_x = D_x, \qquad \eta_x = \boldsymbol{X}\boldsymbol{\theta}[x,\,] + \log(E^c_x). \tag{11}$$

where $[x,\,]$ selects the row for the observation corresponding to age x.

The link function, g, is chosen by the analyst. The canonical link for the binomial GLM is the logit, but other link functions can be used, such as the probit link. The canonical link for the Poisson GLM is the logarithm, but other link functions have been used for mortality work, such as the logit link; see (Currie 2016, Appendix 1) for implementation details of the logit link for Poisson GLMs.

GLMs have a link with 'classical' actuarial modelling since one of the simplest choices of fitted $\hat{\eta}_x$ is a Gompertz function, but this does not extend to other members of the Gompertz-Makeham family (see Forfar et al. (1988)).

GLM's are popular because they are flexible, have nice statistical properties and are linear in the covariates. The binomial and Poisson error structures arise naturally for 'count' data, and age-grouped deaths are examples of 'counts' so these GLMs are, in a sense, natural candidates as mortality models. However, linear dependence on covariates, and the canonical link functions associated with the exponential family, are restrictive, and for large experiences we will often find better-fitting models that are not GLMs (see, for example, the range of models included in Cairns et al. (2009)). In addition GLMs bring us no closer to any foundational concept of a 'mechanism' generating mortality data, so we do not consider them further.

3. Modelling Individual Lifetimes: The Pseudo-Poisson Model

3.1 *Observation of an Individual*

Suppose the ith individual is observed from age x_i until age y_i for total time $v_i = y_i - x_i$. Denote the interval $(x_i, y_i]$ by Δ_i. Observation ends because of either death or right-censoring at age y_i. Define the indicator:

$$d_i = \begin{cases} 1 & \text{if } i\text{th individual died at age } y_i \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Then the random variable observed is the bivariate (d_i, v_i), and the total contribution to the likelihood of these observations, denoted by L_i, is:

$$\begin{aligned} L_i &= {}_{v_i}p_{x_i}\,\mu_{x_i+v_i}^{d_i} \\ &= \exp\left(-\int_0^{v_i} \mu_{x_i+s}\,ds\right)\mu_{x_i+v_i}^{d_i} \\ &= \exp\left(-\int_{\Delta_i} \mu_s\,ds\right)\mu_{y_i}^{d_i} \end{aligned} \tag{13}$$

see Table 1, or (Macdonald et al. 2018, Chapter 5).

3.2 *Age Intervals and Δ_i Notation*

The definition of the interval Δ_i depends on the observational plan and the method of investigation. The important point is that it is time under observation of a single individual, the ith of M individuals. Some examples are the following.

(a) The interval may be the entire period for which the ith individual was observed, potentially spanning many years.
(b) The interval may be that part of a rate interval (for example, year of age) for which the ith individual was under observation.
(c) The interval may be an interval of age on which the hazard rate is assumed to be constant.

Therefore the likelihood (13) based on Δ_i may constitute the whole of the ith individual's contribution to the total likelihood, or only part of it. We will call a contribution to a likelihood of the form (13) a survival model likelihood, whether it forms all or part of the ith individual's contribution, and whether or not hazard rates are assumed to be piecewise-constant.

3.3 *Multiplication and Factorization of Survival Model Likelihoods*

Survival model likelihoods in respect of the same individual over contiguous intervals, multiplied together, give another survival model likelihood. In reverse, a survival model likelihood may be factorized into as many factors of like kind as we please. To see the multiplicative property, suppose the ith individual is observed on the contiguous intervals $\Delta_i^1 = (x_i, z_i]$ and $\Delta_i^2 = (z_i, y_i]$, and that indicators of death $d_i^{(1)}$ at time z_i and $d_i^{(2)}$ at time y_i are defined analagously to d_i above. Then if $\Delta_i = (x_i, y_i]$ as before:

$$\begin{aligned} \exp\left(-\int_{\Delta_i} \mu_s\,ds\right)\mu_{y_i}^{d_i} &= \exp\left(-\int_{\Delta_i^1\cup\Delta_i^2} \mu_s\,ds\right)\mu_{y_i}^{d_i} \\ &= \exp\left(-\int_{\Delta_i^1} \mu_s\,ds\right)\mu_{z_i}^{d_i^{(1)}}\,\exp\left(-\int_{\Delta_i^2} \mu_s\,ds\right)\mu_{y_i}^{d_i^{(2)}} \end{aligned} \tag{14}$$

since necessarily $d_i^{(1)} = 0$ and $d_i^{(2)} = d_i$ if events on Δ_2 are not trivially null. Whether we regard this as factorizing a likelihood on Δ_i, or multiplying two likelihoods on Δ_i^1 and Δ_i^2, does not matter for our purposes.

3.4 *Rate Intervals, Piecewise-constant Hazards and Age-grouped Data*

Recall from Section 2.2 that a rate interval is denoted by Δ_k. Here let Δ_k be the rate interval from integer age k to age $k+1$, that is, $\Delta_k = (k, k+1]$. Age-grouped data may then be denoted by total deaths d_k and total person-years exposure E_k^c falling within rate interval Δ_k.

It is instructive to group data on individual lives to reproduce age-grouped data, and to compare the resulting likelihoods. This is aided by Table 2, which shows the contributions to likelihoods of individual data for two individuals, treated three ways. The ith individual is observed from age 47 until right-censored at age 50. The jth individual is observed from age 47.6 until dying at age 49.3. We list contributions to the likelihood under three combinations of observational plan and model:

(a) Rate intervals Δ_k, and a constant hazard rate on each rate interval, denoted by μ_k^*.
(b) Rate intervals Δ_k, and a smooth hazard rate parametrized by θ, denoted by μ_x^θ (for example, a Gompertz-Makeham function).
(c) Observation of complete lifetimes on age interval Δ_i, and a smooth hazard rate parametrized by θ, also denoted by μ_x^θ.

The contributions are shown in Table 2. In obvious notation, we may denote the contributions to the likelihoods under (a), (b) and (c) above by $L_{i,k}^*, L_{i,k}^\theta$ and $L_i^\theta = \prod_k L_{i,k}^\theta$ respectively. Likewise, collecting all contributions to rate interval Δ_k in columns (a) and (b), we may define the total likelihood contributed by Δ_k by $L_k^* = \prod_i L_{i,k}^*$ and $L_k = \prod_i L_{i,k}$ respectively. This leads to the following observations.

(a) It is obvious from columns (b) and (c) that for any individual, the likelihood over the complete lifetime is the product of the likelihoods over each rate interval, see Section 3.3. In fact we have incorporated this in the notation, $L_i^\theta = \prod_k L_{i,k}^\theta$. It makes no difference if we split the the individual lives data and present them by rate interval. But this does not lead to any simplification, and age-grouped totals d_k and E_k^c play no part, because of the smooth hazard rate in the integrands.
(b) Each entry in column (a) can be regarded as approximating its partner in column (b). For example in the third and sixth lines, we approximate μ_{49+s}^θ by μ_{49}^* for $0 \le s < 1$; we show the sixth line below:

$$\exp\left(-\int_{0.0}^{0.3} \mu_{49+s}^\theta \, ds\right) \mu_{49.3}^\theta \approx \exp\left(-\int_{0.0}^{0.3} \mu_{49}^* \, ds\right) \mu_{49}^* = \exp\left(-0.3\mu_{49}^*\right) \mu_{49}^*. \quad (15)$$

Collecting together all such terms in μ_{49}^* we get the total likelihood:

$$L_{49}^*(\mu_{49}^*) = \exp\left(-\mu_{49}^* \, E_{49}^c\right) (\mu_{49}^*)^{d_{49}} \quad (16)$$

Table 2: Contributions to likelihoods of the ith individual, under observation from age 47 until right-censored at age 50, and the jth individual, under observation from age 47.6 until death at age 49.3, under three observational plans and assumptions: (a) annual rate interval Δ_k, piecewise-constant hazard rates; (b) annual rate interval Δ_k, smooth hazard rate parametrized by θ; and (c) observation of complete lifetime age interval Δ_i, smooth hazard rate parametrized by θ.

		Observational Plan and Likelihood:		
Individual	Rate Interval	Rate Interval Δ_k Likelihood $L_k^*(\mu_k)$ (a)	Rate Interval Δ_k Likelihood $L_{i,k}^\theta$ (b)	Complete Lifetime Δ_i Likelihood $L_i^\theta = \prod_k L_{i,k}^\theta$ (c)
i	Δ_{47}	$\exp\left(-1.0\,\mu_{47}^*\right)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{47+s}^\theta\, ds\right)$	$\exp\left(-\int_{47.0}^{50.0} \mu_s^\theta\, ds\right)$.
	Δ_{48}	$\exp\left(-1.0\,\mu_{48}^*\right)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{48+s}^\theta\, ds\right)$	
	Δ_{49}	$\exp\left(-1.0\,\mu_{49}^*\right)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{49+s}^\theta\, ds\right)$	
j	Δ_{47}	$\exp\left(-0.4\,\mu_{47}^*\right)$	$\exp\left(-\int_{0.6}^{1.0} \mu_{47+s}^\theta\, ds\right)$	$\exp\left(-\int_{47.6}^{49.3} \mu_s^\theta\, ds\right) \mu_{49.3}^\theta$.
	Δ_{48}	$\exp\left(-1.0\,\mu_{48}^*\right)$	$\exp\left(-\int_{0.0}^{1.0} \mu_{48+s}^\theta\, ds\right)$	
	Δ_{49}	$\exp\left(-0.3\,\mu_{49}^*\right) \mu_{49}^*$	$\exp\left(-\int_{0.0}^{0.3} \mu_{49+s}^\theta\, ds\right) \mu_{49.3}^\theta$	

in which the age-grouped totals do appear. Comparing equations (16) and (8), we see that the former is functionally identical to the likelihood from the Poisson model, and yet no assumption about Poisson random variables or distributions has been made in this section. In other words, the Poisson-like nature of the likelihood arises from the fundamental nature of modelling individual lifetimes.

3.5 *Individual versus Age-Grouped Data for Multiple Lives*

Table 2 illustrates how the individual lifetime model is related to age-grouped data, based on rate intervals, exactly without using any approximations. Indeed, columns (b) and (c) show that labelling the data by individual i or by rate interval k is merely a rearrangement. Specifically, the ith individual contributes $L^{\theta}_{i,k}$, possibly null, to the likelihood in rate interval k (fourth column). The outer form of the total likelihood, denoted by L^{θ}, then depends simply on the order in which we take products, as the following identities show:

$$L^{\theta} = \prod_i L^{\theta}_i = \prod_i \prod_k L^{\theta}_{i,k} = \prod_k \prod_i L^{\theta}_{i,k} = \prod_k L^{\theta}_k = L^{\theta}. \tag{17}$$

This informal statement based on Table 2 ('proof-by-example') of course needs to be demonstrated properly. Doing so with the notation to hand is surprisingly detailed, though elementary, and is delegated to Appendix 1. A much simpler proof will be shown when the notation of Section 4 is available (Section 4.7).

3.6 *The Rôle of Occurrence-exposure Rates*

We may arrive at the likelihood based on the age-grouped data (d_k, E^c_k) in two different ways.

(a) We could use the Poisson model with parameter $\mu^*_k\, E^c_k$ (Section 2.4) for rate interval Δ_k.

(b) Within the individual lives model, we could assume that the hazard rate is piecewise-constant with value μ^*_k on rate interval Δ_k. This means assuming that the parameter θ is the vector of hazard rates μ^*_k.

In either case, on rate interval Δ_k, we have a single parameter, which we denote by μ^*_k, and a likelihood that we denote by $L^*_k(\mu^*_k) = \exp(-\mu^*_k\, E^c_k)\,(\mu^*_k)^{d_k}$. In total we have a K-parameter model with likelihood $\prod_k L^*_k(\mu^*_k)$, from which the parameters are estimated independently by the occurrence-exposure rates d_k/E^c_k, which we denote by $\hat{\mu}^*_k$. That is as far as the probabilistic model takes us.

In traditional actuarial terminology, the $\hat{\mu}^*_k$ are 'crude' rates which require to be smoothed or graduated, using no more than the available age-grouped data (Benjamin & Pollard 1980). A convenient way of doing so is to use: (a) the likelihood function $\prod_k L^*_k(\mu^*_k)$; (b) a parametric function μ^{θ}_x for the hazard rate, of much lower dimension than K; and (c) to connect the two with an assumption that $\hat{\mu}^*_k$ estimates $\mu^{\theta}_{x_k}$, for some $x_k \in \Delta_k$, for example $\hat{\mu}^*_k$ estimates $\mu^{\theta}_{k+1/2}$. Note that this smoothing procedure is not part of the probabilistic model, despite its use of the likelihood function. Forfar et al. (1988) show that it is approximately equivalent to the much older minimum-χ^2 method.

In either case, again, the age-grouped quantities approximate exact quantities as follows:

$$(\mu_k^*)^{d_k} \approx \prod_{d_{i,k}=1} \mu_{y_i}^{\theta} \qquad \text{and} \qquad \mu_k^* \, E_k^c \approx \sum_i \int_{\Delta_{i,k}} \mu_s^{\theta} \, ds \tag{18}$$

where $d_{i,k}$ is the number of deaths (0 or 1) befalling the ith individual in rate interval Δ_k, and $\Delta_{i,k} = \Delta_i \cap \Delta_k$ (possibly $\emptyset$). Therefore, inference based upon age-grouped data is close, but not identical, to inference based upon the individual lives data.

The crude hazard rates $\hat{\mu}_k^*$, or more accurately the expected deaths based upon them, $\hat{\mu}_k^* \, E_k^c$, may be used in forming statistics such as deviances, used in testing the fit of a graduation (Forfar et al. 1988, Benjamin & Pollard 1980, Macdonald et al. 2018).

3.7 *Pseudo-Poisson Models*

The various likelihoods that appear in this section, see Table 2, are all Poisson-like, and if a piecewise-constant hazard rate is assumed, indistinguishable from a true Poisson likelihood, an observation that goes back to the earliest work on inference in Markov models, see for example Sverdrup (1965), Waters (1984). However, there are no Poisson random variables. In the likelihood (13): (a) d_i is either 0 or 1, and does not range over the non-negative integers; (b) v_i is random, not deterministic; and (c) d_i and v_i are not independent, the random variable is the bivariate (d_i, v_i).

Many authors suppose, as we did in Section 2.4, that the number of deaths in some model has a Poisson distribution, but without ensuring, as we did, that the exposure times would be non-random; more often the observation of random death times ensures the opposite. This conceptual error is almost always immaterial for inference, precisely because the Poisson likelihood is of the correct form for a survival model, although the survival model is not Poisson. Where it matters is in misdirecting us when we come to extend the survival model, including allowing for: (a) truncation and censoring; (b) more complicated life histories, including multiple decrements; (c) calculating residuals when the expected number of deaths is small; and (d) statistics for multiple lives, see Section 3.5 and Appendix 1.

We suggest it would be clearer and less confusing if the term pseudo-Poisson was adopted, to describe the great majority of models for death counts that appear in the literature.

3.8 *Covariates*

Covariates may be introduced by defining a vector $\mathbf{z}^i$ of covariates for the ith individual and letting the hazard rate be a function $\mu(x, \boldsymbol{z}^i)$ of age and covariates. A common way to introduce such a dependency is to define a vector $\boldsymbol{\beta}$ of regression coefficients such that the hazard rate is a function $\mu(x, \boldsymbol{\beta}^T \boldsymbol{z}^i)$ of age and a linear combination of the covariates. Further simplification is achieved if the hazard rate factorizes as $\mu(x, \boldsymbol{z}^i) = \mu_x \times g(\boldsymbol{z}^i)$, the product of an age dependent hazard rate μ_x (called the baseline hazard) and some function g of the covariates; then the hazard rates of any two individuals of the same age are always in the same proportion, called proportional hazards. Finally, the most common choice of g is an exponential function of a linear combination of the covariates, $\mu(x, \boldsymbol{z}^i) = \mu_x \times \exp(\boldsymbol{\beta}^T \boldsymbol{z}^i)$, which has proportional and non-negative hazards as well as a

Table 3: Three stages in adding structure to a hazard rate that is a function $\mu(x, \boldsymbol{z}^i)$ of age x and a vector $\boldsymbol{z}^i$ of covariates for the ith individual. Each stage is increasingly restrictive, from the most flexible model in Stage 0 to the most restrictive in Stage 3.

Stage	Form of $\mu(x, \boldsymbol{z}^i)$	Description
0	$\mu(x, \boldsymbol{z}^i)$	General function of x and $\boldsymbol{z}^i$
1	$\mu(x, \boldsymbol{\beta}^T \boldsymbol{z}^i)$	Function of age and linear combination of covariates
2	$\mu_x \times g(\boldsymbol{z}^i)$	Proportional hazards
3	$\mu_x \times \exp(\boldsymbol{\beta}^T \boldsymbol{z}^i)$	Basis of the Cox model (Cox 1972)

log-linear dependence on covariates. These steps in adding structure to the hazard rate are summarized in Table 3.

The last hazard structure in Table 3 is popular in medical statistics, where it is known as the Cox model, because the baseline hazard can be ignored and only the regression coefficients need to be estimated, by the procedure known as partial likelihood (see Andersen et al. (1993)). However, actuaries usually wish to estimate the whole model, baseline hazard included, whatever the form of the hazard rate. Then the full likelihood (40) from Appendix 1 becomes:

$$L = \prod_{i=1}^{M} L_i = \prod_{i=1}^{M} \prod_{k=1}^{K} \prod_{\Delta_j \subseteq \Delta_k} \left[\exp\left(- \int_{\Delta_j} \mu(s, \boldsymbol{z}^i)\, ds \right) \right]^{e_{i,j}} \mu(w_j, \boldsymbol{z}^i)^{d_{i,j}}. \tag{19}$$

Clearly any of the hazard rates in Table 3 may be substituted into the likelihood (19). However inspection of the innermost elements of (19), integrals over intervals $\Delta_j \subseteq \Delta_k$, shows that, even if the hazard rate factorizes as in Stages 2 and 3 of Table 3, these factors cannot be collected together to form likelihoods L_k over rate intervals. See also the written comments by A. D. Wilkie in the discussion of Richards (2008).

4. Dynamic Life History Models

4.1 *The Anatomy of a Survival Probability*

We begin with a closer examination of the multiplicative property of survival probabilities, usually expressed as:

$$ {}_{s+t}p_x = {}_tp_x \; {}_sp_{x+t} = {}_sp_x \; {}_tp_{x+s}. \tag{20}$$

We can apply this repeatedly to factorize ${}_tp_x$, with n a positive integer, as follows:

$$ {}_tp_x = \prod_{k=0}^{n-1} {}_{\frac{t}{n}}p_{x+\frac{kt}{n}}. \tag{21}$$

This motivates the first of two questions: what happens as $n \to \infty$? The second (related) question is: how can we express or represent events in the life history as a function of

passing time? We have a compact notation ($_tp_x$, $_tq_x$ and so on) for the probabilities of events in the life history, but no such notation for the events themselves; generally we must express events somewhat clumsily in words. We consider these questions in turn in the next two sections.

4.2 *The Product-integral Representation of a Survival Probability*

From the heuristic $_hp_x \approx 1 - \mu_x\, h \approx \exp(-\mu_x\, h)$, for small h, we have the important product-integral representation as $n \to \infty$ and $1/n \to 0^+$:

$$\begin{aligned} \exp\left(-\int_0^t \mu_{x+s}\, ds\right) &= {}_tp_x && (22)\\ &= \lim_{n\to\infty} \prod_{k=0}^{n-1} {}_{\frac{t}{n}}p_{x+\frac{kt}{n}} && (23)\\ &\equiv \prod_{s\in(0,t]} (1 - \mu_{x+s}\, ds), && (24) \end{aligned}$$

see Appendix 2 or, for example, Andersen et al. (1993). The product-integral has the same Π symbol as an ordinary product over a finite or countable number of terms, but is distinguished (here) by the presence of ds in the integrand and by the variable s ranging over an interval of the real line, $s \in (0, t]$. Then by differentiation of $_tp_x$, the density function of the random future lifetime T_x, denoted by $f_x(t)$, is:

$$\begin{aligned} f_x(t) &= {}_tp_x\, \mu_{x+t} && (25)\\ &= \exp\left(-\int_0^t \mu_{x+s}\, ds\right) \mu_{x+t} && (26)\\ &= \prod_{s\in(0,t]} (1 - \mu_{x+s}\, ds)\, \mu_{x+t}. && (27) \end{aligned}$$

Identities (271) and (27) are important in parametric mortality models, because they allow the likelihood to be specified entirely in terms of the hazard rate.

4.3 *The Counting Process Representation of the Data*

Suppose the ith individual has future lifetime T_0^i, a non-negative random variable. Define the process $N^i(x) = I_{\{T_0^i \le x\}}$. This has value 0 as long as the ith individual is alive, and value 1 if they are dead (including at exact age T_0^i). Thus $N^i(x)$ 'counts' the number of deaths up to and including age x.

Associated with $N^i(x)$ is the indicator of survival, denoted by $Y^i(x)$ and defined as:

$$Y^i(x) = I_{\{i\text{th individual alive at age } x^-\}} = I_{\{T_0^i \ge x\}}. \qquad (28)$$

Thus $Y^i(x)$ is almost, but not quite, equal to $1 - N^i(x)$. Both have value 1 at exact age $x = T_0^i$, the age at death (see Table 4). $N^i(x)$ is right-continuous, while $Y^i(x)$ is left-continuous. This is for technical reasons when forming integrals.

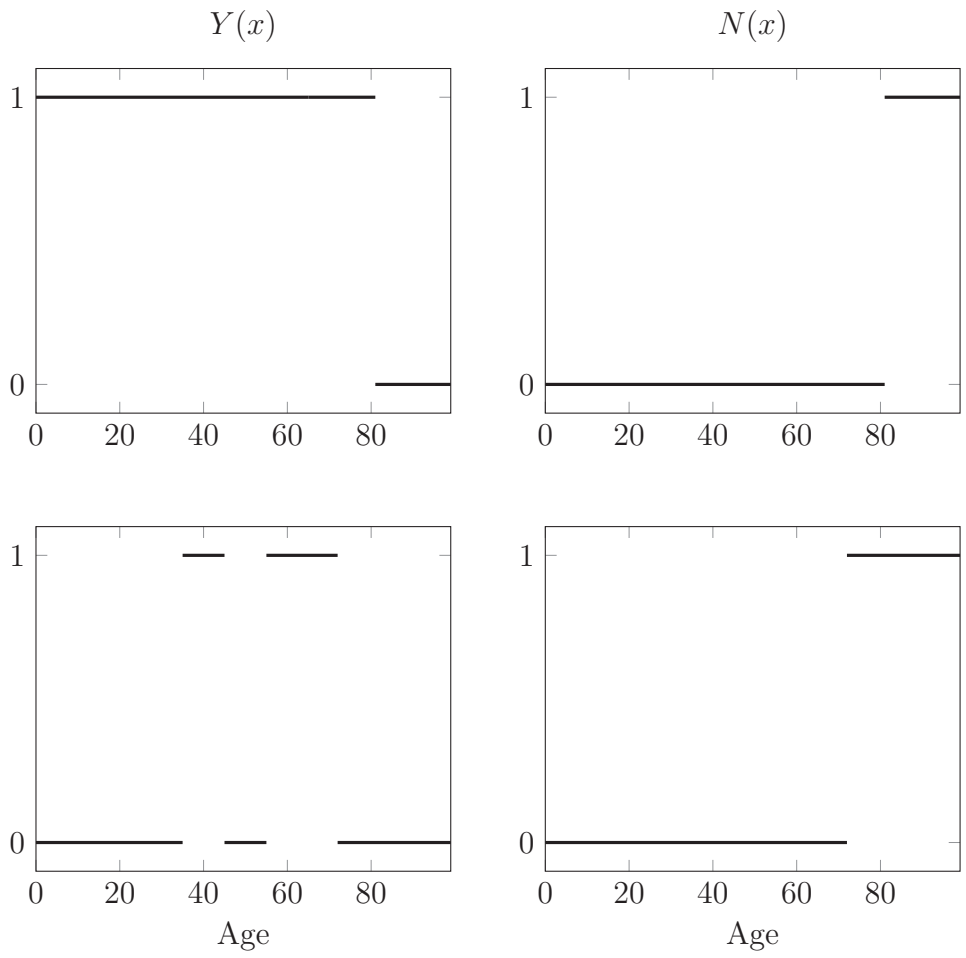


Figure 1: Sample counting-process representations of lifetimes. The first row is a life that enters observation at age 0 and is observed until dying at age 81. The second row is a life that enters observation at age 35 (left-truncation) and leaves at age 45 (right-censoring), enters observation again at age 55 (left-truncation) and is then observed until death at age 72.

Note that $N^i(x)$ represents exactly the same information as T_0^i, but in the form of a stochastic process instead of a random variable. It is purely descriptive, no probabilities or hazard rates appear in its definition, hence we refer to the counting process 'representation', not the counting process 'model'.

Mention the word 'process' to an actuary under the age of about fifty, and it will trigger thoughts of Brownian motion, Itô calculus, stochastic integrals and option pricing. While important and necessary in its place, a counting processes carries none of that baggage[2]. It really is nothing but a parsimonious way to represent an event happening at a random time, by means of zeros and ones. It must have been re-invented hundreds of times by computer programmers needing to represent events in binary.

Nevertheless, with $N^i(x)$ and $Y^i(x)$ representing the data, and μ_x as a model of the underlying 'mechanism' generating the data, we have the key to many problems of survival models.

4.4 *The Multiplicative Model*

Just above, we called μ_x the "... model of the underlying 'mechanism' generating the data ...", referring of course to the heuristic, that the probability of death occurring between ages x and $x+h$, conditional on not having occurred beforehand, is approximately $\mu_x\, h$ (for small h). In fact, we make a small adjustment with a truly profound effect.

Define the stochastic hazard rate at age x to be the product $Y^i(x)\,\mu_x$, also called the Aalen multiplicative model (Andersen et al. 1993) (both names derive from the fact that the hazard rate is multiplied by a stochastic indicator). This represents a hazard rate tailored to the ith individual, that is automatically switched 'on' while they are alive and 'off' at any other time. Where before, we have had to qualify almost everything we said with the mantra "conditional on the life being alive at age x" or the like, this is taken care of by the stochastic hazard rate. This explains the title 'Dynamic Life History Models' of this section; the hazard rates that govern the evolution of the life history are themselves stochastic and changed by events.

4.5 *The Stochastic Probability Function: Back to Bernoulli*

The probability function of the life history up to age x, from equation (27), in its product-integral form, can be denoted by $f_0^i(x)$ and written as:

$$f_0^i(x) = \prod_{s\in(0,x]} \left(1 - Y^i(s)\,\mu_s\,ds\right)^{1-dN^i(s)} \left(Y^i(s)\,\mu_s\right)^{dN^i(s)}. \tag{29}$$

This says, heuristically, that times when the ith individual does not die ($dN^i(s) = 0$) contribute a survival probability $(1 - \mu_s\,ds)$ to the product, while the moment of death ($dN^i(s) = 1$) contributes the death probability $\mu_s\,ds$ (but by convention the ds is not displayed, as $f_0^i(s)$ is then a density function). (In this case the presence of $Y^i(s)$ makes no difference, because the exponents $1 - dN^i(s)$ and $dN^i(s)$ do the job by themselves, but we shall see why it is present in Section 4.6 below. It does matter, however, that $Y^i(x) = dN^i(x) = 1$ when x is the age at death.) Table 4 ('Untruncated/Uncensored')

[2]At least until the Central Limit Theorem is encountered, in which the limiting process is an Itô process, see Andersen et al. (1993).

Table 4: Contributions to the probability function of the infinitesimal Bernoulli trials (equation (29)) from elements of the observed life history, in the absence of left-truncation and right-censoring, and in their presence. Technical point: $N^i(s)$ has right-continuous sample paths and $Y^i(s)$ has left-continuous sample paths (Section 4.3), so at the time of an observed death $N^i(s) = Y^i(s) = 1$.

Period/Time	Observed	$Y^i(s)$	$N^i(s)$	$dN^i(s)$	$df_0^i(x)$
Untruncated/Uncensored					
Before death	Yes	1	0	0	$(1 - Y^i(s)\,\mu_s\,ds)$
Time of death	Yes	1	1	1	$Y^i(s)\,\mu_s\,ds$
After death	Yes	0	1	0	1
Truncated/Censored					
Before death	No	0	0	0	1
Before death	Yes	1	0	0	$(1 - Y^i(s)\,\mu_s\,ds)$
Time of death	No	0	1	1	1
Time of death	Yes	1	1	1	$Y^i(s)\,\mu_s\,ds$
After death	No	0	1	0	1
After death	Yes	0	1	0	1

shows the contributions to the likelihood (29) at different points in the observed life history.

In other words, at every time s when the ith individual is alive there is an infinitesimal Bernoulli trial with probability of death $Y^i(s)\,\mu_s\,ds$. See Gill (1994) on an infinite Bernoulli process. Some authors make product-integration the starting point of survival analysis (Cox & Oakes 1984, Lancaster 1990, Kalbfleisch & Prentice 2002). Actuaries are so strongly oriented towards binomial and Poisson models, however, that we have approached product-integration from there.

4.6 *Left-truncation and Right-censoring*

Left-truncation arises when the first part of a lifetime is unobserved. It is a fundamental characteristic of actuarial data, given that the vast majority of insured lives only become known to the insurer as adults. Right-censoring arises when the lifetime leaves observation before the event of interest (such as death) has occurred. There are many causes of right-censoring; Richards & Macdonald (2024) discuss a wide range of right-censoring events in the context of pensions and annuities.

Left-truncation and right-censoring can, in most cases, be allowed for very simply by adjusting the definition of the indicator $Y^i(s)$, as follows:

$$Y^i(s) = I_{\{i\text{th individual alive and under observation at age } s^-\}}. \tag{30}$$

With this change, everything said in Sections 4.1 to 4.5, including the important representation in equation (29), remains valid. Table 4 ('Truncated/Censored') shows the contributions to the likelihood (29) at different points in the observed and unobserved life

history, where $Y^i(s)$ indicates the presence of left-truncation and right-censoring by taking the value 0. Equation (29) as derived in Section 4.3 describes a purely mathematical probabilistic model. Substituting the indicator processes in equation (30) turns it into the basis of a statistical model involving data.

Figure 1 shows values of $Y^i(x)$ and $N^i(x)$ for two individuals. In the first row, we see observation of the complete lifetime T_0^i from birth to death, here at age 81. This illustrates the $Y^i(x)$ of equation (28). In the second row, one who enters observation at age 35 (left-truncation) and leaves without dying at age 45 (right-censoring), then re-enters observation at age 55 (left-truncation) and dies at age 72. This could happen if, for example, extracts from two different policy files are found during data cleaning to refer to the same individual. This illustrates the $Y^i(x)$ of equation (30).

The impact of this change in simplifying the mathematics is more than its apparent innocence would suggest. An example will be seen in Section 4.7.

4.7 *Example: Individual versus Age-grouped Data for Multiple Lives Again*

In Section 3.5 and Appendix 1 we showed that the likelihoods for individual and age-grouped data were the same, and pseudo-Poisson in form. The method was to derive contributions to the likelihood arising from the smallest possible 'units' of exposure to risk, namely the intersection of the ith individual's lifetime and the kth rate interval. The total likelihood was then the product of all these 'unit' likelihoods over all individuals and all rate intervals. The proof in Appendix 1 is not technically difficult, but is burdened with the notation needed to define intervals and their intersections. By way of contrast we give below an alternative proof using the counting process representation of the data.

The contribution of the ith individual to the likelihood is:

$$L_i = \prod_{s\in(0,\infty]} \left(1 - Y^i(s)\,\mu_s\,ds\right)^{1-dN^i(s)} \left(Y^i(s)\,\mu_s\right)^{dN^i(s)}. \tag{31}$$

Therefore the total likelihood is:

$$L = \prod_i L_i = \prod_i \prod_{s\in(0,\infty]} \left(1 - Y^i(s)\,\mu_s\,ds\right)^{1-dN^i(s)} \left(Y^i(s)\,\mu_s\right)^{dN^i(s)}. \tag{32}$$

Split the age range into rate intervals Δ_k, noting that there is no contribution outside the age range $[r_0, r_K]$:

$$L = \prod_i \prod_{k=1}^{K} \prod_{s\in\Delta_k} \left(1 - Y^i(s)\,\mu_s\,ds\right)^{1-dN^i(s)} \left(Y^i(s)\,\mu_s\right)^{dN^i(s)}. \tag{33}$$

Now change the order of the two outer products:

$$L \;=\; \prod_{k=1}^{K} \left[\prod_i \prod_{s\in\Delta_k} \left(1 - Y^i(s)\,\mu_s\,ds\right)^{1-dN^i(s)} \left(Y^i(s)\,\mu_s\right)^{dN^i(s)} \right] = \prod_{k=1}^{K} L_k \tag{34}$$

noting that the terms in large brackets in (34) are the contributions from each rate interval[3], which we denote by L_k.

The simplicity and directness of the proof above arises from the simplicity of the range of integration of the innermost integral, namely 0 to ∞ instead of an interval defined by the intersection of two other intervals, see Appendix 1. This reinforces an observation by Lidstone (1905) to the effect that it may be simpler to investigate what happens moment-by-moment, than over an extended interval:

> "... it will be found that the formulae are in reality simplified through the absence of any distinction between the beginning and end of the momently intervals under consideration." (Lidstone 1905)

Of course, the work in actually computing such complicated integrals is unchanged, but the greater ease of comprehension makes the task of the theorist (and the reader!) much easier.

4.8 *A Classification of Mortality Models for Actuarial Use*

Figure 2 illustrates how the models underlying survival analysis and occurrence-exposure rates used by actuaries all derive from Bernoulli trials over different time intervals.

(a) The upper branch goes through the instantaneous Bernoulli trial with parameter $\mu_x\,dx$, which through product integration over the age interval Δ_i and simple aggregation over $M > 1$ individuals leads to the survival model based on individual lives.
(b) Similarly, product-integration restricted to the rate interval Δ_k, and a constant hazard rate, leads to the pseudo-Poisson model.
(c) The Poisson model also belongs to the upper branch, but would require a special (and unlikely) observational plan to ensure that E_x^c is deterministic and M is random. An example would be to replace each individual who dies with an identical individual until E_x^c reaches a pre-determined level (Scott 1982).
(d) The lower branch goes through the Bernoulli trial with parameter q_x and leads to the binomial model with $M > 1$ individuals.

Note that both branches lose information about which individuals died in going from the individual model ($M = 1$) to the collective model ($M > 1$) but this will not matter for inference since the likelihood will be changed only by a factor not involving the parameter (for example, in the lower branch, a binomial coefficient). Only in the lower branch, however, is information also lost about the times of death, and therefore the true total of person-years lived. This is genuine loss of information which matters for inference, leading us to prefer any of the models in the upper branch over the binomial model.

[3] A.D.Wilkie, in written comments in the discussion of Richards (2008), advanced a similar argument based on "small age steps h" and a quasi-indicator function similar to $Y^i(x)$; the argument given above is the same as in (Macdonald et al. 2018, Section 5.7).

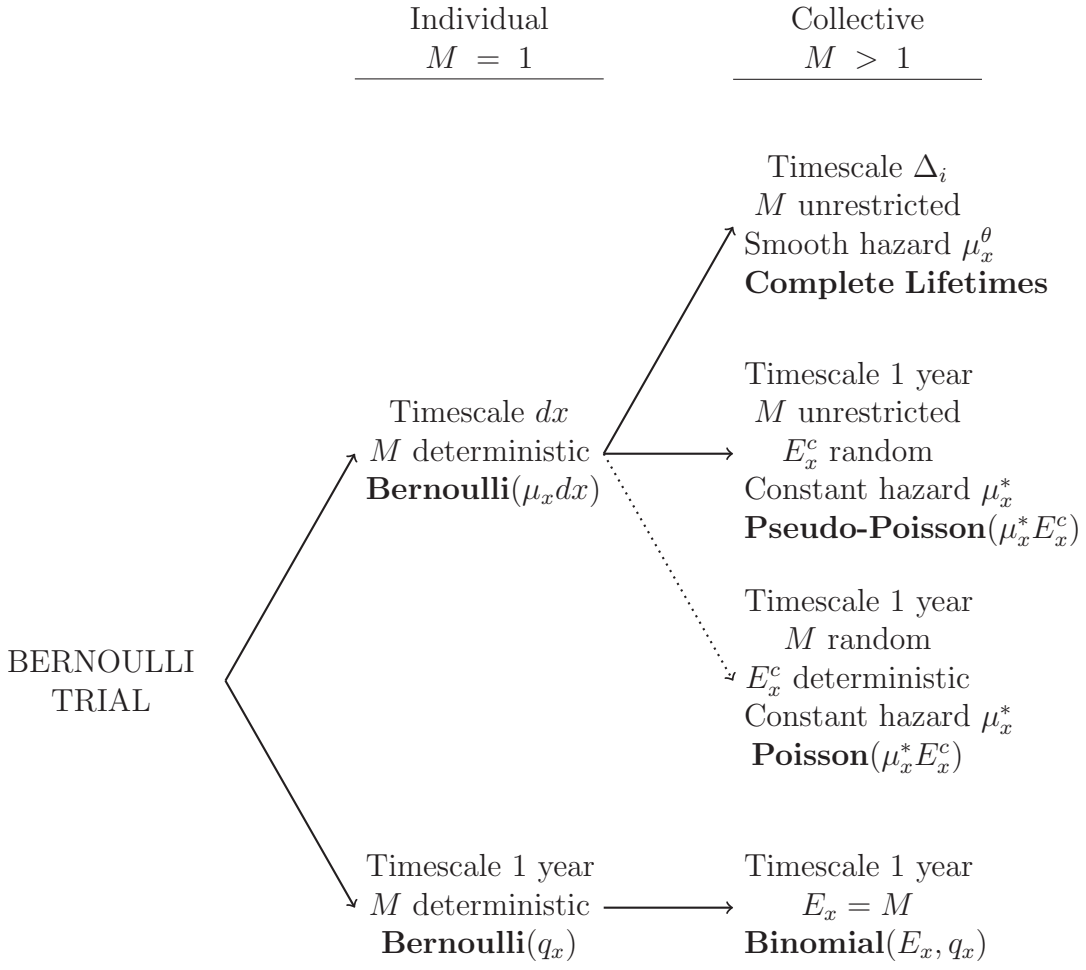


Figure 2: Family tree of models, showing the derivation of individual models, and collective models for individual lives and age-grouped data, (in **bold**) from the basic Bernoulli trial. M is the number of individuals observed, 'M unrestricted' means that M can be either random or deterministic. The tree has two branches, one at the top leading to continuous-time models, including pseudo-Poisson and Poisson models, and one at the bottom leading to the discrete-time binomial model. The dotted arrow indicates that the Poisson model requires the imposition of an observational plan that ensures E_x^c is deterministic, which is unlikely to be realized in practice.

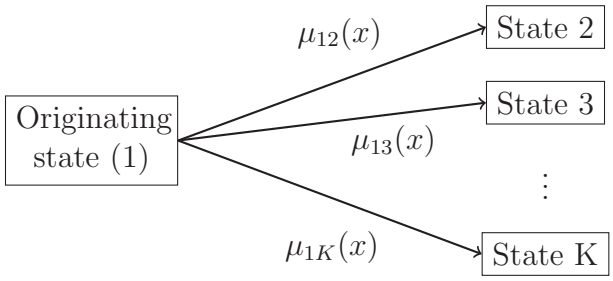


Figure 3: Multiple-decrement model. Note that states $2, 3, \ldots, K$ are absorbing states with no transitions out once entered.

4.9 *Extension to Multiple Decrements*

Equation (29) is easily extended to multiple-state models, with a set of states labelled $1, 2, \ldots, K$. Omitting detailed definitions, process $Y_j^i(x)$ indicates observed presence in state j, and process $N_{jk}^i(x)$ $(j \neq k)$ counts transitions from state j to state k, governed by stochastic intensity $Y_j^i(x)\,\mu_{jk}(x)$, and everything proceeds as before. We will not pursue this in its full generality, referring the reader to Macdonald et al. (2018) (elementary) or Andersen et al. (1993) (advanced), but the simplicity of this extension is an attractive feature of the counting process representation.

However we will sketch briefly the extension to multiple decrement models, since this has been a staple of actuarial textbooks. We have one originating state, labelled 1, in which all life histories begin, and $K-1$ decrement states, labelled $2, 3, \ldots, K$. Transitions are possible from state 1 to any decrement state, governed by intensity $Y_1^i(x)\,\mu_{1k}(x)$ for the ith individual[4]. Figure 3 illustrates this model. Intensities out of a given state are additive, so exit from state 1 is represented by the counting process $N_{1\bullet}^i(x) = N_{12}^i(x) + \ldots + N_{1K}^i(x)$, governed by the total intensity, denoted by $\mu_{1\bullet}(x)$, defined as $\mu_{1\bullet}(x) = \mu_{12}(x) + \ldots, +\mu_{1K}(x)$. The probability function of the life history can be expressed in two rules:

Rule 1: The time of exit of the ith individual from state 1 has probability function similar to equation (29):

$$f^i(x) = \prod_{s\in(0,x]} \left(1 - Y_1^i(s)\,\mu_{1\bullet}(s)\,ds\right)^{1-dN_{1\bullet}^i(s)} \left(Y_1^i(s)\,\mu_{1\bullet}(s)\right)^{dN_{1\bullet}^i(s)}. \tag{35}$$

Rule 2: Conditional on the ith individual exiting state 1 at age x, the probability that the state entered was k is $\mu_{1k}(x)/\mu_{1\bullet}(x)$, $k = 2, \ldots, K$.

Therefore the model is specified completely by a product of infinitesimal Bernoulli trials as in Rule 1, and a simple ratio of intensities as in Rule 2.

If, however, the analyst begins by specifying a binomial-type model, for example based on a time unit of a year, then it is not easy to obtain a convincing representation of behaviour over shorter time periods, and there is certainly no unique solution. The

[4]We could rule all other transitions to be impossible by decree, or assume that all intensities $\mu_{jk}(x)$ with $j \neq 1$ are zero.

classical actuarial approach (see Neill (1977) or Bowers et al. (1997)) involves specifying a hypothetical model of each decrement acting alone, leading to a 'gross' hazard rate acting in the presence of the other decrements, and a 'net' hazard rate acting in their absence[5]. Except in some special cases, further progress is impossible unless 'gross' and 'net' hazard rates are assumed to be equal, but then since 'gross' hazard rates can be shown to be those of the Markov multiple-state model anyway, the modeller is drawn ineluctably towards that destination. Going into more detail would require too much notation, see (Macdonald et al. 2018, Chapter 16).

This illustrates vividly the contrast between the simplicity of the model specified at the 'micro' level, from which, by aggregation, behaviour at the 'macro' level can be deduced; and the perils of specifying the model at the 'macro' level, and then trying to disaggregate it to deduce or intuit behaviour on smaller scales.

4.10 *Further Applications*

We mentioned in Section 4.9 the extension of the counting process representation to multiple-state models, see Macdonald et al. (2018). We mention here other advances and applications based on counting processes, which can be found in the references below.

(a) counting process compensators, martingales, stochastic integrals and central limit theorems (Andersen et al. 1993, Chapter III);
(b) non-parametric estimates including the Nelson-Aalen and Kaplan-Meier estimates (Andersen et al. 1993, Chapters IV.1 and IV.3), (Kalbfleisch & Prentice 2002);
(c) non-parametric kernel smoothing methods (Andersen et al. 1993, Chapter IV.2);
(d) semi-parametric regression models including the Cox model and partial likelihoods (Andersen et al. 1993, Chapter 7), (Kalbfleisch & Prentice 2002);
(e) log-rank comparison tests of survival models (Andersen et al. 1993, Chapter 5), (Kalbfleisch & Prentice 2002);
(f) stochastic reserving models in life insurance Norberg (1991); and
(g) stochastic models of surplus in life insurance, including Hattendorff's theorem (Ramlau-Hansen (1988a,b), Norberg (1991)).

5. Conclusions

In search of sound foundations for mortality models, we began with binomial and Poisson models for grouped counts (Section 2), the basis of traditional graduations of mortality data for actuarial use. Both being models for 'count' data — the number of deaths during a time period, typically a year — they naturally invite questions about mortality at a smaller scale, over a fraction of the time unit. The Poisson model has an answer; its parameter is a hazard rate μ_x, assumed to be constant over the time period, and that defines mortality over short time intervals h as $h \to 0^+$. The binomial model has no such answer; it is up to the modeller to assume how mortality behaves over shrinking time intervals h as $h \to 0^+$. While this is unsatisfactory, the binomial

[5]Thanks to one of the more confusing legacies of actuarial nomenclature, 'gross' and 'net' hazard rates are traditionally called 'dependent' and 'independent' forces of decrement, which have absolutely no connection to statistical notions of dependence and independence.

model does decompose into a sum of Bernoulli trials, each representing the mortality of an individual over the time period.

Modelling the lifetime of a single individual (Section 3) gives the vital insight that the associated likelihood, as in equation (13), has the same form as a Poisson likelihood. Indeed we get the same Poisson-like likelihood if we model individual life histories, or group the data by age (Section 3.5), but in neither case do Poisson random variables feature as part of the model. That is, inference proceeds correctly as if the death counts we observe were Poisson random variables, but they are not. We suggest that all such mortality models based on age-grouped data — which includes most published models — should be called pseudo-Poisson models.

Finally, we identify the fundamental element of a mortality model as the infinitesimal Bernoulli trial; heuristically, an individual alive at age x will die in small time h with probability $h\,\mu_x$, or survive with probability $1 - h\,\mu_x$. To write down probabilities of events over extended time intervals, we need to know how to aggregate such trials, and that requires three ideas new to most actuaries:

(a) the product-integral (Section 4.2 and Appendix 2), as the method of aggregating probabilities of infinitesimal Bernoulli trials over extended intervals;
(b) counting processes (Section 4.3), giving us the natural notation to describe the events in a life history; and
(c) the stochastic hazard rate $Y^i(x)\,\mu_x$ tailored to the life history of the ith individual including left-truncation and right-censoring (Section 4.6).

In conclusion, equation (29), with $Y^i(x)$ as in equation (30), is the 'atom' of a 'continuous-time' survival model, with the qualities listed in Section 1.5 and reproduced below.

(a) It is irreducible, in the sense that it is composed of (infinitesimal) Bernoulli trials.
(b) It is based on behaviour at the 'micro' time scale.
(c) It is based on individual lives.
(d) Aggregated, over time and over individuals, it explains the Poisson-like nature of likelihoods, therefore estimation based on the collective at the 'macro' time scale.
(e) It allows for left-truncation and right-censoring.
(f) It is easily extended to multiple-decrement and multiple-state models.

Acknowledgements

We are grateful to Mr Gavin Ritchie for comments on a draft of this paper.

Competing Interests

None.

Appendix 1

The Likelihoods for Individual and Age-Grouped Data

We show that the likelihood obtained by modelling individual lifetimes (see Section 3.5) is the same as that obtained from age-grouped data for each rate interval. To do so we use the factorization in Section 3.3. Note three points before we begin:

(a) The equality of likelihoods means that inference has the same results using either approach, although each may have advantages for other reasons.

(b) We show equality assuming an arbitrary form of hazard rate μ_x, making no parametric assumptions. This includes as a special case the assumption of piecewise-constant hazard rates, constant on each rate interval, usually made in conjunction with the assumption that total deaths D_x in each rate interval are Poisson random variables.

(c) The result here, which resides in the equality of expressions in equations (40) and (41), is quite detailed, and follows by selecting elements of three different partitions of the age range in precise ways. This contrasts with Section 4.7, where the same result follows as an easy consequence of the definitions in the counting process representation of the data.

Suppose we have M lives, not all identical, the ith individual being observed between ages x_i and $x_i + v_i$ and the random variable d_i indicating death or censoring at age $x_i + v_i$. Define $y_i = x_i + v_i$, and let Δ_i be the interval $(x_i, y_i]$ on which the ith individual is observed.

We wish to introduce a set of rate intervals and write down contributions to the likelihood for the ith individual over those rate intervals that intersect Δ_i. We need some detailed definitions, which we introduce in three steps.

Step 1: Intervals: Let the sequence of ages $r_0 < r_1 < \ldots < r_K$, with $r_0 \leq \min_i x_i$ and $r_K \geq \max_i y_i$, define the rate intervals $\Delta_k = (r_{k-1}, r_k]$ $(k = 1, \ldots, K)$. Let $w_0 < w_1 < \ldots < w_J$ be the sequence formed by the (ordered) union of the three sequences $x_1, \ldots, x_M$, $y_1, \ldots, y_M$ and $r_0, \ldots, r_K$ and define $\Delta_j = (w_{j-1}, w_j]$ $(j = 1, \ldots, J)$. Hence we have rate intervals Δ_k, and the ith individual exposed to risk on interval Δ_i, and the intervals Δ_j are formed from the intersections of all Δ_i and Δ_k. Figure 4 illustrates these three sequences of intervals.

Step 2: Indicators: For $(j = 1, \ldots, J)$, define the sequences of indicators:

$$e_{i,j} = \begin{cases} 1 & \text{if } \Delta_j \subseteq \Delta_i \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

and:

$$d_{i,j} = \begin{cases} d_i & \text{if } w_j = y_i \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

Therefore $e_{i,j}$ indicates that the ith individual was exposed to risk during $(w_{j-1}, w_j]$, and $d_{i,j}$ indicates death or censoring of the ith individual at age w_j.

Step 3: Likelihoods: We can now replace likelihood (13) with the following:

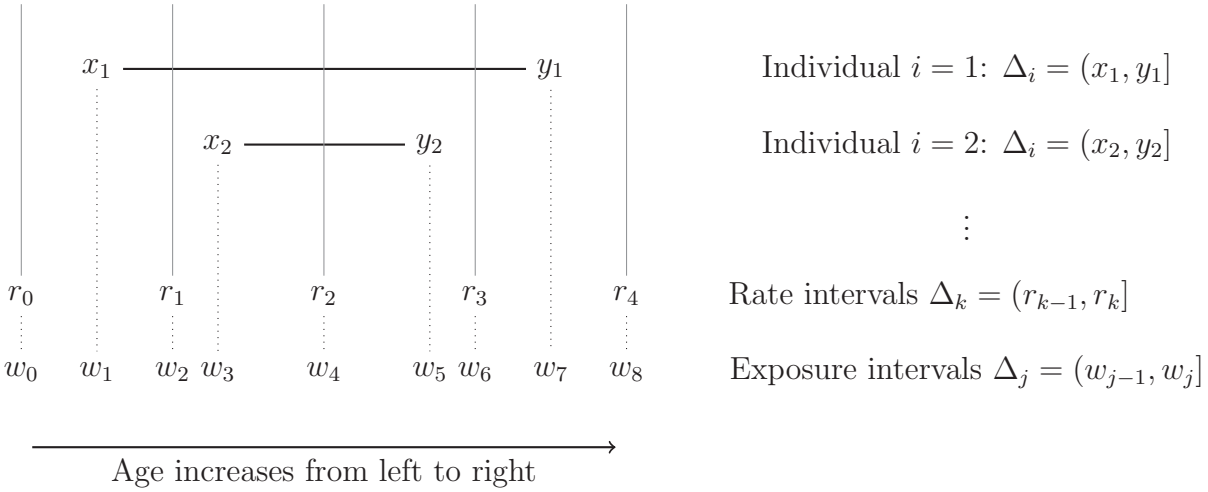


Figure 4: Rate intervals, individual observations and exposure intervals. Rate-interval boundaries, $\{r_k\}$, are set by the analyst, and here are not necessarily integers, nor evenly spaced. The data are paired ages of the start and end of individual observations, (x_i, y_i). The set of exposure-interval boundaries, $\{w_j\}$, is defined as the ordered union of $\{r_k\}$, $\{x_i\}$ and $\{y_i\}$. One consequence is that each exposure interval Δ_j is always completely contained within a corresponding rate interval Δ_k.

$$L_i = \prod_{j=1}^{J} \left[\exp\left(- \int_{\Delta_j} \mu_s \, ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}} \tag{38}$$

and collect together all those intervals Δ_j that are part of the rate interval Δ_k:

$$L_i = \prod_{k=1}^{K} \prod_{\Delta_j \subseteq \Delta_k} \left[\exp\left(- \int_{\Delta_j} \mu_s \, ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}}. \tag{39}$$

The payoff from all this careful defining of points and intervals comes when we form the total likelihood over all M individuals, denoted by L:

$$L = \prod_{i=1}^{M} L_i = \prod_{i=1}^{M} \prod_{k=1}^{K} \prod_{\Delta_j \subseteq \Delta_k} \left[\exp\left(- \int_{\Delta_j} \mu_s \, ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}}. \tag{40}$$

Then by reversing the order of the two outer products, we can collect together contributions to each rate interval instead of contributions to each lifetime:

$$L = \prod_{k=1}^{K} \prod_{i=1}^{M} \prod_{\Delta_j \subseteq \Delta_k} \left[\exp\left(- \int_{\Delta_j} \mu_s \, ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}} = \prod_{k=1}^{K} L_k \tag{41}$$

where we define L_k as:

$$L_k = \prod_{i=1}^{M} \prod_{\Delta_j \subseteq \Delta_k} \left[\exp\left(- \int_{\Delta_j} \mu_s \, ds \right) \right]^{e_{i,j}} \mu_{w_j}^{d_{i,j}}, \tag{42}$$

which we recognize as the total contribution for rate interval Δ_k. The desired equality is shown above, namely $\prod_i L_i = \prod_k L_k$. We note that the intricacy of the definitions and the argument (including the presence of two sets of indicators $d_{i,j}$ and $e_{i,j}$) stems from the need to handle both points and intervals of time, in different combinations. This need is largely abolished, as far as algebra is concerned, by the definition of the process $Y^i(x)$ in Section 4.3.

Appendix 2

The Product-integral

The ordinary integral is familiar to actuaries, the product-integral less so. However, every time an actuary multiplies survival probabilities of the form ${}_tp_x$, she uses a product-integral. It is clear that she uses an ordinary integral, and moreover uses its additive property, since:

$$\begin{aligned} {}_{t+s}p_x &= \exp\left(- \int_0^{t+s} \mu_{x+r} \, dr \right) \\ &= \exp\left(- \int_0^{t} \mu_{x+r} \, dr - \int_t^{t+s} \mu_{x+r} \, dr \right) \\ &= \exp\left(- \int_0^{t} \mu_{x+r} \, dr \right) \exp\left(- \int_0^{s} \mu_{x+t+r} \, dr \right) \\ &= {}_tp_x \, {}_sp_{x+t}. \end{aligned} \tag{43}$$

This suggests the exponential function as a link between functions with additive and multiplicative properties, and indeed it is. Start with the following identity, proved in most courses on real analysis (see (Hardy 1992, pp. 410–411)):

$$\lim_{n \to \infty} \left(1 + \frac{1}{n} \right)^n = e. \tag{44}$$

More generally, assuming we may exchange logarithms and limits and then taking just the first-order term of the Taylor expansion $\log(1+s) = s - s^2/2 + s^3/3 - \ldots$ (convergent on $-1 < s \leq 1$):

$$\log \lim_{n \to \infty} \left(1 + \frac{s}{n} \right)^n \approx \lim_{n \to \infty} n \, \frac{s}{n} = s \tag{45}$$

implying (44) and more. This is homogeneous, in the sense that $n \times s/n = \sum_1^n s/n$ is a sum of n equal summands. Suppose we have a well-behaved function $f(s)$ on an interval $(a, b]$. Partition the interval into n equal sub-intervals denoted by $\Delta_1 = [a, a+h), \ldots, \Delta_n = (b-h, b]$ where $h = (b-a)/n$, and let $f(s_k)$ be the function value at an arbitrarily chosen $s_k \in \Delta_k$ $(k = 1, 2, \ldots, n)$. Then by the same reasoning:

$$\log \lim_{n\to\infty} \prod_{k=1}^{n} (1 + f(s_k)\, h) \approx \lim_{n\to\infty} \sum_{k=1}^{n} f(s_k)\, h = \int_a^b f(s)\, ds. \tag{46}$$

It only remains to replace the interval length h with the more general ds in the limit and exponentiate both sides, giving us the important representation:

$$\prod_{s\in(a,b]} \big(1 + f(s)\, ds\big) = \exp\left(\int_a^b f(s)\, ds\right). \tag{47}$$

Choose the function $f(s) = -\mu_{x+s}$ on the interval $(0, t]$ as in a survival probability, and we have the product-integral representation of the familiar identity:

$$\prod_{s\in(0,t]} \big(1 - \mu_{x+s}\, ds\big) = \exp\left(-\int_0^t \mu_{x+s}\, ds\right) = {}_tp_x \tag{48}$$

which is equation (24).

The above is intuitive and heuristic; for a rigorous account see Gill & Johansen (1990) and references therein.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993), Statistical Models Based on Counting Processes, Springer, New York.

Batten, R. W. (1978), Mortality Table Construction, Prentice-Hall.

Benjamin, B. & Pollard, J. H. (1980), The Analysis of Mortality and Other Actuarial Statistics, Heinemann, London.

Bowers, N., Gerber, H., Hickman, J., Jones, D. & Nesbitt, C. (1997), Actuarial Mathematics, 2nd edition, Society of Actuaries, Shaumburg IL.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. & Balevich, I. (2009), 'A quantitative comparison of stochastic mortality models using data from England and Wales and the United States', North American Actuarial Journal **13**, 1–35.

Cox, D. R. (1972), 'Regression models and life-tables', Journal of the Royal Statistical Society, Series B (Methodological) **34**, 187–220.

Cox, D. R. & Oakes, D. (1984), The Analysis of Survival Data, Chapman-Hall, London.

Currie, I. D. (2016), 'On fitting generalized linear and non-linear models of mortality', Scandinavian Actuarial Journal **2016**, 356–383.

Dickson, D. C. M., Hardy, M. R. & Waters, H. R. (2020), Actuarial Mathematics for Life Contingent Risks, International Series on Actuarial Science, third edn, Cambridge University Press.

Forfar, D. O., McCutcheon, J. J. & Wilkie, A. D. (1988), 'On graduation by mathematical formula', Journal of the Institute of Actuaries **115**, 1–149.

Gill, R. D. (1994), Lectures on survival analysis, in P. Bernard, ed., 'Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Fleur XXII–1992', Springer-Verlag, Berlin, Heidelberg.

Gill, R. D. & Johansen, S. (1990), 'A survey of product-integration with a view toward application in survival analysis', Annals of Statistics **18**, 1501–1555.

Hardy, G. H. (1992), A Course in Pure Mathematics, Cambridge University Press.

Kalbfleisch, J. D. & Prentice, R. L. (2002), The Statistical Analysis of Failure Time Data (second edition), John Wiley & Sons, Hoboken NJ.

Lancaster, T. (1990), The Econometric Analysis of Transition Data, Cambridge University Press, Cambridge.

Lidstone, G. J. (1905), 'Changes to pure premium policy-values consequent upon variations in the rate of interest or the rate of mortality, or upon the introduction of the rate of discontinuance (with discussion)', Journal of the Institute of Actuaries **39**, 209–252.

Macdonald, A. S. (1996), 'An actuarial survey of statistical models for decrement and transition data I: Multiple state, Poisson and binomial models', British Actuarial Journal **2**, 129–155.

Macdonald, A. S., Richards, S. J. & Currie, I. D. (2018), Modelling Mortality with Actuarial Applications, International Series on Actuarial Science, Cambridge University Press.

Neill, A. (1977), Life Contingencies, Heinemann, London.

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized linear models', Journal of the Royal Statistical Society. Series A (General) **135**, 370–384.

Norberg, R. (1991), 'Reserves in life and pension insurance', Scandinavian Actuarial Journal **1991**, 3–24.

R Core Team (2021), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Ramlau-Hansen, H. (1988a), 'The emergence of profit in life insurance', Insurance: Mathematics and Economics **7**, 225–236.

Ramlau-Hansen, H. (1988b), 'Hattendorff's theorem: A Markov chain and counting process approach', Scandinavian Actuarial Journal **1988**, 143–156.

Richards, S. J. (2008), 'Applying survival models to pensioner mortality data (with discussion)', British Actuarial Journal **14**, 257–326.

Richards, S. J. & Macdonald, A. S. (2024), 'On contemporary mortality models for actuarial use I: Practice', Working Paper .

Scott, W. F. (1982), 'Some applications of the Poisson distribution in mortality studies', Transactions of the Faculty of Actuaries **38**, 255–263.

Sverdrup, E. (1965), 'Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between states of health', Skandinavisk Aktuaritidskrift **48**, 184–211.

Waters, H. R. (1984), 'An approach to the study of multiple state models', Journal of the Institute of Actuaries **111**, 363–374.

Wintringham, T. & Blashford-Snell, J. N. (1973), Weapons and Tactics, Penguin Books Ltd., Harmondsworth.

Contact

More information including case studies, latest features, technical documentation and demonstration videos can be found on our website at www.longevitas.co.uk

24a Ainslie Place, Edinburgh, EH3 6AJ
Telephone 0131 315 4470
Email info@longevitas.co.uk

LONGEVITAS™