ABSTRACT OF THE DISCUSSION

Mr S. J. Richards, F.F.A. (introducing the paper)**:** I shall give a short summary of the paper, with some extra details which are not in the paper. My comments are about the need for statistical modelling in modern actuarial work, and about the core of the paper, specifically the so-called geodemographic models for mortality, the criteria for picking a particular model, and the difference between a statistical model and a model suitable for application in financial services. The need for modelling is perhaps best shown in Table D.1, which comes from Richards & Jones (2004).

The table shows the financial impact of different rating factors: gender, lifestyle, time since retirement, size of pension and United Kingdom region, and it shows the impact of step-by-step changes, changing one rating factor at a time. The base case is female, high income, high status, in the South East of England. Making a single change to male changes the annuity factor by over 11%. Changing from high status to low status changes the annuity reserve by a similar amount. What we see is that every one of these rating factor changes is highly material in a financial sense. These are annuity factors at age 65, and, given that the pricing margin for an annuity is perhaps around 5%, every single one of these rating factors is very significant for pricing or profitability. Errors in any one of these could make a substantial dent in the profitability for a life company in the annuity market. Mortality modelling is important, because pricing margins are relatively thin in comparison with the financial impact of some of these rating factors.

My preferred method of mortality modelling is to use survival models. These models deal with mortality at the level of the individual, not of the group, and they model people throughout the whole of the time when they are observed in a portfolio. Lives are not observed for a particular calendar year, or for a particular policy year; they are observed from the moment when the pension starts being paid, to the moment when they cease being observed. This is dealt with further in Section 5.

The difference between the modelling in the paper and classical survival models is that survival models in wider application usually assume that you are modelling from the very start of an individual's life. In life insurance and pensions work, people usually take out contracts much later on in life. This leads to a complication known as *left-truncation*, and this is also dealt with in Section 5.

It is extremely important when building statistical models to carry out deduplication, and to turn benefit or policy-orientated records into a data set of actual people. There are problems in how to recognise duplicate records, and these are dealt with in Section 2.

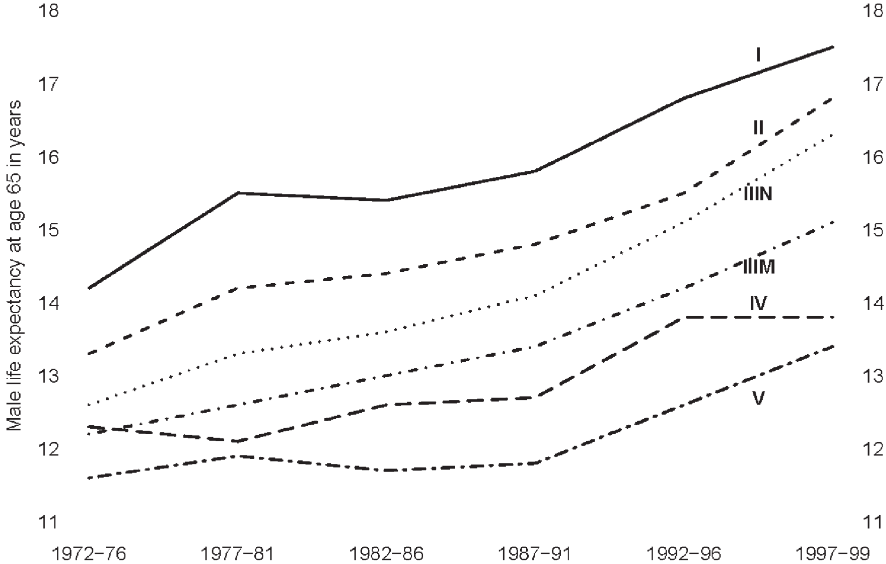
Table D.1. The financial impact of lifestyle

The financial impact of mortality rating factors

Factor	Step change	Reserve	Change
Base case	—	13.39	—
Gender	Female-male	12.14	–9.3%
Lifestyle	Top-bottom	10.94	–9.9%
Duration	Short-long	9.88	–9.7%
Pension size	Large-small	9.36	–5.2%
Region	South-North	8.90	–4.9%
Overall	—	—	–33.6%

Source: Richards & Jones (2004), page 39

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press



Source: ONS Longitudinal Survey

Figure D.1. Retirement life expectancy by socio-economic group

Figure D.1 shows the persistent and, indeed, widening socio-economic differentials among males aged 65 in England and Wales. What we are looking for is a way of tagging people for their socio-economic group. The problem is that the ONS data use the occupation to determine the socio-economic group, and that is seldom recorded in the payment systems.

The traditional actuarial approach to dealing with social differentials in mortality is to weight calculations by pension size. This is an increasingly unreliable way of getting at someone's socio-economic group, not least because people's much-increased tendency to switch employers means that their pensions reflect their service period as much as their salary. The problem is that, to look at a particular policy or pension in isolation, gives you only one part of someone's total wealth or income. What we do in the paper is to use a person's address or postcode to derive their geodemographic profile. This is more reliable, since, regardless of how large or small a person's particular pension is, their address will be the same. This is further developed in Section 3.

One of the last things which we do before we build models is to look at the Cramer's V statistic (Cramer, 1999), which measures the strength of the association between two variables, and this is described in Section 4. Having checked the Cramer's V statistic, we are almost ready to fit a model, as is discussed in Section 8.

There is a wide choice of actuarial mortality laws which we could fit, each named after the actuary who proposed it. These are all models proposed by British actuaries over the past 200 or so years, as described in Section 6, and these actuarial mortality laws fit very neatly into the statistical modelling framework which is used in the paper. Of particular interest are the Beard and the Makeham-Beard models, both of which incorporate ρ, which we call the Beard parameter. We use this as a marker as to whether or not a model has explained most of the variation present. As long as ρ is significantly non-zero, it tells us that there is more variation to be found in the dataset beyond that which our model has by way of rating factors.

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

As with Cox (1972) and elsewhere, we pick one group as the baseline, and measure the mortality effects against this baseline. For example, we might pick females to be the baseline, and measure the effect of being male (or vice versa). Equally, we could pick non-smokers to be the baseline, and measure the effect of being a smoker. Whenever we fit a categorical variable, we need one fewer parameters than there are levels. So, if there are two genders, we would need one parameter to measure the effect of being male; if there were seven socio-economic levels, we would need six parameters to fit the socio-economic groups. Equation (8) shows an example where the baseline would be female non-smokers. We would measure the effect of being male as a departure from this, and the effect of being a smoker as a departure from it as well, and this is discussed further in ¶7.2. The advantage of this approach is that there is no limit to the number of risk factors which you can incorporate into the model. The statistic which we use to choose between models is the Akaike Information Criterion (AIC) (Akaike, 1987), and this is discussed in ¶¶7.4 and 7.5.

What we find in the paper is that a combination of postcode-driven lifestyle and pension size does much better at explaining mortality variation than either factor on its own. In fact, the combination of the two is better than you would be led to believe, just by looking at each one on its own. This is a phenomenon called enhancement, where the more significant rating factors which you add to the model, the greater the power of the existing ones.

The last area of concern is the difference between a statistical model and a model suitable for financial applications. Statistical models are lives based by definition, whereas financial liabilities are anything but lives based, and these are discussed in Section 11.

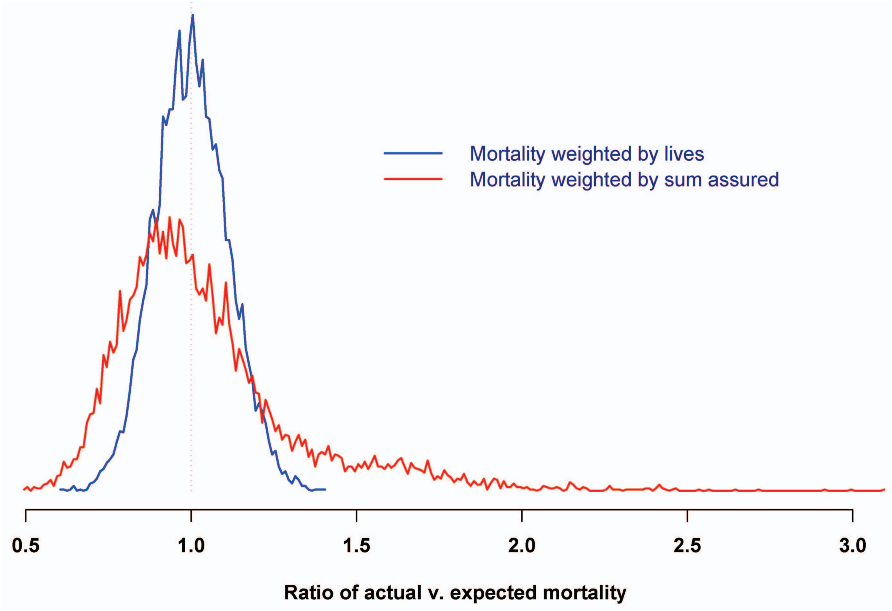
Table 14 shows the concentration statistics for a large life office portfolio of annuities and a large portfolio of defined benefit pensions. In both cases we see a common theme, namely a tendency for the top 10% of pensioners to have around half of all the pensions in payment, representing a tremendous concentration of liabilities. The problem for any statistical lives-based model is that the members of this group have only a 10% representation in a statistical model, yet they carry five times the representation in terms of their financial impact on the portfolio. In other words, they have exactly the same model weighting as the bottom decile, which has only one hundredth of the financial impact. One must be very careful when using statistical models for financial applications, because of this tremendous concentration of risks within a fairly small sub-population. One way in which we can test this is to use bootstrapping, where we sample portfolios from our data randomly, and we use a model to profile and to predict the probability of these people dying. We then check to see whether or not they did die. So, we use the model to predict for data where we know the answer, and we compare what the model predicts with what actually happens. We repeat this many times, so that we can look at the variation of actual versus expected mortality, and we can weight the calculations by both lives and amounts.

Figure D.2 is for a term assurance portfolio. We can see that this model does a broadly decent job; the ratio of actual deaths to expected deaths predicted by the model is around one, certainly for the lives-based approach. It is slightly biased, slightly off-centred, for the amounts-weighted ratio.

However, what is of particular interest in this figure is the long, fat tail towards the right. Although the model in this example is, perhaps, good enough for financial applications, the fat tail on the right suggests that the mortality is slightly heavier than is predicted by the model. This may require you to make some adjustments to the model before you use it in financial applications. Other examples are given in Section 11.

In summary, insured data are natural fits for individual survival models; you know the exact date when the benefit started being paid, and you know the exact date when it stopped. Careful data preparation is extremely important to ensure the validity of the independence assumption. We found that geodemographic models provide substantial improvements for models for mortality and other risks, over and above what you might expect from just using the benefit size. In particular, the combination of the benefit size and the postcode usually provides a much better fit than just using either factor on its own.

Mr M. R. Kipling, F.I.A. (opening the discussion)**:** The evaluation of mortality risk in all its



Source: Bootstrapped experience for portfolio of 50,000 term assurances, repeated 10,000 times; Longevitas Ltd calculations using model for mortality experience of a portfolio of nearly 1 million term-assurance policies between 2002 and end-2006; model is Age + Gender + SelectPeriod + Smoker + JointLife + Product + Size

Figure D.2. Bootstrapping a term assurance model

forms has always been a fundamental part of the profession of an actuary. However, to those of us who worked through the last three or four decades of the last century, it seemed to be a risk which was under control. Our attention was turned more to newer and more exciting problems: the rise and fall of inflation and interest rates; the apparently endless rise and then eventual fall of the equity market; and the need for ever greater transparency and fairness for consumers of financial services.

We developed new techniques for dealing with these risks, most prominently stochastic modelling and the use of ever more targeted and complex hedging vehicles. We believe that we have the behaviour of these newer risks understood and controlled, to the extent of our risk appetite.

So why, despite all these far more interesting possibilities, are we discussing pensioner longevity? Moreover, why are we at the end of a run of no fewer than five consecutive sessional meetings in England and Scotland on this very topic? It is because, suddenly, mortality has become interesting again!

Masked by high inflation and high investment returns, and by the relative youth of pension schemes and life insurance annuity portfolios, the initial acceleration in the rate of improvement of longevity did not feature usually amongst the most important risks to control and to understand. With the masking now stripped away and with the maturing of the owning vehicles, longevity has recently leapt up the list of risk managers' priorities.

We cannot rely on the comfortable techniques of the last century to address this re-emergent

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

risk. However, if market solutions are to be found to transfer the risks from their current unwilling and over-exposed owners to those with the appetite and the capital to take them on for reward, then modern techniques, on which sophisticated investors are prepared to rely, are needed. We need 21st century solutions to tackle this 21st century problem, and we thank the author and others like him that we now have the opportunity to see these techniques emerge.

There are many things which need to be right when pricing annuity risk. As the Pension Regulator's (TPR's) recent consultation document emphasises, two crucial demographic elements are a best estimate of the current rates of mortality and a best estimate of the speeds and the directions in which these will change in future. An essential accompaniment for most actuarial applications is the estimated distribution of actual future longevity outcomes around these best estimates.

The paper focuses on the first of these elements, the current level of mortality, and, in particular, on how this differs according to a number of measurable proxy factors for direct, causal influences. The other sessional meetings this year have also largely been about this element, although 2007, of course, saw the publication of the Continuous Mortality Investigation's (CMI's) library of mortality projections, and TPR has also recently pronounced on the subject.

An important aspect of the paper is its use of survival models. This is a technique which Professor Angus Macdonald drew to our attention in 1996, although it was first applied to mortality by Cox (1972) long ago. The author and colleagues more recently presented specific applications of survival models applied to t-year survival probabilities in their paper Richards *et al*. (2007). This demonstrated, in particular, the ability of this method to separate out time trends from cohort trends in longevity improvements. Tantalising glimpses were also given of the possibility of introducing other parameters into the model besides age, birth cohort and time. The present paper reveals more.

The paper can be divided into two parts. Sections 2, 3, 4 and 8 provide an illuminating and occasionally amusing, canter through the process of cleaning up a life office or a pension scheme database to provide reliable data on which to work. Anyone embarking upon analysing their own database would be well advised to use this paper as a checklist. Sorting data into numerical order and looking at the extreme values is one which I recommend highly. I was recently doing this on a portfolio of deferred annuities, and it revealed the highest planned retirement age as 285; almost certainly an error.

The paper then proceeds to explain the methodology which has been adopted, carefully relating it to the 'laws' of mortality with which we are all familiar. The meat of the paper begins in Section 7, where the concept of the Gompertz α and β parameters is extended to individual lives, and is sub-parameterised in relation to the risk factors pertinent to each life (gender, birth cohort, postcode, etc.). A method of avoiding over-parameterisation is also introduced, with the AIC enabling this to be done objectively.

Table 6 presents the first set of results, comparing the ability of the different models to fit the data most efficiently. It can be seen clearly from Figure 4 that models which allow for late-life deceleration, or, to use non-technical parlance, droop a bit at the top, fit better.

Gender, not surprisingly, results in a significant improvement in fit (the AIC reduction of 1800 to 1900) and a further improvement in pension size (around 1250), even using just three optimally-computed bands. The size bands determined here are divided in the low:mid:high ratio of 28:54:18.

In February 2008 the Institute discussed the first ever CMI graduation of occupational pension scheme data (*B.A.J.* **14**, 208-227), and it is interesting to note that the proposed CMI 'Light' amounts table is to be based on approximately the top 25% of their exposure, not too far away from the author's 18%. The CMI 'Light' table produces an expectation of life at age 65 of 18.8 years, compared with only 16.3 years on the 'Heavy' table (based on about 75% of the CMI exposure). This is a 15% longer expectation of life for the 'lighter' group, even with this broad differentiation. Put alternatively, the valuation difference in an annuity at 5% interest is nearly 11%.

The paper continues by examining a lifestyle factor derived from postcodes and Mosaic types.

This is a more efficient fit for the data than size, and also seems to account for more of the difference between the different mortality laws. An even better fit is achieved when both factors are used together, suggesting that they are by no means complete proxies for one another. The author then finds, in Table 11, a significantly wider differential in life expectation and annuity values between the highest and lowest male categories than between the CMI 'Light' and 'Heavy' tables (30% on expectancy and 21% on value).

These material differences show the importance of correctly categorising the members of a pension fund or the incoming annuitants of an insurance company. Indeed, more granularity than may be necessary can be used if it is considered likely that the lives concerned fall at the extreme ends of the size (or any other) categorisation. There might also be a message for the CMI: "Would it be better to issue three tables, 'Light', 'Medium' and 'Heavy', covering, say, the top 20%, the middle 50% and the bottom 30% of the mortality spectrum, respectively?"

In Section 10 the author turns to patterns observable from the residuals. Figure 5 reveals clearly observable initial selection and time dependencies, and possibly some more subtle age/cohort effects. Attempting to strip these out by adding more parameters to the survival model is partially successful, although the patterns in Figure 6 suggest that some further work could achieve more. The less observant reader should note that the scale change on the 'time' graph makes the benefit of adding this parameter appear less than it really is. The clear evidence of initial selection may be something which the CMI would want to consider for future graduations, especially as it appears to be driven, in part, by an increasing prevalence of underwriting.

Finally, in Section 11, the author suggests a practical approach to test whether there are concentrations of risk factors in a particular portfolio which might not be picked up by a model fitted to a wider experience, particularly if relatively few factors have been built into the model. The case of industry-specific mortality is given as a classic example.

Areas where the discussion might usefully focus include:

— What are the most useful factors to which to fit the models? Are there much data about past occupations, for example? Perhaps the bulk-buyout experts can enlighten us.
— At a scheme level, what account needs to be taken of the different risk factors for management, relative to those of blue-collar staff, given the former's typically high benefit levels?
— Are there problems caused by people changing address around the time of retirement? Is the effect of 40 years living in Glasgow eliminated by a move to Torquay, and can we allow for it anyway?
— How can we allow for the increasing number of people who retire abroad? What is the Mosaic type of Marbella?
— How good are the links between proxy factors, such as amounts and addresses, and causal factors, such as health and past occupation?
— Over time, will enough data on the health of those taking out underwritten annuities become available to be useful as survival factors?

Perhaps most importantly of all, how can we link this work in with that on the other crucial mortality unknown, future improvement rates? Can, for example, survival models cast any light on whether mortality differentials between social categories will continue to widen or begin to narrow, or can they add to the accuracy of *P*-spline projections by giving greater insight into the future progress of time and cohort factors?

Also, there will be those here who are quite happy with using more traditional (although not all that traditional) GLM techniques on q_x. It would be useful in the discussion to hear from them, and whether their findings agree with, or differ from, those in the paper.

Mr M. F. J. Edwards, F.I.A.: One aspect of the approach in the paper is the extra power which it gives to us to allow for any non-linearity of mortality within individual years of age. This will be very useful when analysing portfolios of centenarians, or if writing annuities to people in their 90s. If we do have very old portfolios and this non-linearity is of concern, there is nothing to

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

stop us from applying generalised linear model (GLM) techniques to study monthly, rather than annual, mortality.

I have started a comparison with GLMs, because they are the obvious point of comparison in considering the usefulness of survival models, and they have been used in the field of mortality analysis for quite a few years now. In any discussion of survival models, we should ask whether they offer anything substantial which GLMs do not. I notice a general lack of reference to GLMs in the paper, engendering a sort of 'Banquo's ghost' feeling.

The author has, in the past, contended that GLMs cannot be used with many years of data, because we would be violating an independence requirement. This is a contention disagreed with by all GLM practitioners whom I know; the reason why there is not a time series dependence problem follows, quickly and simply, from the fact that we are modelling conditional probabilities. Indeed, if there were any such dependence problem, one could argue that it would invalidate all of the CMI's normal multi-year investigations. It has also been contended that linearity constraints would, in some way, invalidate the use of GLMs, but this results from a curious misunderstanding of what the word 'linear' in GLM means. For reference, I outlined, in the January/February 2008 edition of *The Actuary* magazine ('A new perspective on GLMs'), what the word means in this context.

That is a theoretical perspective on why we are happy with multi-edge GLMs, and, therefore, would want strong reasons to move to a different methodology. The theory behind any particular model, and here I am using 'model' to mean a particular set of numbers describing the experience of a particular portfolio, is irrelevant. Mortality analysis is a completely empirical field, and the only measure of success is what works, by which I mean, being demonstrably predictive, where we test predictiveness on another, independent portion of the data. We work in a completely empirical field in this regard, and, in that vein, we can show, empirically, that multi-year GLM analyses give us far more predictive models than single year analyses, when tested against independent portions of the data. So, whether looking at the question from a theoretical, or the more important empirical perspective, there is no problem with using GLMs on many years' data. Therefore, it becomes more interesting to ask what survival models can offer which GLMs cannot.

I will return to the empirical perspective later, because it is fundamental to any discussion of mortality, but I would first like to consider the parametric formulation of the survival model, and the highly debatable contention in the paper that: "A major advantage of fitting a statistical model is that smoothness is built in ...". Is this a good thing or a bad thing?

In a recent article on survival models, the author reformulated the concept by saying that survival models offered advantages over GLMs, because we can derive better estimates where data are sparse. Alas, the field of data analysis is something of a 'zero sum game'. We cannot analyse what we do not have. For the sake of example, if we do not have any 83-year olds, the only way to get an estimate of q_{83} (or m_{83}, if you prefer) is by imposing a parametric relationship and inferring q_{83} from what is observed at nearby ages. This means that we are deliberately distorting the model by forcing it to fit our a priori parametric preconceptions; very often in this field preconceptions turn out to be misconceptions. The GLM has the advantage of being less parametric; it is far better to look at the crude rates produced by GLMs, which, in practical terms, function as non-parametric methods, and if necessary, because of data weakness, we can then fit *B*-splines (Bessel splines). The ability to run GLMs with some variates described by *B*-splines is very powerful indeed, and gives us the best of both worlds — we first see what the data are saying, then — if we want to — we fit splines where data are sparse. In any case, I strongly recommend working with crude estimates when thinking about which standard tables fit the observed experience.

The wider question of parametric, as opposed to non-parametric, methods is an interesting one. When I was writing some of the ActEd course notes on survival models ten years ago, and was tutoring the survival model/mortality subject, I thought that they offered a promising non-parametric way forward, and I find it disappointing that the approach outlined in the paper follows the parametric route, exposing us to the potentially 'distortive' effect of the modeller's preconceptions.

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

Of course, whether survival model results do, or do not, contain any parametric bias is the sort of thing which we can test with reference to an independent part of the data, which brings us back to the empirical perspective which is so extremely important.

The underlying theory behind any particular model is fundamentally irrelevant, be it a GLM, a survival model, or reading tea leaves. There is no magic 'best formula' to describe how mortality works. The only thing which matters is how predictive our particular results are, where we test how predictive they are using an independent testing part of the data. This is especially important when working with complex multivariate models, and even more so if experimenting with a new methodology (which is the case here).

This empirical aspect, the testing of predictiveness, is fundamental, not only in testing, in absolute terms, the validity of any one model, but also in comparing different approaches. If you should wish to say that one method is better than another, we would wish to see some sort of comparison of predictiveness. Again, the empirical aspect is what is completely fundamental here if we should wish to choose between different methods. The theoretical underpin of any particular model is of no relevance. What is important is the empirical aspect, the predictiveness, and demonstrating that predictiveness, whether we are looking at one model in isolation or wishing to compare different models.

This empirical perspective is particularly important in checking that we have not overfitted the model to the data, and some of the categorisation procedures outlined in the paper do carry with them this very risk of overfitting — ending up with a self-prophesying model of the postcode factor, in particular. Any general insurance pricing actuary can explain what the dangers are when grouping such factors, and how to avoid the problem.

On the subject of segmentation and the geodemographic factors, what matters in a mortality analysis, done for the sake of portfolio pricing, is what will have a material and valid effect on the net present value (NPV) of the portfolio cash flows. In that context, it is normally the case that any segmentation over and beyond segmentation by benefit amount, although it may appear interesting in abstract mortality terms, will not have a material and valid effect on the NPV. We can come back later to this question of 'valid', which is how much we can project observed historical multiplicative relationships accurately.

However, supposing that we do wish to do something intelligent with postcodes. What should we do? The socio-economic data discussed in the paper are just proxies (in fact, proxies with a very non-transparent, 'black box' origin; with apologies to any guests from those particular firms).

If modelling mortality, we should be basing our postcode groupings on mortality, not on a proxy. These simplistic proxies may serve as a useful 'beginners' introduction' to the subject, giving a point of comparison for proper postcode groupings, but any serious modelling in this field should focus on the mortality characteristics of the portfolio in question.

A better approach is to calculate the residual mortality effect in every geographical region, and, from there, derive mortality-based postcode groupings. The work is quite fiddly, because of the need to avoid the self-prophesying model trap, but it is not difficult to come up with predictive postcode groupings which are based on the mortality experience of the portfolio in question. This is a more valid approach than basing our work on black-box off-the-shelf proxies, and will, in general, give more predictive categorisations. It is a particular concern of mine in the current bulk purchase annuity market, where we have the strange 'group-think' situation of a very competitive market, but many players, not only using the same 'recipes', but recipes which seem to be inherently flawed, because they are based on simplistic proxies and dangerous categorisation techniques.

I have two further observations. First, I was surprised by the treatment of the year of birth in the paper, where we seem to be breaking things down into just three very broad year of birth groups. With GLMs, we normally find no problem with a year-by-year treatment of the year of birth, and this can bring out surprising features of annuity portfolios and pension schemes. Very often we see almost the complete opposite of what people generally consider to be the 'cohort effect'. This is obviously very useful information in thinking about improvement assumptions.

Secondly, there is a very thorough and impressive section on data handling, which will, no doubt, become a standard reference on this for some time to come. This is, clearly, a vital part of any mortality analysis, but there is a risk of going down the wrong track as regards grouping together the benefit segments held by individuals. Although, intuitively and statistically, it is right to group benefit segments, if we conduct our analysis on data regrouped in that way, but if our mortality results are then applied to 'raw' ungrouped valuation data, then there is an inconsistency which could easily make the whole process wrong.

I suspect that this is a situation which may have been happening a lot, with a 'disconnect' between the mortality analysis data preparation and how the cash flow projection teams treat the data. The paper, itself, is not saying anything wrong here, but it is an obvious issue which many people may have neglected.

My comments have focussed on the differences between GLMs and survival models, and there are quite a few differences. GLMs are more robust and more flexible, are more tested in this field, are less exposed to parametric bias and the modellers' preconceptions, are much more transparent, and offer other applications, such as generating closed-form solutions in asset and liability management (ALM) contexts. The differences are dwarfed by the similarities. They are both mathematically inelegant ways to fit crude multiplicative models to the observed mortality of large amounts of data, using brute force numerical algorithms to find the optimal fit.

Their similarities in this regard mean that they share notable defects. First, we risk assuming the constancy of recently observed multiplicative relationships over the next 30 to 40 years. This is clearly wrong; a socio-economic mortality differential of two, observed when a portfolio's average age was 70, will, of course, not be two in ten years' time, if we are talking about a closed book, which we normally are in this portfolio cash flow valuation context; hence the scepticism about the worth of such segmentation in the context of portfolio cash flow valuations.

Secondly, the methods encourage over-segmentation, which will often be spurious from a financial perspective, and will lead to extra operational risk when the results are used.

Thirdly, and most importantly, we still do not have a clue about mortality improvements, although we may get strong hints from the year of birth or calendar year of exposure trends indicated by the models. Any differences between the GLM and the survival model results will be absolutely swamped by this future variability.

We have two actuaries arguing about the relative merits of two conceptually somewhat similar mortality modelling approaches. We work in an empirical field, and the theory underlying any particular model is irrelevant compared with the empirical question of validated predictiveness, and any differences will be completely swamped by the uncertainties of what the future brings.

I suggest that we do not devote time and effort to dreaming up new mortality modelling techniques, given that survival models do not offer a single advantage over GLMs, and, indeed, offer some disadvantages. I do not see much purpose in reinventing the wheel, or in replacing the wheel with something which may be inferior. A more fruitful line of research would be to come up with analytic techniques which tie in better with recent thinking about future mortality trends.

Since the most promising thinking on improvements seems to be coming from the area of disease-based modelling, it would be useful to start thinking about analyses which can model the experience in a similar way, and whether generalised additive models (GAMs) may offer us a useful way forward, here, in modelling the different building blocks of mortality, and then adding these blocks together.

Dr L. M. Pryor, F.I.A.: There are concerns about the main sections of the paper. However good your model, if it is based on poor data, or if you are failing to model the right thing, you have problems. This paper provides valuable examples of data problems and of how the blind application of a model can be unwise. It brings out some useful points, which are much more widely applicable.

It is extremely useful to have such a clear explanation of a method of geodemographic mortality modelling. We have been hearing about this for a long time, but there has been no readily accessible explanation of how it actually works in practice.

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

Mr A. D. Smith: I have a few observations, the first of which is to comment about the difference between statistical significance and financial significance. The author mentioned how, potentially, you might weight observations by amounts to give something like a by-amount analysis, and that puzzles me. If amount is in the explanatory variables, and you are doing just a regression analysis of y against x, if x is in your explanatory variables, then you are going to hit, not only the mean value of y in the forecast, but also the mean value of x times y. So, I did not see why the weighting was necessary or why they might not align in the first place. That may be my weakness in understanding what was written.

There is one way where we could gauge the financial significance a lot more easily. That would be if the author had chosen to disclose the fitted parameters, as is customarily the case when describing the goodness of fit in models. Instead, what we have is the AIC, which is interesting, but does not really tell us about whether parameters are financially significant. So, for example, you may have a demographic effect which only changes the mortality by ½% of the force of mortality, but is still statistically very significant. Since the number of parameters involved is of the order of five or ten in each case, I assume that it was not reasons of space which prevented their publication.

Another thing which puzzles me is the interpretations of the Beard parameter ρ, which the author contends diagnoses heterogeneity within the population. I tried to figure this out. I was hampered by the fact that Appendix 1 seems to have a bit of a notation conflict, where, for example, α switches into a and then back again. z starts off as a random variable. When I finally unscrambled it, I worked out what underlines this, which is a frailty model, where different individuals have their own personal multiple of the standard table, and the effect then is, if you have a very wide range of individual frailties, the people with the highest frailties die off and you are left with the people with the lower frailties. So, instead of getting an exponentially increasing force of mortality, you get a force of mortality which levels off.

It seems to me that heterogeneity is not the only possible explanation for that effect, and it seems not entirely convincing to allocate a non-zero value of ρ to indicate heterogeneity. It might be just that the initial first guess at the shape was wrong, and that adding an extra parameter made it a bit better.

Mr Richards: The opener raised the question about people changing address upon retirement. It is true that people often change their Mosaic type code when they move from one address to another. What we find, in practice, is that people may change their Mosaic type codes, but that they tend not to change their optimised mortality group.

I take as an example where my parents live, where I live now, and where I lived as a student actuary; although the Mosaic type code is different in each case, all of these type codes belong in the same lifestyle group optimised by the software. While address changes might risk being an issue, in practice people tend not to move between lifestyle groups, especially if you reduce the 61 Mosaic types to a smaller number of groups.

The opener also mentioned the ability of survival models to look at time trends separately for a socio-economic group and separately for other risk factors. This can be done. I did not do it in the paper for reasons of brevity, but, as a rule, I tend to find much stronger time trends for males than for females. These can be identified separately in survival models by fitting time as a variable. If the dataset is large enough, then you would be able to pick up separate time trends by lifestyle group.

One last point about the opener's comment on data; I would not like to take all the credit for the data preparation, as it was Mr Gavin Ritchie who wrote the database code.

To turn to the points raised by Mr Edwards, I disagree with much of what he said. I have nothing in particular against GLMs, and I have published several papers using them. I used them extensively in my initial phase as a consultant. The reason why GLMs are not mentioned in this paper is because survival models are more appropriate. GLMs have a number of limitations when dealing with individual data, and I have yet to find a situation, in practice, where a survival model is not better than a GLM.

I was interested to note that Mr Edwards said that you can model mortality on a monthly basis instead of on an annual basis. Why stop at monthly? Survival models, effectively, model mortality on a daily basis. If you take a GLM for q_x, and progressively shorten the time period when you are modelling q_x, you approach a model for μ_x, which is a survival model.

I take issue with the point about smoothness not being desirable. Smoothness is highly desirable, and a very useful feature of a parametric model, in that smoothness is automatic. Intuitively, it makes sense that mortality for an individual will increase smoothly with increasing age, and a model which imposes that as a constraint automatically gives smoothness in the fitted rates without any separate smoothing step necessary. Where data are sparse, it enables a reasonable structure for mortality to be imposed and permits the 'borrowing' of information from adjacent ages.

Regarding the comments on postcodes, I have analysed in excess of 15 separate life assurance and pension scheme portfolios. In every case I have found the Mosaic type to be a highly useful explanatory variable for mortality differences. I have also looked at regional models of mortality, just using the first one or two characters of the postcode, and whatever significance I have actually seen in that rating factor in the past usually disappears when a postcode-driven lifestyle element is added to the model. For example, the life portfolio which lies behind the examples in this particular paper shows no regional variations in mortality in the presence of a postcode-driven lifestyle group. The reason for this is quite simple; most of what we think of as regional differentials in this country are largely differences in socio-economic group mix. The reason why the North East and North West of England have higher mortality is because there are proportionately more people of higher-mortality lifestyle groups in those parts of the country. What we think of as regional variations are predominantly socio-economic variations.

To address the points of Mr Smith, the use of the weighting is largely applicable where there is an extreme concentration of benefits. In most portfolio cases you do not need to have benefit size as a rating factor *and* to weight the parameters. The term assurance example showed that the unweighted model did an almost ideal job. For pensions and annuities, however, the concentration is usually so extreme that the financial impact of a very small number of individuals is understated by a lives-based model, whether a GLM or a survival model. That is simply because they are such a small proportion of the overall headcount, but such a large proportion of the liabilities. For most product lines the weighting is, perhaps, unnecessary. For something like pensions' portfolios, where there is such an extreme concentration, you often find that it is useful to do this additional weighting step over and above using a benefit size or a pension size in the model.

I agree with Mr Smith that heterogeneity is not the only thing which may lie behind what I have suggested for the Beard parameter. I am merely suggesting that it is one possible explanation for this, one which has a backing in the mathematical derivations in Appendices 2 and 3. With a succession of models, I showed how this Beard parameter, which I suggested could be used as an indicator of unexplained variation, did, in fact, shrink from model to model, as we added risk factors and improved the model fit. I fully accept the point that heterogeneity is not the only interpretation for that parameter. Mr Smith asked about why there are not any fitted parameters. There are some, but I wanted to avoid the paper being a 'wall of numbers'.

Mr P. D. G. Tompkins, F.I.A.: We are seeing the use of postcode rating and geodemographic work in many of the pricing activities going on for those who are buying and selling annuity products. In particular, it is being used at the interface between the life offices which are providing the solution for some of the pension schemes which want to lose annuities. The predictive nature of the work which the author and others are doing is, in part, leading to an assessment of what those who are doing the pricing are trying to do when they come up with quotations for the costs of taking the annuities onto their books. Perhaps there is a danger of a herd instinct developing around the field of postcode analysis.

It was interesting to see the comments that the author put in the paper about a range of possible choices of databases. There are many thousands of postcodes and associated subgroupings, and there is much change happening. If you are adding, shall we say, 10% to the housing stock of the U.K. in the next 15 years, as well as changes in the quality and the location

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

of housing around the country, it is probably quite important, both for your analysis and for the work which goes on with the various agencies which do the analysis of the subgroupings which they produce, that this is kept under review. This is in order that models can be refined and made better and more accurate, so that different parties can assess whether prices for, say, bulk annuity purchases, are reasonable.

Mr J. M. Lowes, F.I.A.:. With 60-odd geodemographic categories to start with, why did the author choose to divide that number into three groups for the purposes of analysis, rather than, say, four or five groups?

Mr A. G. Sharp, F.F.A.: If I understood the author correctly, he talked about needing a model to have an ever-increasing force of mortality with age. Those of us who were at the recent self-administered pension schemes (SAPS) presentation (*B.A.J.* **14**, 208-227) saw what the SAPS Committee have graduated and have chosen as an arbitrary fixed rate of mortality at the very old ages, because of the very sparse amount of data.

However, looking at what data there are at the very old ages will become ever more important. The SAPS data did pick up on what we can find from the ONS data. I wonder, perhaps, whether we try to make our models too theoretical. I have quite a lot of sympathy with the word 'empirical', which was used earlier.

Mr C. D. O'Brien, F.I.A.: I understand the benefits which life insurers have in using postcodes as an indicator of mortality. I am not quite sure why the same needs to apply to occupational pension schemes, which, presumably, have such information. The author mentioned the increasing differentials by socio-economic class in terms of the expectation of life. The ONS data from autumn 2007 suggested that actually that might be diminishing. So, it may be that there is some lessening of that trend. We have talked about the way in which the postcode is a proxy for the underlying drivers of mortality. The Profession will benefit from understanding the results of medical studies where there are longitudinal surveys of individuals and their mortality.

We have seen data which have given some indication of the link between lifestyle factors of mortality, in particular, tobacco consumption, alcohol consumption, the amount of exercise and also diet. If we understand those drivers, those links between behaviour and mortality, then that will put us in a better position to predict future changes in mortality. We can start thinking about what might be the changes in exercise, diet, and so on, in the future, and, it is to be hoped, we can understand the impact which they will have on mortality.

It raises the issue as to how we underwrite, certainly as regards annuities. It may be quite useful to underwrite annuities by asking someone for their postcode, which seems a fairly neutral question. It may be rather impertinent to start asking questions about how much time they spent at the gym last week, and so on, but that might give better information about mortality.

Mr B. K. Wilson, F.I.A.: On behalf of the CMI SAPS Committee, we are now collecting postcode data, so that all Scheme Actuaries should do their very best to send in data with postcodes, so that we can start analysing data on that basis.

Mr R. J. Houlston, F.I.A.: The opener commented that we would like to see 21st century solutions to mortality. The recent graduations produced by the CMI have been produced by dividing the portfolio and graduating parts of it. I would like to see the Profession move forward, and suggest that the CMI considers using survival models or GLMs to try to get more information from all of the data. However, it may leave the CMI trying to arbitrate between two warring factions.

Mr Edwards: Clarifying one of the earlier points about postcodes, I used the word 'region' in the abstract sense, meaning postcode or postcode sectoral district, whichever the data suggest. On the question of postcodes versus proxies, you should be looking at mortality at the postcode

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

level, deriving the residual mortality effects at that level, and then working from there. That will give mortality-based postcode clusters, which will be more predictive than any proxy. Whether or not they are is something which we will check empirically.

This list of self-prophesying models is of particular interest to me, as I have a general insurance background as well as a life background. General insurance people will be familiar with the problem that it is very easy to overfit a model to the postcode factor, in particular, so that you end up with a model which actually loses hugely in its predictiveness, because you have not gone through the right hoops.

The problem of self-prophesying models is a classic trap, into which it is very easy for analysts, without a full, wide range of experience, to fall, especially if the resulting models are not properly checked for predictiveness against an independent part of the data.

Describing the problem in general terms, it consists of deriving models which are fine tuned to replicate the observed experience, in a way which systematically magnifies distortions arising from random fluctuations, consequently making the model much less predictive of future events than it should be or it could be.

We can illustrate this by way of an example. Suppose that we have an observed experience of around 10,000 deaths and a geodemographic classification involving around 10,000 sectors, with 'sector' being defined in a completely abstract sense. Assume that we have a fairly even exposure distribution. We see, on average, the order of around one death in each sector, but whether we have zero, one or maybe two or more deaths in any sector is highly susceptible to random fluctuations. People have a tendency to live or to die in an 'integer' sort of way, without paying due respect to all of the mortality formulae in the established literature.

If we use these results to order thousands of sectors in some way and then combine them into a smaller number of groups, our groupings will obviously be influenced by this randomness. At this stage we do not necessarily have a problem. However, if we now use the same data to quantify the mortality effects associated with each of these groups, the effect of the random fluctuations feeds in again, in a systematic and highly distorting fashion, to give us a model, which will be fine tuned to replicate the observed experience, but at a cost of being much worse at predicting future events.

In Section 7 there is an automated grouping procedure described, which initially comes across as very clever. However, if you apply a very clever automating procedure, then that procedure is likely to exacerbate the effect in question about as much as it is possible to do; therefore giving us a model which will be especially clever in trying to replicate historical random fluctuations and especially unclever at predicting future mortality. The problem will get worse if we are starting with a very granular decomposition of the portfolio, or if we are dealing with a modelled event with a low claim frequency; that is, mortality as opposed to motor third party.

In the present instance we have both of those conditions together, so that the potential problem is about as bad as it can get. It is a well known trap to be aware of in the general insurance field, and there are some standard ways to treat the data to avoid falling into this trap. Most professional multivariate software packages provide the necessary tools to take users through the hoops, and to avoid falling into this self-prophesying model fallacy.

It is interesting to consider what the financial effects of that sort of error might be. Basically, we are getting the geodemographic factors very wrong, because we have produced a model which replicates random fluctuations at the cost of being much less predictive. If we are deriving substantially wrong geodemographic factors, but are applying these to the relative mortality experience of subgroups within a portfolio with a well-estimated overall level of mortality experience, we might expect the overall basis to be materially incorrect, but not disastrously so. In effect, we are moving cards around the deck, and getting the order of the cards wrong in some way, but we are not miscounting the total number of cards. That is a fairly poor analogy, but it is the best that I could think of.

If, however, we take substantially wrong geodemographic factors, and apply them in isolation to estimate the mortality experience of some other book of business, where we do not know the overall level of mortality reliably, then the areas discussed will feed directly through to give a potentially very wrong mortality basis, and they are likely to be even more substantially wrong where the geodemographic profile of the book in question differs materially from that of the datasets used to derive the factors.

Obviously, if that mortality basis were then to be used in pricing some bulk annuity bid, the results could be mispriced by many millions of pounds.

Mr N. D. V. Bodie, F.I.A. (closing the discussion)**:** One of the principles which I understood when I studied economics was that there might be different ways of reaching a similar answer. We should not seek to throw away either of the methodologies which have been espoused so powerfully, because we need to know the answers to the question.

As Mr Edwards pointed out, there is no point in starting from the wrong place as well as needing to know about what future improvements are going to do. Our ability to project future improvements, our understanding of the linkage between diseases, is such that we are really groping very much in the dark when the Regulator has talked about making evidence-based decisions.

We have a fighting chance of getting that right in terms of current mortality. In the absence of a proven form of time travel, we have no evidence about future improvements, and so we have to develop our own theories. We are going to have to debate them. We must get the starting point right.

There was some discussion over the question of whether smoothness is desirable. I sit with Mr Wilson on the SAPS Committee, and we certainly found that, particularly when looking at all pensioner data, smoothness does not exist. Once you have gone below age 50, you go from a grouping of people, some of whom are ill-health pensioners and some of whom are normal health pensioners, and at age 49 down you have none but ill-health pensioners. So, I do not find myself persuaded by a desire for smoothness, because, quite clearly, the way in which pension schemes operate can, at certain points, militate very much against smoothness.

The point was also raised: "Can occupational schemes not provide their own data with regard to regional fluctuations in mortality?" There is a significant problem with this. The largest schemes will tend, quite often, not to be focused in particular areas. There will be some schemes which will be focused in particular areas of manufacturing, say, but they will also have sales forces which are spread over a wide range of the country. Other large schemes will have their workforces spread all over the country, and we then have the same problem trying to identify regional variation, and will end up, probably, using postcodes again.

However, the smaller schemes are the ones which have the real problem. They are local, but they do not have sufficient data to be able to do their own projections. So, we have to come up with some model which will enable us to use more broadly-based indicators of what mortality rates are before we then go into modelling the future.

Mr Richards (replying)**:** To bring out the point raised about the CMI using survival models, the CMI does already use survival models, and has done for many years; they are used for PHI experience, and for the graduation of mortality tables, such as the '00' series and the SAPS tables. The software which the CMI uses fits models for the force of mortality, i.e. a survival model, albeit that this is done for group data rather than for individual data, as here. The CMI has been using survival models for, perhaps, 25 to 30 years.

Mr Lowes asked why I had chosen three groups. One reason is simplicity. The dataset here would, indeed, support more lifestyle groups on a statistical basis. One reason for not using more is that, although they are significantly different statistically, there are actually some very small numbers of very high mortality individuals of not particularly great financial significance, so, while I could have fitted five, or even seven, mortality groups to this particular data set, some of them would have been very small groups, and not of any financial significance. I merely chose to keep things simple.

There was a question about using postcodes as a proxy. Obviously, something like smoker status would be useful. However, the reason for using postcodes is twofold. First, they are easily available, whereas smoker status for pensioners is usually not. The second reason for using it for pricing is that it is an unobtrusive piece of data which a life insurance company will already

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

collect, for example, whereas collecting things like smoker status will be a more intrusive process, and will require asking for new information (or for information which is simply unavailable).

One of the advantages of using postcodes and pension sizes is that this is information which you already have for existing business. You do not need to do anything new to change your underwriting process.

Finally, now that Mr Edwards has defined self-prophesying models, I can assure him that this does not apply to the survival models in this paper. In particular, the 'small groups' issue which he raised does not apply to the way in which these models were actually fitted. Each lifestyle group contained thousands, if not tens of thousands, of deaths. The aggregation takes place through mapping post code to Mosaic type, then mapping Mosaic type to one of three lifestyle groups, and only *then* fitting the model. Mr Edwards outlined fitting a model on very large numbers of postcode regions with very small numbers of lives and then aggregating that. That would, indeed, lead to exactly the problems which he outlined, but that is not the procedure followed here. The models in this paper are, therefore, not 'self-prophesying', and do not suffer from any of the flaws which Mr Edwards claims.

The President (Mr N. J. Dumbreck, F.I.A.): We have had a good discussion and some controversy, which I hope will continue in the *Journal* and in the pages of *The Actuary*.

It now remains for me to express my thanks and the thanks of all of us to the author, the opener, the closer and to all those who participated in this discussion.

WRITTEN CONTRIBUTIONS

Mr J. Armsworth: Geo-demographic profiling schemes such as Mosaic have long been used to segment the unwieldy and volatile 1.8 million postcodes in the U.K. into a small number of broadly homogenous and stable categories. The data on which Mosaic is based incorporate social, financial, health and other measures applicable across the whole of the U.K. population. The foundation of Mosaic lies in U.K.-wide and fully compliant data sources, such as property values, council-tax bands, company directorships, and many others. The methodology behind the creation of Mosaic is a matter of peer-reviewed research, and is therefore public record. In contrast to Mr Edwards' assertion, Mosaic is a transparent and well-understood categorisation, and has proved its effectiveness in citizen profiling for over 25 years.

Professor S. Haberman, F.I.A.: The author has done a very good job in describing how survival models can be used to represent the variation of mortality within pensioner populations, according to a range of covariables. The modelling aspects are particularly interesting, and I will focus my comments on these aspects.

The author starts his model building with the proportional hazards framework due to Cox (Cox, 1972). The actuarial application of these models can be traced back to Bernard Benjamin's contribution to the discussion at the Royal Statistical Society of Cox's seminal paper. Later applications include Renshaw (1988) and England & Haberman (1993), both looking at impaired lives data sets from life insurance.

In Table 5, the author presents a set of mortality laws which are of interest. Clearly, some of these are non-linear in their parameters. In the statistical literature, there is considerable discussion about the difficulties of fitting non-linear models (see, for example, Ratkowsky, 1990), and the difficulties of obtaining stable and reliable parameter estimates. Often, models need to be re-parametised in order to facilitate the fitting process. This is discussed, for example, in Butt & Haberman (2002, 2004), and more recently by Doray (2008), and represents an important issue of which readers should be aware. It is not clear that the author has fully addressed these difficulties.

Finally, the author touches, in ¶6.5, on the identifiability problem regarding the Beard and Perks laws. There are many structures which can lead to the Beard and Perks models. Thus,

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

Hougaard (1984) notes that: "The frailty distribution is not identifiable, if the frailty is an individual quantity ... It is not possible to divide the variation into that within and that between individuals, if there is only one observation per individual." In a similar vein, Yashin *et al.* (1994) show that, without additional covariate information or assumptions, the fixed frailty Gompertz (or Makeham)/gamma model cannot be distinguished from a range of models incorporating changing frailty at the individual level. At a fundamental level, Hoem (1990) has proved that many mortality laws (at an individual level) and frailty distribution combinations can produce the *same* aggregate mortality law.

References

Butt, Z. & Haberman, S. (2002). Application of frailty-based mortality models to insurance data. Actuarial Research Report No. 142. Cass Business School, City University.

Butt, Z. & Haberman, S. (2004). Application of frailty-based mortality models using generalized linear models. *ASTIN Bulletin*, **34**, 175-198.

Doray, L. (2008). Inference for logistic-type models for the force of mortality. Paper presented to the Society of Actuaries 'Living to 100' Symposium, Orlando, Florida.

England, P. & Haberman, S. (1993). A new approach to the modelling of excess mortality. *Journal of Actuarial Practice*, **1**, 85-117.

Hoem, J. (1990). Identifiability in hazard models with unobserved heterogeneity: the comparability of two apparently contradicting results. *Theoretical Population Biology*, **37**, 124-128.

Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing heterogeneity. *Biometrika*, **71**, 75-83.

Ratkowsky, D. (1990). *Handbook of nonlinear regression models.* Marcel Dekker, New York.

Renshaw, A.E. (1988). Modelling excess mortality using GLIM. *Journal of the Institute of Actuaries*, **115**, 295-312.

Yashin, A., Vaupel, J. & Iachine, I. (1994). A duality in ageing: the equivalence of mortality models based on radically different concepts. *Mechanisms of Ageing and Development*, **74**, 1-14.

G. Ritchie: First, I am not an actuary, but a consultant with long experience of software development in financial services. Second, I must declare an interest — I am, along with the author, a developer of the Longevitas software used to fit the models described in the paper.

This contribution stems from a desire to correct some of the misapprehensions raised by Mr Edwards. He stated that "the only thing which matters is how predictive our particular results are, where we test how predictive they are using an independent testing part of the data''. This statement might be taken to suggest that the predictiveness of the models in the paper had, in some way, gone untested. However, this empirical perspective — the testing of predictiveness — is exactly what the paper covers in Sections 10 and 11. Residuals are used to check the lives-based model fit, and then the randomised bootstrapping technique is used to validate both the lives-based and financial (amounts-based) applicability of the model.

He then observed that, in the paper, there is an automated grouping procedure described at the end of Section 7, which initially comes across as very clever, and, if you apply a very clever automating procedure, then that procedure is likely to exacerbate the effect in question about as much as it is possible to do. The 'clever automating procedure' referred to is simply the search for a statistically optimal classification which involves fitting many hundreds or thousands of models. Automating tasks with many hundreds of steps allows them to be undertaken in a repeatable, consistent and efficient fashion. The approach outlined in the paper does not exist to demonstrate cleverness, but rather to show what is involved in creating optimally fitting lifestyle groups, without requiring subjectivity or rules of thumb.

Mr Edwards then commented that the socioeconomic data discussed in the paper are just proxies (in fact proxies with a very non-transparent 'black box' origin). If modelling mortality, he suggested that we should be basing our postcode groupings on mortality, not on a proxy. These

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

comments show a misunderstanding of the role of geodemographic types in the technique outlined in the paper.

Geodemographic profiling schemes, such as Mosaic, have long been used to segment the unwieldy and volatile 1.8 million U.K. postcodes list into a small number of broadly homogenous and stable categories. The data from which these schemes are derived incorporate social, financial, and health measures, including property values, council tax bands, company directorships, and more. Such profiling systems have demonstrated their usefulness in many applications across the public and private sectors, and enjoy continued use after more than two decades of real-world testing and examination.

The fundamental misunderstanding which Mr Edwards expresses is in suggesting that the mortality lifestyle groupings are somehow based on Mosaic itself, rather than on mortality. In fact, the groupings of Mosaic types discussed in the paper are based solely on the mortality characteristics of the portfolio in question. The author chooses not to group postcodes, but, instead, to group the more manageable and useful geo-demographic classifications derived from those postcodes.

The link between socio-economic class and longevity is well recognised (by the ONS, amongst others), and would presumably not be disputed. The question is whether geodemographic profiles are a sufficient proxy for socio-economic class to be useful as an explanatory variable when examining mortality differentials in insurance portfolios. In my work, developing the software used to fit the models within the paper, I have seen this demonstrated repeatedly for a wide variety of data sets. These models are further validated by purely empirical and non-parametric methods, such as Kaplan-Meier survival curves. The paper seeks to bring a valuable methodology, based on demonstrably provable techniques, to the attention of the Actuarial Profession.

Professor A. D. Wilkie, C.B.E., F.F.A., F.I.A.: In formula 5 the author gives a log likelihood function (I change the notation slightly):

$$\mathrm{L} = -\sum_{i=1,\mathrm{N}} \mathrm{H}_{xi}(t_i) + \sum_{i=1,\mathrm{N}} d_i \ln \mu_{xi+ti}$$

where (formula 3) the integrated hazard function is:

$$\mathrm{H}_x(t) = \int_0^t \mu_{x+s} ds$$

individuals are numbered $i = 1$ to N, individual i starts at age x_i and ends at duration t_i, if i ends by death $d_i = 1$, else $d_i = 0$.

Now imagine that we split the times lived into small age steps, h, so that i enters in some age interval $m_{1i}h$ such that $m_{1i}h < x_i < (m_{1i} + 1)h$ (with some arbitrary rule to deal with equalities); i ends in some age interval $m_{2i}h$ such that $m_{2i}h < x_i + t_i < (m_{2i} + 1)h$.

We now set up an indicator function R_{im} defined so that: if i enters in period $mh = m_{1i}h$ then $\mathrm{R}_{im} = 0.5$; if i lives through period m $(m_{1i} < m < m_{2i})$ then $\mathrm{R}_{im} = 1$; if i ends in period $mh = m_{2i}h$ then $\mathrm{R}_{im} = 0.5$; and otherwise, i.e. if $m < m_{1i}$ or $m_{2i} < m$, then $\mathrm{R}_{im} = 0$, indicting that i is not observed living in this age step.

We then set up another indicator function A_{im} where $A_{im} = 1$ only if i exits by death in this age step, and $A_{im} = 0$ otherwise. Thus if i does not exit by death, so that $d_i = 0$, then all $A_{im} = 0$ and if i does exit by death, so that $d_i = 1$, then $A_{im} = 1$ in the one period in which i exits, and $A_{im} = 0$ otherwise.

Now we could approximate the integrated hazard if $x = j_1 h$ and $t = j_2 h$ exactly by

$$\mathrm{H}_x(t) = \int_0^t \mu_{x+s} ds \approx \sum_{j=j1,\, j2} \mu_{jh+h/2}$$

that is, we split t into k small steps and use an approximate integration rule very like the trapezium rule, but taking the value of μ at the mid-point of each interval, rather than the mean of the values at the ends of the interval.

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

But usually x_i and $x_i + t_i$ are not exact multiples of h so instead we put $m_{1i}h < x_i < (m_{1i} + 1)h$ and $m_{2i}h < x_i + t_i < (m_{2i} + 1)h$ as above and we count 0.5 for the start and end steps, giving:

$$\mathrm{H}_x(t) \approx 0.5.\mu_{m1h+h/2} + \textstyle\sum_{j=m1i+1,m2i-1} \mu_{jh+h/2} + 0.5.\mu_{mih+h/2}.$$

We can see that this is equal to:

$$\begin{aligned} &= \mathrm{R}_{im2i}.\mu_{m1h+h/2} + \textstyle\sum_{j=m1i+1,m2i-1} \mathrm{R}_{im}.\mu_{jh+h/2} + \mathrm{R}_{im2i}.\mu_{m2h+h/2} \\ &= \textstyle\sum_{j=mLo,mHi} \mathrm{R}_{im}.\mu_{jh+h/2} \end{aligned}$$

where m_{Lo} and m_{Hi} are the highest and lowest age intervals in the whole data set. Remember that $\mathrm{R}_{im} = 0$ if i is not observed in period m.

We now go back to Richards' formula 5:

$$\mathrm{L} = -\textstyle\sum_{i=1,\mathrm{N}} \mathrm{H}_{xi}(t_i) + \sum_{i=1,\mathrm{N}} d_i \ln \mu_{xi+ti}.$$

We replace $\mathrm{H}_{xi}(t_i)$ by $\sum_{j=mLo,mHi} \mathrm{R}_{im}.\mu_{jh+h/2}$ and we replace $d_i \ln \mu_{xi+ti}$ by $\sum_{j=mLo,mHi} \mathrm{A}_{im} \ln \mu_{jh+h/2}$, that is, taking the value of $\ln \mu$ at the mid-point of the step in which i dies (or does not die, in which case $\mathrm{A}_{im} = 0$).

We now have:

$$\mathrm{L} = -\textstyle\sum_{i=1,\mathrm{N}} \sum_{j=mLo,mHi} \mathrm{R}_{im}.\mu_{jh+h/2} + \sum_{i=1,\mathrm{N}} \sum_{j=mLo,mHi} \mathrm{A}_{im} \ln \mu_{jh+h/2}.$$

Now, provided μ_x depends only on age, x, and not on other characteristics of i, we can reverse the order of summation, summing over individuals first and periods next to give:

$$\mathrm{L} = -\textstyle\sum_{j=mLo,mHi} \sum_{i=1,\mathrm{N}} \mathrm{R}_{im}.\mu_{jh+h/2} + \sum_{j=mLo,mHi} \sum_{i=1,\mathrm{N}} \mathrm{A}_{im} \ln \mu_{jh+h/2}.$$

But $\sum_{i=1,\mathrm{N}} \mathrm{R}_{im}$ (which we can call R_m) is simply a count of the time lived ("exposure") in period m by all available lives, counting a half for each period in which the individual enters or exits (including deaths), and $\sum_{i=1,\mathrm{N}} \mathrm{A}_{im}$ (which we can call A_m) is simply a count of all the deaths that have been observed in period m. So we get:

$$\mathrm{L} = \textstyle\sum_{j=mLo,mHi} \{-\mathrm{R}_m.\mu_{jh+h/2} + \mathrm{A}_m \ln \mu_{jh+h/2}\}.$$

In the usual CMI mortality investigations the age step is as long as one year (this may be a bit too long, but it is what we have), so $h = 1$ and m is in years of age. The exposed to risk is calculated by averaging the numbers in force at age m at the start and at the end of the year. This is approximately the same as counting all those that go through the year as 1 each, and all those that enter or exit as a half each; strictly it counts all those that survive as one half in age m and one half in age $m + 1$, but it is intended as an approximation to a count of all the periods at age m going all the way through the year. The deaths are those at age m.

In Forfar, McCutcheon and Wilkie, considering the method of graduation of μ-type rates, the log-likelihood function given (6.3.5) is (omitting reference to the parameters and changing the notation a little):

$$\mathrm{L} = \textstyle\sum_{x=xLo,xHi} \{-\mathrm{R}_x.\mu_{x+1/2} + \mathrm{A}_x \ln \mu_{x+1/2}\}$$

which, if we put $h = 1$ and replace m by x, is just the same formula as Richards'.

However, if μ_x does depends on other characteristics of i as well as age, so would be better denoted by $\mu(x, i)$ then we can't in general reverse the order of summation. We could do so if the individuals fall into a small number of categories (e.g. male/female by smoker/non-smoker), and we add up the categories separately; but if there are many categories (postcodes) or there is a

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

continuous variable (policy size) we have to process each individual separately. We can still use the summation method, using approximate integration, but having an integral function for $H(x, i)$ may make it much quicker. Remember that we if we wish to maximise the log likelihood, and there are many parameters to be optimised, we may need to go through the file very many times. However, this is then just a matter of computer time.

In processing IP recoveries and deaths, where the time primary variable is duration since sickness, usually called z (though age at time of sickness, x, is an important auxiliary variable), we can work with $h =$ one day. We may also use variable sizes of h at different points in the duration, daily at low durations, moving to weeks (7 days), months (28 days) or years (365 days) as duration increases. We also treat entries and exits a little differently. Because of the deferred period, all claims are assumed to start at the beginning of, e.g. the 29th day for a DP4 weeks case; also there are continuations from the previous year, which start at the beginning of 1 January. All exits, even recoveries and deaths, are treated as going out at the end of the last day of claim. Thus there are no half-days. But when we move to weekly or monthly periods, we count the number of days exposed, exactly, so there are fractional weeks or months (but not half-weeks). This is just a feature of how the data is recorded. The same would be true if we had exact days for mortality data, as one might have with individual policy data.

The way Cox models were first suggested is that, for an individual i, with characteristics a vector $\boldsymbol{z_i}$, the hazard rate at time t is of the form:

$$\mu(t, i) = \mu_0(x).\exp(\boldsymbol{b}.\boldsymbol{z_i})$$

where $\mu_0(x)$, the baseline hazard rate, depends only x, and $\boldsymbol{b}.\boldsymbol{z_i}$ is a linear function of z_i. Cox models have since been substantially extended so that $\boldsymbol{b}.\boldsymbol{z_i}$ is replaced by any function $b(z_i, x)$, where the function $b(.)$ can be any combination of parameters, $\boldsymbol{b}$, the characteristics of the individual, z_i, and also time x. Often $b(.)$ has many linear components, but they may include polynomials in x or in any elements of z_i that are numeric rather than categorical.

There are two advantages of this formulation. First, by using the partial likelihood approach one can estimate the parameters of $b(.)$ without estimating the baseline hazard rate, $\mu_0(x)$, at all. Secondly, given $b(.)$, one can construct crude hazard rates for time, split into suitable units, by using adjusted exposures, so that the estimated rates are:

$$\hat{\mu}_0(x) = \mathrm{A}(x)/\sum\nolimits_{i=1,\mathrm{N}} \mathrm{R}_i(x).\exp(-b(z_i, x))$$

where $\mathrm{A}(x)$ are the actual events (deaths, ...) in age interval x, and $\mathrm{R}i(x)$ is the exposure of individual i in age interval x.

Richards uses a different structure, which in some cases coincides with this, but in general is different. He assumes that $\mu(x, i)$ is dependent on x through a set of J basic parameters $\boldsymbol{a}$, where each element of $\boldsymbol{a}$, say a_j ($j = 1, J$) is a function (linear in all his examples) of the characteristics of i, z_i. Thus each $a_j(i) = \sum_{k=1,K} a_{jk}.z_k(i)$, where there are K characteristics (which I suppose could be numeric or categorical), the characteristics for i are $z_k(i)$, $k = 1, K$ and a_{jk} is the parameter for characteristic k in parameter j.

This could be much more general than the Cox method, but seems to require estimation of all parameters at once, rather than in two steps. There might be a lot of parameters. Richards, probably for convenience, uses only functions for $\mu(x, i)$ that can be integrated over x, but the first part of this note shows that the approximate technique could also be used, at the expense of more computer time.

An advantage of a Cox model that is linear in the parameters, combined with a GM(0,s) formula for $\mu_0(x)$, so that it $\mu_0(x) = \exp(\text{polynomial in } x)$, therefore also linear in the parameters, is that this simplifies to a purely linear model for $\ln \mu_0(x)$, which can be fitted by many of the standard GLM programmes. This saves having to write special purpose programmes, but if one is willing and able to write such programmes, there is no need to restrict oneself to such a linear model.

Reference

Forfar, D.O., McCutcheon, J.J. & Wilkie, A.D. (1988). On graduation by mathematical formulae. *Transactions of the Faculty of Actuaries*, **41**, 97-269.

The author subsequently wrote: There was not time for me to respond to every claim made by Mr Edwards during the discussion. I have opted to make a written contribution in order to do all these points justice.

Deduplication is important wherever people have multiple benefit records, otherwise there is overdispersion and the fitted model will be invalid. The greater the number of benefit records per person, the worse the overdispersion will be. Annuity portfolios are particularly in need of deduplication, with an average number of policies per person varying from 1.2 to 1.5. The assumed independence of the events being modelled is a critical foundation for even the most basic of statistical models, and to claim that "independence is not important in practice" risks being negligent. Where deduplication is not possible, other routes to dealing with the problem of multiple policies per person is necessary. For example, the CMI sometimes deals with this in its graduations using an overdispersion parameter. A classic example was given to me by Professor Wilkie, and which is documented in the graduation of the standard table a(90). The value for q_{94} was many times higher than expected, and this was found to be because the information collected was on policies, not on individuals:

> "of 54 identified deaths in the combined pre-1957 and post-1956 data some 41 were of a Mr A and 3 of a Mr B, so there were only 12 separate lives for the 54 policies."
>
> CMI (1976)

From experience, the average number of duplicates tends to rise with age, possibly because of the tendency for longer-lived wealthier people to hold multiple policies. The above example shows that failure to control for duplicates can have unwelcome consequences for a mortality model.

If the comment that "non-linearity is only important for centenarians" is intended to apply to the progression of mortality with age, Figure 4 shows that this is not true: non-linearity is quite obvious after age 90. Furthermore, the failure of a model to follow the non-linear pattern affects, not just the older ages, but it will distort the fit at younger ages as well. For example, the left panel of Figure 4 also shows that the linear assumption results in systematic under-fitting from age 70 onwards, so non-linearity is clearly not just important for centenarians.

Mr Edwards' claims that GLMs are better than survival models was not backed up, nor did he state which type of GLM he has in mind. He might be talking about contingency tables, or the logistic regression of Richards & Jones (2004), or Poisson models with age as a continuous covariate. As it happens, GLMs and survival models are not opposites, and there is overlap between the two; a GLM for Poisson counts is a model for the force of mortality, for example, and is therefore closely related to a survival model for the force of mortality. There are other connections between GLMs and survival models, of course. For example, a Gompertz survival model for the force of mortality is directly equivalent to a binomial GLM for q_x, with a complementary log-log link.

I found Mr Edwards' objections to parametric models perplexing. What kind of GLMs does he fit if they do not have parameters? What is he getting from his GLMs, if not parameter estimates? I dispute the implied claim that GLMs are any less parametric than survival models. The logistic regression used by Richards & Jones (2004) is a GLM, and it contains parameters which are the direct analogues of the parameters in Table 13. When fitting an equivalent mortality model, the GLM parameters in a logistic regression have the same sign and order as those in the survival model, the only difference being that the survival models are more powerful, use all of the data, and tend to produce smaller standard errors around the parameter estimates.

Mr Edwards is correct that splines can be used in GLMs, but this is equally true of survival models. When modelling mortality at the level of the individual, I do not share his belief that

GLMs and survival models 'share flaws', I believe that survival models are simply better. When presented with longitudinal data for individuals, it makes more sense to model the hazard rate μ_x (and thus the survival function ${}_tp_x$), than it does to model the mortality rate q_x. The reasons for this are spelled out in Macdonald (1996a, 1996b), and survival models are clearly the most practical option where there is more than one decrement. Survival models cannot 'over-segment' without this becoming patently obvious in either the AIC or the standard errors on the parameter estimates. Indeed, the automated procedures outlined in the paper do not 'exacerbate flaws'— on the contrary, the use of the AIC as a target ensures that over-segmentation cannot happen.

I fundamentally disagree with Mr Edwards' assertion that smoothness is unimportant. It is probably not putting it too strongly to say that smoothing is what lies behind nearly all of statistics, from calculating a mean to fitting a complex 2-D surface to a mortality array. Comments were also made about so-called 'self-prophesying models', which seem to be a warning about fitting parameters to small sub-groups then re-aggregating. This is to misunderstand the methodology used in this paper, which maps lives to lifestyle group via postcode *before* any model is fitted, not afterwards. In any case, the smallest lifestyle group in the model in Table 13 contains 416,937 life-years of exposure and 13,545 deaths, which I think that most people would agree is quite large enough.

Mr Edwards asserts that GAMs are the way forward. GAMs are indeed flexible, and can provide a good fit in the presence of non-linear relationships. However, this flexibility must be used with great caution. Look at the right panel of Figure 4 — why use numerous B-splines when a simple, two or four-parameter mortality law will fit the data just as well? It is important to evaluate whether the added complexity of GAMs is really required, and this is usually not the case for pensioner mortality. Given a comparable fit of the models, the simpler mortality law approach is preferable to the more complex GAM. For example, a GAM with B-splines centred at five-year age intervals would involve substituting eight spline parameters for each of the six used for age-related mortality in Table 13. This would result in a net increase of 42 parameters ($42 = 8 \times 6 - 6$), and I strongly doubt that the model would be improved by enough to warrant such a large increase in complexity, even allowing for any reduced effective dimension due to smoothing.

Mr Edwards advances no evidence for his assertion that mortality-based postcode groupings are better than Mosaic. I have fitted models based on postcode grouping, models based on Mosaic, and models with both simultaneously. In every case I have found Mosaic (or Acorn) to be a much more powerful predictor of mortality than any other postcode-based grouping. Models using Mosaic or Acorn type are also easier to deal with than the hundreds of thousands of post codes, in addition to passing the significance tests. As to the claim that the use of postcodes is unnecessary and that pension size is all you need, consider a pension of 3,800 p.a. to a male aged 65. This puts the pensioner in the highest of the three pension bands in Table 11. However, the difference between this pensioner being in a higher or lower lifestyle group is 7.1% for an escalating pension. Since a typical pricing margin for an annuity or bulk buy out is around 5%, being able to use postcodes properly is essential for profitable pricing. To put it another way, a company writing annuity business without using postcodes will lose money competing against companies which do. At the time of writing, three of the largest insurers in the U.K. — Legal and General, Norwich Union and Prudential — have publicly stated that they use postcodes for underwriting individual annuities, and I expect that they are all using geodemographic systems.

It was not quite clear what was meant by "the proposed structures are new and unproven". If the charge was levelled at commercial geodemographic profilers, these date back to the late 1980s, and have been used with great success for direct marketing. Indeed, they have been so successful that a public domain geodemographic profiler is even available from the ONS. As for being 'unproven', geodemographics are well established in academic literature. If the charge was levelled at survival models, then proportional hazard models have been around since they were proposed by Cox in 1972. Cox's paper is one of the most cited in mathematical literature, with around 600 citations a year. Formal study of the survival curve has been around since Kaplan &

https://doi.org/10.1017/S1357321700001732 Published online by Cambridge University Press

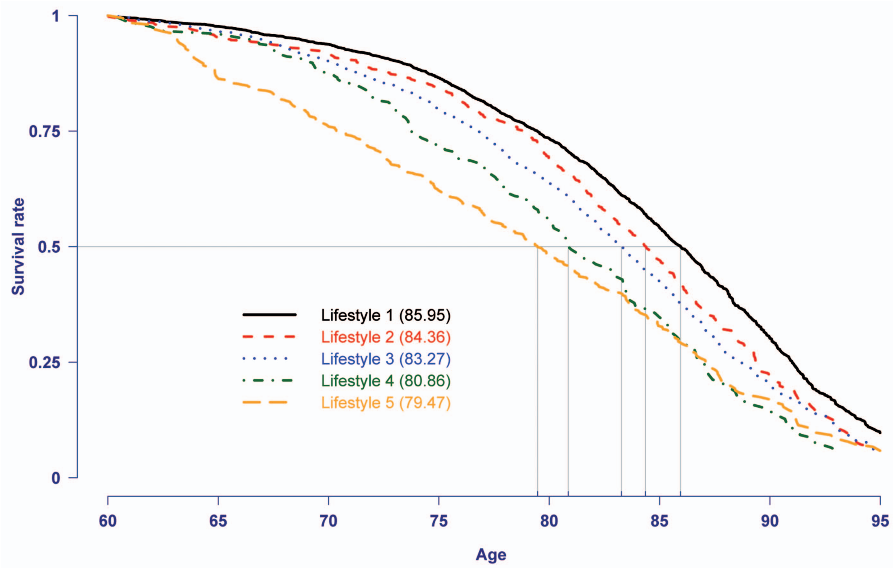


Figure D.3. Empirical survival curves for five lifestyle groups

Meier (1958), which is even more frequently cited than Cox's paper, and, typically, appears in the top five most-cited scientific papers of all time. Closer to the actuarial world, modelling the force of mortality lies behind all recent CMI table graduations, including the '92' Series, the '00' Series and the new SAPS Series 1 tables. This paper shows how existing and proven techniques from other disciplines can be combined and applied to actuarial problems. To paraphrase Nelder (1977), these methods may be accepted, in whole or in part, or rejected entirely, or replaced by something better, as other ideas are put forward. Other actuaries will decide. Meanwhile, Mr Edwards might do worse than consider the possibility that the methods in this paper represent useful techniques in modelling mortality, techniques which it would be worth his while trying to understand.

Mr Sharp expressed worry that we make our models too theoretical. To reassure him, Figure D.3 shows the empirical survival curves for the same annuity portfolio, but with five lifestyle groups instead of the three used in the paper. Following the methods of Kaplan & Meier (1958), these survival curves are formed purely from the experience data, and involve no parameters or model fitting whatsoever. As we can see plainly, the different lifestyle groups based on Mosaic type have very different empirical survival probabilities.

In his written contribution, Professor Steven Haberman mentioned the difficulties of obtaining stable and reliable parameter estimates for non-linear models. In practice we have found only two difficulties, both of which had work-arounds. First, in order to keep parameters well scaled and stable between models, we re-expressed the calendar time y by measuring from 1 January 2000. Thus, our regression coefficient δ applies to $y - 2000$, not y itself. The CMI has done similar offsetting and re-scaling in the past for age, as in CMI (1999), but we prefer the neatness of using age directly without transformation. The second difficulty lies in the limitations of finite precision computer arithmetic. For example, the formulae in Table 5 must be implemented carefully to avoid arithmetic over or under flow at run-time, especially the formulae for the integrated hazard function $H_x(t)$.

References

Continuous Mortality Investigation Bureau (1976). *CMIR 2*, Institute of Actuaries and Faculty of Actuaries, 69.

Continuous Mortality Investigation Bureau (1999). *CMIR 17*, Institute of Actuaries and Faculty of Actuaries, 4.

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

Nelder, J.A. (1977). A reformulation of linear models (with discussion). *Journal of the Royal Statistical Society, Series A*, **140**, 76.