

# What Longevity Predictors Should be Allowed for When Valuing Pension Scheme Liabilities?

---

## Abstract of the Discussion

[Institute of Actuaries, 28 September 2009]

### Contact

Steven Baxter, M.A., F.I.A., E-mail: [steven.baxter@hymans.co.uk](mailto:steven.baxter@hymans.co.uk)

This abstract relates to the following paper:

Madrigal, A.M., Matthews, F.E., Patel, D.D., Gaches, A.T. & Baxter, S.D. What Longevity Predictors Should Be Allowed for When Valuing Pension Scheme Liabilities? *British Actuarial Journal*, doi:10.1017/S1357321711000018.

**The President (Mr N. B. Masters, F.I.A.):** The main item on the agenda this evening is a paper by Steven Baxter, Andrew Gaches, Ana Madrigal, Fiona Matthews and Deven Patel. It is entitled: “What Longevity Predictors Should Be Allowed For When Valuing Pension Scheme Liabilities?”. To begin, may I ask Dr Fiona Matthews to open the presentation?

**Dr F. E. Matthews** (Panel member, opening the presentation): This presentation will cover a brief introduction and description of the available data, a non-statistical description of the methods and a short presentation of some results. I will then hand over to Steven Baxter who will discuss some of the issues and context of the analysis.

The focus of the talk is on the ability to use routine data to detect differences in longevity predictors and to give some examples, rather than a full description of all the differences found within our available data.

It is well known that there are life expectancy differences found in different regions of the country and that much of this difference is due to differences between social groups.

Within the actuarial framework the estimate of baseline longevity used in valuations will show variations between different groups. Currently, baseline longevity is either produced using a single or a group-adjusted estimate.

Pension schemes, by their very nature, provide a rich source of information on life expectancies and deaths. Together with routine data collected as part of the pension requirements, this can be used flexibly to investigate differences in mortality and then to estimate realistic individual or scheme projections for baseline longevity.

The potential to investigate differences within routine data is shown. Different schemes in our database are plotted with men’s life expectancy against women’s (see Figure 2 in paper). It is shown that there is clearly a wide range of average life expectancy by scheme from around 80 years to over 85 years for men, with similar differences for women. The fact that the men’s and women’s rates are so closely related would suggest that the schemes themselves have some similar characteristics. Men and women share similar characteristics within schemes and these are the same factors that reflect differences between the schemes. It is these differences that we wish to investigate.

The key questions we want to address are these: do mortality differentials manifest themselves within pension scheme data? Are routine pension data sufficient for investigating differentials? Do differentials between schemes and within schemes have common causes that can be identified? Are potential models robust enough to enable estimation of differences to be made that are larger than the error estimates produced from the models themselves?

The data used is the Club Vita dataset (as at the time of paper submission), with 91 pension schemes and over 1 million living pensioners and dependants, together with 500,000 historical deaths over a period of up to 15 years.

All regions in the nation are represented, though some schemes will be from focussed geographical locations. However, although there is a range of locations, the data cannot be viewed as truly nationally representative.

The last three years will be used for the purpose of the analysis presented today, that is, the years 2005–2007.

A summary of the 91 schemes in terms of size distribution is shown. There are ten very large schemes, though these are balanced by the larger numbers of smaller schemes. By pooling the schemes more consistent patterns can be investigated. However, due to the number of different schemes, the larger schemes should not dominate the effects seen within the smaller ones.

In terms of individuals, there are over a million men and women pensioners within the analysis, with more male deaths. This raises some issues in the analysis. Testing statistical significance in the presence of such large numbers is problematic as very small effects can produce ‘statistically significant’ results, that is a p-value of less than 0.05 while the actual effect is of negligible importance either clinically (in terms of death) or in actuarial terms (in terms of the impact on life expectancies and annuities). The presentation will focus solely on men from now on, though similar effects can be differentiated within women.

Traditionally, within pension schemes, a fair amount of routine data on potential mortality factors is collected. At retirement it is usual to know the age, sex, health status (whether retired in good or ill-health), final salary at retirement, and pension amount. The collection of occupation type is closely related to the type of scheme and year. This variable, whilst useful for investigating differences, is more problematic. Due to difficulties about the meaning and coding of occupation type, we will not focus on occupation within this discussion. In addition, in the most recent period, all schemes measure postcode. Postcode is a geographical indicator of the location where an individual lives in retirement.

Ideally, we would like to have a measure of the lifestyle of an individual that relates to their mortality profile. Unfortunately, pension schemes do not have available information on smoking status, drinking and other risk profiles known to affect mortality. However, there are aggregated data sources available that do collect this information using market research surveys and routine customer information. This information can be synthesised into producing a lifestyle profile for a particular postcode. These lifestyle groups, whilst not at an individual level, will reflect the general situation around where the individual lives in retirement. They have been found, in other settings, to be related to health outcomes even after adjustment for individual social factors.

We show that information from lifestyle types can be combined with information on mortality to produce five longevity groups for use in the analysis.

And so to the investigation of mortality differentials: there is a well-known difference between the mortality of those who retire in ill-health and non-ill-health. However, the difference appears to change with age (see Figure 3 in paper). The increased mortality is more apparent at younger ages than at older ages, though data does become sparse at extreme ages. The pattern of mortality also appears to be different, so we want to stratify by health status at retirement. Statistical terms reflecting all these differences could be included in the model, but the model would be very complex and not easy to understand.

The question remains as to whether other factors show differentials at retirement in the same way, such that the shape of the change differs with age, or are there constant differences that can be more easily modelled?

Consider Figure 5 in the paper. The first entry shows the well-known and very large impact of age on mortality. The second entry is an example where the ill-health group appear to be of lower mortality than those of the non-ill health group. This is entirely due to different age patterns, and hence, standardising for age, results in higher mortality rates for those in ill-health. Crude and age-adjusted effects suggest higher mortality for those (a) in manual occupations, (b) with lower salary, (c) with lower pension, and (d) in the most deprived lifestyle group. These differences, whilst interesting, are univariate in nature, and it is probable that some of these factors are related to each other. The purpose is not to stratify all the potential groups, but to model them. This means that investigations must be undertaken with a greater number of covariates, because the effects need to be investigated together.

The rest of the presentation shows results for the subgroup of men in good health only, though the same methods apply to each stratification group. There are a number of methods for investigating differences.

The traditional method of producing homogeneous groups based on stratification of known mortality differences is sound but limits the scope of the investigation in that only group covariates can be investigated. For very large data sources the scope is not significantly penalised, but in the smaller schemes stratification will provide unstable estimates.

Statistical modelling will stratify some variables where the pattern of change is very different (such as between the sexes or the ill-health groups) but will then model the other subgroups together. This means that not only can continuous data be included (either age or salary), but also a larger number of variables can be modelled even within smaller pension schemes. Data available are individuals with a known date of birth and, where applicable, date of death. Individuals enter and leave schemes on known dates. For any one year, the status of an individual (alive/dead) is known.

Two potential statistical modelling approaches commend themselves for this investigation. Both fit the data available. One, a logistic regression, models the outcome of death or no death and the covariate profiles associated with them. The other models time to death using a survival analysis framework.

The data available will enable either modelling framework, but the focus of the talk is on differentials within the covariates so further discussion will use the general linear model (GLM) only. Non-parametric survival modelling using Cox Regression did not fit the data satisfactorily hence further investigation is ongoing to formulate the correct parametric shape for the survival model, though the identification of covariates related to mortality is similar in both modelling frameworks.

So for the analysis method: GLM is akin to simple linear regression except there is a transformation of the 0/1 outcome variable to a continuous scale. The logistic link is used which links the probability of dying to the covariates. The last three years of data have been used to have sufficient data to enable us to smooth the effect of age at extreme old age, where data are sparse. Some individuals will be included more than once due to this pooling. However, they will have aged one year and will now be “at risk” of death in the same way as all others of the same age. Sex and health status have been stratified.

One of the most important factors is how the mortality patterns change with age (see Figure 6 in the paper). There is clearly a strongly increasing trend with increasing age when looking at the crude rates. Initially a linear trend was fitted. This works well during the middle age range, but does not fit the apparent curvature at the beginning and end of the age ranges. Further investigation shows that using a quadratic model is better at younger ages, but the oldest ages still do not fit well. The cubic model fits the data much better indicating that the cubic expression for age fits well. It curves, both at the younger ages and at the oldest ages. Of course, in practice, we choose the best model for age using statistical model choice rather than by eye (see Table 6 in the paper). But it is comforting when it also fits by inspection.

Once we feel we have a pattern for age that best fits the data, we want to include other factors. The idea is, initially, to include the factors singularly, but then to combine factors in a sparing way (that is, not check all possible factor combinations). Initially we look at differentials univariately by age.

In the univariate model for lifestyle and age for men in non-ill health at retirement, with age effects modelled using the cubic, the lifestyle group with the highest mortality is very different at all ages from that of the lifestyle group with the lowest mortality (see Figure 7A in the paper). These lifestyle groups reflect not the individual’s situation, but that of their location, though they do reflect the raw differences used during the formation of the groups.

A similar graph, relating salary bands to mortality, shows effects which are very similar to those seen for the lifestyle factors (see Figure 7B), raising the issue of what is the more important predictor for mortality and whether all have to be included within the modelling framework.

This raises the issue of model choice. There are a number of statistical methods that can be used to investigate model choice. As stated earlier, in very large datasets, whether a factor is statistically significant does not necessarily entail mortality differentials that are important. Those factors graphed earlier did seem to suggest differences, but are they the same differences being expressed via two different variables? What is needed from the analysis is the simplest model that accurately reflects the differences seen within the data, but that does not over-engineer the relationships. Two methods have been used within the paper: the AIC and the BIC. Both methods examine the sequential fitting of models and investigate how much improvement is seen (in terms of the log likelihood) compared to how much additional complexity has been added in terms of the model parameters. This means that variables with just two levels do not need to improve the model as much as a more complex variable with lots of levels, because, even without stratification, more unnecessary terms in the model will inflate the errors seen.

For example, the model first fitted with age (up to the cubic terms) produces a value for the log likelihood. Investigating three potential factors – pension, geo-demographic (lifestyle factor) and salary – we can show clear improvement in the model using any of these variables, demonstrated by

the size of both the BIC and AIC results. The BIC penalises parameters more and hence gives a smaller improvement, but can be compared with the other BIC values. Figure 8 in the paper shows that the BIC and the AIC suggest that both lifestyle and salary improve the model by a similar amount, with pension showing less improvement. What is then needed is an investigation as to whether including more factors together further improves the model.

So starting with the model with age and lifestyle factor, we then add in, alternatively, salary and pension. We can show a clear improvement in both models (see Figure 8). Salary is a much better risk factor than pension amount. And it is an improvement to include both salary and lifestyle factors. Whilst there is some overlap in the differentials they are investigating, as the change in BIC and AIC are not the sum of the individual effects, there is sufficient difference to suggest that both factors are important.

In the full model of age, lifestyle factor, salary and pension amount, there is little additional change from the inclusion of pension amount in addition to salary. In fact, the BIC for the full model is higher than that for the model of age, lifestyle, and salary, so adding pension amount makes the model unnecessarily more complex.

Therefore, in summary, the final factors for the model of men in good health would be one incorporating age (in cubic terms), lifestyle group and salary. The mortality differentials within these groups can then be plotted. For instance, we can show the effect of salary given a lifestyle group (see Figure 9 in the paper). The effect on probability of death is clearly different between the salary groups within those of a similar lifestyle. We can show confidence intervals which clearly indicate differences.

Conversely, and from the same model, we can also plot the effects of longevity on a group for a given salary amount. In salary band 2 (£15,000–£22,500, the most common), there are clear differences between the longevity groupings. To fully investigate this using stratification would not have been possible due to the continuous age model fitted. However, the stratification of longevity and salary would have meant 25 different stratification groups for men of good health alone.

All these groups can be shown on one graph (see Figure 10 in the paper). With increasing longevity group and salary differentials, it can be clearly seen that there are interesting patterns. The use of a model has smoothed some of the potential fluctuations due to sparse data as consistent effects are modelled between the two factors. This therefore reduces the error shown around any one of the life expectancy estimates, whilst keeping the patterns evident in the data. Of note here are the mean life expectancy and the error around the mean. The differences in mean value between groups far exceed the error around the estimates themselves.

**Mr S. D. Baxter, F.I.A.:** Dr Matthews described how we can ‘graduate’ post-retirement mortality tables to different populations by simultaneously assessing the impact and importance of different factors.

We have seen how this works in practice for male normal health retirees – but what about other strata, and how can pension actuaries handle issues such as absence of information in the data they receive? I will seek to address some of the practicalities of using the results of the statistical analysis described in our paper.

Firstly, the results of similar analyses for other strata are summarised in Table 12. In all cases, the range of life expectancies arising from using the most predictive factors is material – indeed over the

strata and profiles we see a variation in life expectancies from age 65 of over 10 years. This is equivalent to 40%–60% on annuity values depending on your baseline.

The recommended factors share some common themes in that age and geo-demographics always appear. There are also some subtle differences between the groups in relation to the importance of affluence. For example, within female pensioners we see age, geo-demographics and pension being the important longevity predictors to allow for when valuing liabilities – albeit that the decision between pension and salary is relatively borderline. In contrast we see that for widowers there is little benefit in using any affluence measure, with geo-demographics being the key predictor to use where available.

I would stress though that by ‘recommended’ we mean factors we would suggest allowing for when they are known and available in a scheme’s data. To consider what this means in practice, we look at an example member of a typical pension scheme – John, whose member record is summarised as: a pensioner, who retired on 31st March 1995 from active service and not on grounds of ill health. His current pension is £3,500 p.a., and his final salary at retirement was £18,000 (which is £25,600 in current terms). His postcode is CV8 2AD.

The information on the member record – rating factors in the language of insurance actuaries – can be used to identify the most applicable set of mortality rates amongst those shown by Dr Matthews. In this example the rates would allow for his retirement health, salary revalued to current terms and postcode based geo-demographics.

Our analysis suggests that full-time equivalent salary is a better predictor of longevity than pension for men. However, salary may not always be available. Whilst it is usually held on administration records it can be missing for some individuals. Suppose it is missing for John. In such a scenario our preferred model cannot be used as we no longer have salary data. In this situation pension schemes could use a mortality model which ignored salary, but this could mean ignoring affluence altogether. Instead it would be logical to consider using alternative longevity predictors from those available within the data held on John.

Figure 8 shows that pension offers additional insight to simply using postcode. Put another way, adding pension into our model gives better predictions of mortality and hence longevity. It would therefore be sensible to consider using mortality rates which allow for the level of John’s pension.

A second example covers Jane who is an active member, with a salary of £25,000, 5 years accrued service and an accrued pensions of £2,000. Her postcode is HA9 6RE. Once again care is needed in the application of our analysis.

For example, in Table 12 we identified, for women, pension as our preferred affluence based predictor for post-retirement mortality. However, to use pensions for active members where the pension is only partially accrued runs the risk of using mortality rates which are unnecessarily heavy. Instead, it would make sense to use salary as the affluence measure for active members of both genders, provided the statistical analysis Dr Matthews described suggests that salary adds sufficient predictivity for women to warrant its use. I can confirm that our analysis has found salary is sufficiently predictive over and above purely postcode to include in a model of female mortality when you do not want to use pension.

I described an example of an active member and a current pensioner member deliberately.

A potential benefit of the rating model approach we have used is that it reduces the need to make any subjective assumptions upon which traditional experience investigations often rely. For example, we do not need to assume how the future pensioner population differs, or indeed does not differ, from the current pensioner population.

The importance of this can be illustrated by considering the car manufacturing or printing industries. These industries have undergone very considerable changes in recent decades, and I would suggest that it is not unreasonable to assume that the longevity characteristics of the current workforce may be very different from previous generations.

By understanding what are the most powerful longevity predictors, actuaries have the potential to use these to determine an appropriate baseline mortality assumption for future pensioners that automatically captures any changes which have occurred over time in the profile of the populations being valued.

The analysis in our paper can result in profiling a pension scheme into more groups to which distinct mortality assumptions would be applied than is, perhaps, typically the case at present.

As an example, for male pensioners we have described five salary bands and five lifestyle groups giving rise to 25 different mortality tables. Clearly, a balance needs to be struck between extra precision and extra complexity when we introduce more assumptions.

The statistical methods described by Dr Matthews go some way to achieving this. They penalise for the complexity of adding an additional factor in order to prevent that factor's inclusion where it does not sufficiently explain any residual mortality differences.

However, we accept for some actuaries seeking to use our results that there may be additional practical constraints, for example on computational complexity, which influence the decision as to how many mortality tables to use. We would suggest that the transparency of our approach to trustees and sponsors of pension schemes alike mitigates this argument. Further, our experience in applying this analysis to a variety of schemes is that the refinements made to assumptions are often financially material and so worth the investment in using more tables.

In practice, pension schemes do not only use mortality assumptions for funding. For example, they are embedded in administration calculations such as transfer values and commutation terms. Thus we should consider whether, if actuaries and trustees use lots of different mortality assumptions for funding, these should also be used in actuarial factors? One area where I suggest that this may be particularly relevant is transfer values given the requirement for them to represent an 'expected cost' to the scheme.

Where it is deemed inappropriate to use lots of tables in actuarial factors, an alternative may be to use a 'bottom-up approach'. This involves profiling members in terms of the longevity predictors described in our paper to obtain the best estimate  $q_x$  values for each member. These individual best estimate mortality rates can then be used to construct an appropriate composite, or average mortality table, which truly reflects the characteristics of the underlying population. Such a table would enable actuaries to use truly scheme specific mortality in the occasional situation where multiple tables are not practical.

In conclusion I would summarise our observations as follows:

Firstly, we have observed a wide variety of baseline longevity between schemes and this is unsurprising. Each scheme has a different mix of individuals each of whom is heterogeneous in terms of longevity characteristics.

Secondly, the records of these individuals contain valuable predictive information that can be used to determine baseline mortality. Examples include, but are not limited to, retirement health, gender and affluence – via salary and pension. A wide range of additional factors can also be obtained via the use of postcodes and the commercial data to which postcodes provide access. Our paper shows how these factors can be analysed to identify the significant predictors of mortality and hence longevity.

Thirdly, we saw how postcode based geo-demographic measures are very powerful predictors and, where available, should be taken into account. We showed that the choice of affluence measure is important – for example, salary at retirement or earlier exit revalued to current terms is a more powerful predictor than pension, at least amongst men. The effects of both geo-demographics and affluence attenuate, i.e. decline with age, and do so in relative rather than absolute terms. In particular the impact of these rating factors cannot be captured by a simple age-independent scaling factor.

Finally, we illustrated for male normal health pensioners, that the differences in fitted life expectancy between groups with like characteristics are material, and that these differences are much larger than the parameter uncertainty therein. The authors believe this suggests that actuaries should be explicitly allowing for these characteristics of individuals when valuing pension scheme liabilities.

**Mr M. F. J. Edwards, F.I.A.** (opening the discussion): I would like to congratulate the authors on a paper which will, I hope, become a new reference point for practitioners engaged in the field of multi-factor mortality research. I particularly appreciated the way the paper spanned the whole arc of theoretical background, practical implementation of that theory, and lessons derived.

My main purpose in making some comments tonight was to offer some generally supportive parallels with similar work we have been doing, and I hope these parallels will be of interest.

We have just finished a long piece of work analysing a multi-pension-scheme dataset of similar size to that discussed in this paper, albeit pensioners only rather than also with deferred and active members. The dataset has approximately 3 million person-years of exposure and 100,000 deaths, and some of our findings from this work are very close to those presented in tonight's paper. In particular, as regards the various groupings of pension size, your proposals are very similar to what we found for male and female pensioners.

We also found a similar explanatory power in the main socio-demographic postcode factor we used, explaining a variation of mortality rates of (in multiplicative terms) around 160%. (For reference, I use percentages here in the sense that a figure of 100% would mean no variation, rather than a doubling of mortality.)

On a trivial note, we also found that data duplication, or data de-duplication, was not a material problem in the pension schemes in our dataset, although equivalent work with life office annuity portfolios often does require substantial effort to be applied at the de-duplication phase.



One of the most interesting aspects of this sort of investigation is the ability to quantify the degree to which the perceived lifestyle effect varies by age. Our investigation showed that the explanatory power of the main socio-demographic factor varied substantially over the typical pensioner age range, with explanatory power of over 200% for pensioners in their 60s down to around 160% for the 70s age band and further down to around 130% above the age of 80. So far as I can tell from your graphs that is similar to what you have found.

One aspect of the paper which surprised me slightly was the overall treatment of age, where the authors were able to model mortality across a very wide age range (from 16 years onwards, if I understood correctly) with a cubic polynomial in log-space. We preferred using Bessel splines (combining, typically, five cubics across the age range) to give a closer-fitting shape, although in many such analyses, where the end model is based around some close-fitting standard table for reasons of communicational convenience, the precision that might be obtained with a tailor-made age factor becomes somewhat irrelevant.

The point about final pensionable salary as opposed to pension is an interesting one. Intuitively one would expect salary to be more predictive than pension, but we have generally been constrained by the data available to use pension (which is, fortunately, still highly predictive of mortality), and it would be interesting to understand more about how much of the dataset gave reliable salary information.

The only area where we would diverge from the authors is in their treatment of postcode, and in particular their reliance for postcode clustering on just one off-the-shelf set of socio-demographic clusters. Clearly, any set of such clusters – or almost any set – is likely to lead to some form of observed mortality effect, and the empirical findings presented in the paper and in other work tie in with what we might intuitively expect.

There is a potential problem, though, in the degree of mortality heterogeneity left ‘unprocessed’ in any given postcode cluster, especially where the clusters have been derived with reference to proxies to mortality (e.g. financial status and/or lifestyle) rather than with reference either to mortality or to something similar, for instance some indication of health status or what we might call ‘health style’. The degree to which this residual heterogeneity is a problem of course depends to a large extent on how the results are to be used – if we are looking at the mortality of a scheme in isolation, then it may not be much of a problem, whereas if the results are to be used in any subsequent analysis predicated on postcode as a proxy to mortality, the potential mismatch could be substantially misleading.

The approach we have chosen is partly based on the idea that there is no one best approach. Rather than use one set of off-the-shelf financial-cum-lifestyle postcode clusters, we have found it more powerful to base the analysis on a combination of four factors, two of which are off-the-shelf clusters, one based on financial/lifestyle information and one based on some perception of ‘healthstyle’. As an aside, we found that the effect of our main healthstyle indicator varied substantially by sex and our clustering using that factor had to be done taking into account different sex effects. I am curious how you have done your own clustering with your own lifestyle factor. Was that sex-differentiated?

We combined those off-the-shelf sets of postcode clusters with two mortality-based clusters, one based on the observed postcode effects within the dataset, one based on observed postcode effects

from general population mortality in the age 50+ bracket. Deriving such clusters is impossible at pure postcode level, for obvious data reasons, but using lower level super output area instead gives very good results. Even here there are still problems with low volumes of data, and to get round this we have found it necessary to use some form of credibility-based spatial smoothing algorithm to join the micro-regions in a statistically valid manner.

The main point though is not how good or bad any one particular set of clusters is, as much as to note that combining several postcode clustering mechanisms, and ensuring that some of those clusters have been derived directly from mortality experience rather than only from proxies, can improve the predictive power of such models considerably.

In conclusion, with the exception of our somewhat different thoughts regarding postcode clustering mechanisms, I found that our work generally supports much that the authors report in their paper.

**Mr S. J. Richards, F.F.A.:** I am the author of a similar paper produced last year to which the authors have been kind enough to make several references. I find interesting the comparison between the two papers. The authors use different modelling techniques – GLMs for  $q_x$  instead of survival models – a different data set, defined benefit or occupational scheme pensioners compared with annuitants and a different geo-demographic profiler (Acorn instead of Mosaic). Nevertheless, we both arrive at the pretty much the same conclusions. First, that geo-demographic profiles are very useful in explaining mortality patterns among the subgroups and, second, that a combination of geo-demographic profile and pension size is a more effective assessment of socioeconomic status than either on its own.

As well as confirming the results of Richards (2008), the authors add a new finding: that last-known salary is a more effective rating factor than pension size.

In ¶3.3.1 the authors sensibly tested for the presence of duplicates to ensure the validity of the independence assumption and found no more than 4% of lives had more than one record. This is much lower than the 16.7% in the annuity portfolio used in Richards and Currie (2009), but how did the authors de-duplicate their data? And even if the number of lives with duplicate records is 4% overall, is this evenly spread throughout the sub-groups? Richards and Currie (2009) found that the presence of duplicates was highly correlated with important risk factors, with some sub-groups having essentially no duplicates while other sub-groups had, on average, two policies per person.

I also have some suggestions for the authors. In ¶3.2.2 they state that they have excluded the last month of data to allow for unreported deaths. This is a smaller allowance than I typically use. I typically ignore the six months of experience prior to an extract, but that is working with annuity data sets. Defined benefit (DB) pension schemes may be different. DB schemes may have shorter reporting delays. I did think that one month was perhaps a little on the light side for late reported deaths.

In ¶3.1.8 the authors state they are using the Acorn profiling system. I am wondering if they have modified this or used it off-the-shelf. The reason I ask is because the standard Acorn profiler has at least 150,000 commercial postcodes which have been erroneously assigned to a geo-demographic type, and these need correcting before putting into any kind of mortality model.

I have found that every geo-demographic profiler has some kind of wrinkle and requires some kind of modification for optimal use in mortality modelling, and that each profiler is different in the

changes required. For example, in my work I use a version of Mosaic with several bespoke modifications to improve mortality modelling. This does not, however, in any way invalidate the conclusions to which the authors have come.

In ¶4.5.5, and elsewhere, it sounds like the authors are fitting a separate model for each subgroup or stratum, e.g. separately for each gender and – according to ¶5.1.1 – health status. Why not simply include gender and health status as risk factors in a single, over-arching model? The models fitted in Richards (2008) were unitary models across the whole dataset encompassing every risk factor as a risk factor rather than stratifying by some of these risk factors.

In ¶4.6.3 and Section 5 the authors write about using linear extensions on the logit scale to extrapolate to higher ages. It is not clear to me why this is needed as the logit formula in ¶4.6.4 can be applied at all ages and this can be used without any subjective extrapolation. One reason for the extrapolation might be the instability of the polynomial terms outside the data range, but this would be an argument to replace the polynomial terms rather than add any kind of subjective extrapolation.

In ¶5.1.2 and Table 4 the authors have included a complex polynomial in age. The only reason this seems necessary is because of heterogeneity due to the missing risk factors. Indeed, the need for many of these polynomial terms vanishes in ¶5.1.7 as the authors add further risk factors and the heterogeneity reduces as a result.

On a technical note, in Table 6, I suggest rescaling the parameterisation of the inverse polynomial terms. Some model-fitting algorithms may not be stable when fitting parameters ranging from the scale  $10^{-5}$  (which is your cubic term in age) to  $10^6$  (which is the cubic term in age interacting with lifestyle). That is eleven orders of magnitude in difference across the parameter set. When I am building models I try to structure things to keep the difference in scale, the magnitude between the parameters, to three or fewer orders. That is fairly easily achieved simply by applying some kind of scaling factor to the covariates being used.

**Dr L. M. Pryor, F.I.A.:** Even those of us who cannot necessarily follow the details of the statistics could understand the general message, which reinforces the overall importance of baseline mortality. People are very concerned about mortality improvements, which of course do have a huge effect, but where are you improving from? The base situation is, in my opinion, equally as important.

As the two previous speakers have pointed out, the results given in this paper are not only very convincing but also add to the weight of evidence that we are gradually seeing from a number of sources that factors such as affluence and lifestyle, as indicated by geo-demographic variables, have a significant effect on mortality.

This is important because what you are doing when you are looking at mortality in this way or indeed in any other way, even by using ordinary mortality tables, is essentially modelling mortality. If you are not using these sorts of factors, you have to be very aware that your model has some quite serious limitations.

We were told today that there were differences of 40%–60% in annuity values, which is significant in anyone's terms.

It is therefore important that if you are doing any sort of work with mortality, especially if you are presenting information on the basis on which other people make decisions, such as pension fund trustees in this case, that they understand the limitations of the information they are being asked to use. So if you use straight mortality with just age and sex, say, you would have to accept that although you know that geo-demographic variables have a significant effect no allowance for them has been made. This is going to increase the amount of uncertainty in your results.

Also, you have to think very carefully about the added sophistication that can be provided through this sort of analysis which was not possible even probably ten years ago due to huge leaps in computing power over the past ten or 20 years. You have to balance the sophistication against the need of decision-makers to understand the decisions they are making. In other words, to understand the information they are using for their decisions.

I very much doubt that some of the phrases and explanations I have heard, both from the presenters and from the two previous speakers, could be understood by an average pension trustee, or possibly even by a very well informed pension trustee. So it is important that actuaries think about communicating, and think about communicating not necessarily the technical terms but the basic message of what they are trying to get across.

I think that this will be a very useful paper. Let us hope that actuaries in future take note of the lessons it gives.

The President: Dr Pryor, were those purely personal remarks, or should we read those as at least semi-official remarks in your role as Director of the Board for Actuarial Standards?

**Dr L. M. Pryor, F.I.A.:** I am sure that any remarks I make are influenced by the work that I do.

**Mr T. J. Gordon, F.I.A.:** I should like to start by praising the authors for publishing a paper on longevity. This is an area that is key to the very existence of our profession. Yet, strangely, it is one where we seem to be short on data and expertise. It is reassuring to see that we are all covering the same sorts of issues. In particular, I like the attention paid to parameter uncertainty.

I am going to limit myself to making two specific points on the model set out in the paper and one more general point.

First: weighting. The ultimate purpose of this mortality model in an actuarial context is to value liabilities. The title of the paper itself makes this clear. This means that fitting lives-based models can be misleading because an error in high liability value mortality has a greater impact compared with an error in low liability value mortality.

Accordingly, the requirement to do the statistics, including a weighting reflecting liability value, is inescapable. But the model in the paper does not appear to do this.

Second: “predictivity”. There is no such word, but I think there ought to be. The statistical tests described in the paper are fine as far as they go, but they do not test features of such a model that are likely to be used in practice. We can do this test. For instance, if we have mortality experience of, say, 90 pension schemes in our databank, we can fit the model to the first 89 of them and then test how well the model predicts the known mortality experience of the 90th. We can do this 90 times by

cycling through the schemes and derive a statistic that tests how predictive the model actually is. Here is the rub: our experience is that using the methods described in this paper can result in a model that actually performs poorly at predicting individual scheme mortality.

Finally, as a more general point, I am not sure I understand why actuaries persist in modelling  $q$  rather than  $\mu$ . I know that point is covered in the paper: that is, why are we modelling probability of death over one year compared with the hazard rate? We are dealing with survival, so let us just cut to the chase and use survival models. The maths and coding are simpler and more reliable, and the interpretation is easier. If nothing else, users of  $q$ -based models should be required to test their choice of an additional dimensional parameter that they have introduced.

**Professor R. Macve, FCA, Hon F.I.A.:** I am from the London School of Economics. I am no pensions expert and I may have missed this point in the presentation, when Mr Richards addressed the question of how you use this approach for projecting future mortality.

The point has already been made that we are looking here at ‘baseline mortality’, i.e. as a base for dealing with potential future mortality improvements. However, I was not quite clear about the properties of this ‘baseline’ (and so this is a question rather than an observation). How do you deal with the fact that people in the current sample populations are themselves very different and with differing ‘mortality’ characteristics? At one extreme, the ones who died aged 95 a couple of years ago were born around 1910. Some of the other people retiring on ill-health grounds may not have been born until about 1950. They faced very different ‘life chances’ that even the inflation-adjusted salaries are not going to capture. They capture the change in the cost of living but they do not capture the changes in the standard of living that was available to those people for most of their lives. So your samples of recent deaths are not from a homogeneous ‘baseline’ group. It is a question for a bit more clarification if possible on how you are controlling for that, or would, in making future projections.

**Mr A. G. Sharp, F.F.A.:** I am speaking with my CMI hat on tonight. This paper certainly has helped bring some of the newer techniques much more to the forefront and make them much more accessible to many actuaries.

There are quite a number of common themes, I think, between what we have also been looking at in the CMI with at the paper. Certainly, I was very encouraged by the quantification the authors have given to the predictive powers of pension amount which, as you know, we have featured in CMI publications. We do not, as yet, have salary data to do so. That is one difference between our two surveys.

I really think that the question of data is key to anybody’s investigation into features like this, and the nature of the dataset needs to be very fully explained, whoever is putting forward such information.

The authors have given quite a few statistics about their dataset. I would make just one observation, assuming I looked at the right page on National Statistics. Table 3 gives median salaries. 2008 median national earnings were something like 50% higher than figures given for the median earnings in the survey. I wonder if the authors are able to make any comments about that.

Finally, in terms of predictors I think we do need to do what the authors have suggested, and that is yet more research into the factors affecting mortality improvements coupled with baseline mortality. That is certainly an area for further research. The CMI will be looking at that. I hope others will, too.

**Mr O. J. Lockwood, F.I.A.:** The paper proposes the use of mortality assumptions in pension scheme valuations that vary by salary or pension amount and by geo-demographic profile as well as by the well-established variables of age, gender and health status at retirement. I agree with the authors that it is important to use mortality assumptions which vary with these variables, and vary in a consistent way. Coming from the life insurance industry rather than the pensions industry, I have first-hand experience of the significant shifts over time that can occur in the membership profile of pension schemes. Starting from a position where the majority of the active members are clerical staff, the industry has moved to a position where the customer service functions have largely been outsourced and the majority of active members are professionals. Suddenly we have a very different profile of salaries and geo-demographic characteristics. Most of these changes have happened within the six years I have been involved with the life insurance industry, demonstrating how rapidly such shifts in membership profile can occur. I agree with the authors that there is a need to allow for these changes in a more systematic manner than by applying a simple adjustment factor to the assumed mortality of current pensioners to obtain assumptions deemed to be appropriate for the future mortality of the current actives. At the same time, I agree that fitting a completely separate mortality model for each class of member is not appropriate as it will inevitably lead to inconsistencies between the mortality rates for the different classes.

I would like to raise some more detailed points. I have not set out to discuss the relative merits of GLMs and survival models, as these have been debated on previous occasions. I would, however, note that whatever type of model we choose to use, it is necessary to have a clear statement of the assumptions made. Section 4.3.1 states that the estimation of initial mortality rates includes part years of exposure. This requires an appropriate assumption to be made about how mortality varies with age over the age range  $x$  to  $x + 1$ , a piece of information lost by focusing solely on  $q_x$ . If the authors have calculated their initial exposed to risk figures in the way described in the course notes when I took the Survival Models exam, then they have used the so-called Balducci assumption. The advantage of this assumption is that it gives a particularly simple formula for the initial exposed to risk. The disadvantage is that it results in the force of mortality for a life just before their 71<sup>st</sup> birthday, for example, being lower than that for a life just after their 70<sup>th</sup> birthday. So the mortality curve has jumps at each integer age. This is a counter-intuitive assumption.

Section 4.3.4 states that GLM modelling assumes the observations are independent and that the independence assumption is violated here because the same life may be exposed to risk for more than one year. It is stated that it is possible to demonstrate the impact of this violation is not significant using the methodology of generalised linear mixed models. I should have liked to see a little more detail on this in the paper, as there may be other circumstances where the impact is significant.

I would emphasise that I do not believe concerns over such detailed technical issues should deter pension schemes or life insurance companies from using longevity assumptions which vary by geo-demographics and by salary or pension amount, along the lines suggested in the paper. I consider this paper will help actuaries to recommend mortality assumptions to their clients which are much more appropriate to the lives who will be in receipt of the pensions or annuities in the future.

**Mr L. Churchill (visitor):** Much of the discussion so far has been about methodology. I wanted to ask some questions about impact. To the non-actuary in the hall it seems to me that your findings are capable of being very significant, and that a number of pension schemes may be under-valuing their technical provisions reasonably significantly, depending on the particular demography that they have within the scheme.

Conversely, of course, it is possible that some may be over-valuing the technical provisions at the moment.

I just wonder whether, perhaps, some feel for the impact of this can somehow be given from the schemes that you have looked at within your data samples.

**Mr P. N. Thornton, F.I.A.:** I, too, welcome this paper, for taking us a bit further towards understanding these complex issues of mortality. I want to home in on a particular issue around postcodes. As the paper says, the mapping has been developed for marketing purposes and then found to be quite useful for predicting mortality. I am just wondering what experience other members might have had of that.

In routine valuations of pension schemes you have a chance to sort things out next time round. To the extent to which you over or under estimate mortality, it comes out in the wash in the next valuation. There is a lot of interest currently in closed mature pension schemes transferring liabilities, particularly for current pensioners, into the insurance market, or otherwise into the capital market. That is a once and for all transaction. You do not get a second chance to come back and to think about it again.

In a recent case in which I was involved, the scheme actuaries had done a lot of work to analyse in considerable depth what they thought should be the correct price. The preferred provider had also done a lot of work on pricing. There were many actuaries involved. They analysed the recent experience of deaths and why there was a significant difference between the two prices in percentage terms.

Some of the issues were to do with correct interpretation of the benefits structure. One of them, however, was to do with the base mortality assumption, and my reading of it, as an interested party, but not the Scheme Actuary, was that the insurance provider was, perhaps quite reasonably, using postcode analysis to reserve for and price its mortality book, whereas the Scheme Actuary had been estimating the mortality rates for the particular scheme in the light of its own experience.

That leads me to question whether we are putting too much reliance on postcodes for predicting mortality in such circumstances. In the case of that particular pension scheme, it seemed to me entirely possible that the subset of people from the particular industry living within all of the postcodes actually did experience heavier mortality than the other people living in those postcodes.

I think there is a little bit of a trap here. I will be interested in the authors' comments on whether they have had any other experience of people stumbling on this issue.

**Mrs S. Bridgeland, F.I.A.:** I was not planning on speaking, but I am moving house on Friday and my postcode will change as a result. So my question, with my trustee hat on, is: how far do the authors think we should go in using member-specific mortality assumptions? Should transfer value calculations depend on postcodes, too?

I am thinking about the potential disputes from members: those whose estranged other halves have moved to another part of the country and may believe they have a better socio-economic group as a result, and so their bit of the pension should be valued in a different way; to a senior executive who chooses which address to use when he is having his transfer value calculated; or even how

higher rate taxpayers should assess the value of their pension accrual for tax purposes. So, how far should we go?

**Mr P. J. Lee, F.I.A.:** Just a quick comment on the computation side. I can see that for asset liability modelling purposes it might be a significant factor, but my experience is that for valuation calculations, even for some very large schemes, if you split the mortality groups down by a factor of 25, you might find that you are multiplying something like two or three minutes by 25 for valuation time, which is not necessarily a real problem. That would be for a majority of schemes. Unless one is talking about very complex public sector schemes which might be different, but for most private sector schemes, even with thousands of members, the valuation run time need not be more than one hour or two, even with 25 fold multiplication of the number of mortality groups.

I agree that for asset and liability modelling purposes you do not want to multiply ALM calculations by 25.

**Mr C. G. Lewin, F.I.A.:** I just wanted to comment on an area that the paper deliberately does not cover, which is the way you allow for moving forward from this base to what the mortality will actually be in the future.

It seems to me that what one needs to think about are the underlying causes of the differences which exist today and the differences which will exist in, say, 20 years' time.

It has occurred to me, for example, on postcode differences, that, as the medical profession becomes ever more conscious of the variations in mortality from one part of the country to another, health campaigns will be directed towards the worst areas, which may mean that gradually they will have bigger improvements than the areas which are good at the moment.

One other little point occurred to me: the difference between pension and salary as predictors of mortality today. It seems to me that in any pension scheme a person might have quite a small pension because they have changed jobs, yet their total pension, if you were to look at all pension schemes, would be quite high. That is one of the reasons why salary is a better prediction of affluence than pension.

**Mr S. J. Jarvis, F.I.A.:** I have one comment on the paper, picking up on a comment that has been made already a few times this evening. This relates to the question of revaluing members' salaries to date. As I was reading the paper, it occurred to me that revaluing with the Retail Prices Index (RPI) was a strange thing to do. What I thought I would have done, perhaps naively, would have used been to revalue with national average earnings (NAE). But it now seems to me that using either RPI or NAE may not work well – what you really want to do is to assess where somebody falls within the national distribution of salary when they retire. The change in pay for those in the bottom 10% of this distribution may be quite different from the change for those in the top 10%.

So perhaps one could just look at where somebody falls within that distribution at retirement, and do away with the revaluation entirely. This sounds challenging but there is in fact quite a lot of data available on the distribution of pay across the population. The Government has published data on salary deciles across the population through time. This is often used to track progress on income inequality for example. I believe that you can download this data from the websites of the Office for National Statistics, or the Institute for Fiscal Studies.



**Mr M. A. Pomery, F.I.A.:** I just wanted to mention that the International Actuarial Association has had, for a year or two now, a task force on mortality. It has collected together a group of interested people from all around the world. They are planning to do a series of presentations at the International Congress in Cape Town next March. I would urge members of the U.K. Actuarial Profession who are involved in mortality to contribute fully to those discussions.

My impression, from my involvement in the international scene over the last five years, is that we do have something of a lead in mortality matters in the U.K., and we have much to contribute internationally. I have no doubt if you go there you will also learn a lot from people from other countries.

The President: Thank you all very much for an excellent discussion.

I want to dwell on one thing which occurred to me as I listened to this. It is to reiterate the comments that Dr Pryor made. That is, that we now have some really powerful predictors of the base case mortality. Some of the new actuarial standards that are either now in force, or will soon be in force, will mean that we do have to take cognisance of that in a way that we might not have done before. I think it is important that the profession, building on the authors' work and some of the other speakers tonight, does look again at what constitutes good practice in this area. We need to think hard about these findings and the importance of this paper and the papers by the other authors, which we have heard about tonight and which are referenced.

It also worries me, in exactly the same way as Mrs Bridgeland indicated, that your postcode in some way might influence your transfer value.

I will now ask Nigel Bodie, our closer, to summarise the proceedings.

**Mr N. D. V. Bodie, F.I.A.** (closing the discussion): I will start with a note of an item at a pension scheme trustee's meeting which I was asked to discuss. The heading of the item was "Longevity: Good News for People; Bad News for Actuaries."

I objected strongly to this. Contrary to popular opinion, actuaries are people, too. We are only too pleased by the recent improvements in longevity. It is just that our pleasure is tempered by the understanding of the financial costs of increasing lifespans.

I studied economics in the early 1970s, and the left-wing writers of the day were gleefully predicting the death of capitalism. By the end of the 1970s company earnings and share dividends were, in real terms, at all-time lows, and over the previous five years inflation had swung between 26% and 8% and was on its way back up to almost 22%.

Pensions actuaries in the early 1980s were far more worried about inflation and investment returns than they were about mortality. It is just depressing to realise that I can speak from first-hand experience of those times.

If we were to time-warp our 1980 actuarial audience forward to the present time, they would be astounded at three things. First, the number of papers discussing the issue of mortality. Second, the depth and quality of the research that has been reported in this room and elsewhere. I am delighted to say that this paper is no exception to the high quality of the work that has been done.

Third, I think they would be surprised at the remarkable disparities in mortality experience between different groups of members even within individual pension schemes.

A paper presented here in the early 1990s by Messrs Thornton and Wilson on the funding of pension schemes suggested that there might perhaps be a two-year age difference between the life expectancies of people in white collar and blue collar employment. We would now say “and the rest”, but at the time there was some discussion as to whether it was even appropriate to make differential assumptions about the various groups that one might find within pension schemes. We have only to look at today’s paper to see how important it was to start establishing that differential.

The outside world has, I think, come to terms with the fact that actuaries are not capable of predicting the actual date of death of every individual in their pension scheme.

I also believe that people would accept that actuaries have not focussed with quite the current intensity on the trends in mortality but very few other people have either. Demographers in the 1980s generally did not predict the subsequent dramatic changes in post-retirement life expectancy that have occurred. Rates of improvement are very hard to pin down and are the subject of widespread debate.

But, as a profession, we would have little excuse to offer if we failed to carry out detailed and accurate analysis of current rates of mortality, and that is the area where this paper will be of considerable value to actuaries. I am pleased that the authors have indicated already that there are other aspects which they are looking at on the basis of the data that they have and that further analysis is already being carried out. I for one certainly wish to encourage them in doing this. I will also add that I think that they are particularly fortunate to have data regarding salary at exit. As other speakers have said, this is important information that we are often unable to obtain, and I think that this is a particularly important area on which to focus future efforts.

I have one small quibble. In Figure 11 there is a chart of the mortality rates that emerge from the analysis. There are discontinuities in certain places, particularly at the point where the high ages stop in the graduation and when the smoothing into the long-term assumptions occurs. It seems to me that these discontinuities are an artefact of the method of graduation rather than a reflection of the realities of the data. I would like to see the authors consider more carefully the process by which the mortality rates are smoothed from the ends of the distribution that has actually been graduated into the long-term assumptions that they have used.

Turning back to the discussion, I sympathise with Dr Pryor in that it is going to be difficult to define our mortality assumptions. The CMI’s model for future improvements does have the default series of settings but it can also be altered quite substantially. We are at risk of having a very large number of mortality assumptions for current mortality as well. In terms of the lay user, I think that we can focus in on certain aspects or explanatory statistics that encapsulate all these assumptions:

1. What is the expectation of life at normal retirement age?
2. What do we think it will be in ten years’ time?
3. What do we think the expectation of life is at age 80?
4. How much is current mortality and how much is assumed future improvement?

With a few simple parameters like that, I think many trustees and other users of these statistics would be able to come to terms with the significance of what we are talking about.

Mr Thornton commented on differential mortality within postcodes. I have certainly seen this in the work that I have been doing. Mr Lockwood commented on the possibility of there being significant temporal shifts in the populations of schemes. I can again confirm that I have certainly seen this in my own experience of dealing with mortality questions.

I would like to thank the authors for an excellent contribution to the current mortality debate and look forward to many more papers to come.

**Mr Gaches** (responding): Many thanks to all of the contributors, both those from the floor, and also to Mr Bodie for his comments. I will attempt to respond to as many points as possible. There will inevitably be some to which I will not do justice this evening. However, I know that my co-authors and I are more than happy to pick up any such points outside this meeting or on another date.

I am going to comment briefly on a few of the statistical issues raised to begin with before moving on to some of the wider issues which have been raised, alluded to or made previously.

Before I do that, I might just try to cover a few quick points.

Mr Gordon rightly made a point around the danger of using a lives-based analysis if there was no method of incorporating a liability weighting into the resulting application of that analysis. That is the reason why traditionally amounts-based analysis has been very popular. The approach which we have taken here is a lives-based analysis, but because it results in individual baseline assumptions which can be applied to individual members, you end up applying a high life expectancy assumption to an individual with a large liability. So in the application of results you can naturally get liability weighting coming through.

A number of individuals, Professor Macve, Mr Lewin and the Mr Bodie, I think, made comments around future improvements. For such an important question I am going to say very little, simply to acknowledge that it is a huge issue. We have deliberately not sought to address that in this paper. We fully accept that it is a big issue which needs to be dealt with, but we come from the same point of view of those who have said that we should not be distracted by the size of the issue of future improvements and shy away from doing what we can in an objective way about the baseline, where we can bring in much more objectivity.

Turning to a few of the statistical comments, Mr Gordon and Mr Lockwood both raised the question of the relative merits of survival models and GLMs. Ultimately, survival models and GLMs are simply two different statistical approaches to the same problem. Both methods in our view are equally adequate for answering the question posed in the title of our paper, and we can very effectively model the kind of data we have here using either of those two approaches. In the case study of normal health pensioners that Dr Matthews described, we drew the same conclusions as to the key predictors of longevity under both methods, but we have presented only one here.

Akin to Mr Richards' papers on the subject, we have also observed the importance of the parametric form if any form of survival modelling is used, and it is certainly very valid to be using survival models in these kinds of applications.

On the question of duplicates, Mr Edwards commented about the effects, as did Mr Richards. To the extent that a member has pensions in multiple pension schemes or multiple records within

a scheme, we do have a non-independence of observations, and it is natural to ask if duplicates have influenced or biased the analysis.

For example, if the duplicates were biased between lives and deaths, then we may have an issue. I hope you will be relieved to know that we have looked into the individual records for whom a matchable identifier, such as a national insurance number, occurs. It occurs in around 89% of records and within that group there is a duplication level of around 4%, and we are not seeing any obvious bias coming through from on this account.

The key question, then, is what level of duplication would be needed before the conclusions of this paper could change? We have simulated the effect of various different levels of duplication using stochastic processes. They suggest with the data in question, the impact of the 4% duplication could be of the order of 0.1 years on life expectancy at age 65. So it is really just confirming that the impact of the duplication we are seeing is not terribly significant relative to other issues.

Mr Lockwood raised the question of using multiple years of observation. In particular, there could be a concern that an individual who is alive for more than one of the three years forms a non-independent observation. As Dr Matthews mentioned, in a model based on individual years no one person contributes to the same estimate more than once. This alone should give some reassurance. Additional analysis, adjusting for year of exposure in the age-only model, the mixed linear model, confirmed that any impact of the non-independence over calendar years was negligible.

We would stress that there are practical advantages in the GLM framework of using multiple calendar years as it has the substantial advantage of smoothing out the effects of mild and harsh summers and winters which would be undesirable to embed within the baseline mortality.

Moving, then, to an issue that Mr Edwards raised first of all, in relation to whether postcode clustering would result in a model that could be problematic for onwards analysis. Mr Gordon also raised the question of whether such models would work in individual schemes. My description here of some of the additional checks we have done may sound rather similar to what he suggested. It certainly is an important point.

Ultimately, one of the applications of the methodology underlying the analysis presented in our paper, is to provide mortality rates which can be used to reliably set baseline mortality rates for individual schemes.

One of the considerations is whether the resultant rates are appropriate to use for individual schemes; either to those schemes whose data has contributed to the analysis or for schemes where the membership data has not contributed to the analysis.

The usual statistical method applied in such situations is cross-validation, and that is part of the testing we have carried out. The idea of cross-validation is to verify the “predictivity” by partitioning the data, say, into 20 components. Models are then fitted to 19 of those components combined and the “predictivity” tested on the 20th. That approach also provides an indication of the sampling error of the resultant estimates.

We have applied those kinds of additional checks. To date those methods suggested that our methods do give the degree of robustness for which we are looking.

Picking up on the issue of data, and the cleanliness of data, raised by various contributors, including Mr Richards and Mr Edwards, it does surprise many, and this includes pension managers, to see how much data they hold, some of which they are unaware. Salary at exit is one of those examples.

Clean data is absolutely fundamental to these analyses. For example, the postcode based lifestyle effect, is extremely localised with significant effects down to the full six digit postcode. So it is important that postcodes are accurate to the full six digits.

Just to put that into perspective, we have found around a 10% correctable error rate with postcodes resulting in over 300,000 corrected postcodes being passed back to administrators over the last year. So for anyone using postcode models one of the key things that it is important to do is validation of the postcodes before reliance is placed on them.

Mrs Bridgeland also talked about postcodes. One of the things she raised alludes to one of the potential perceived weaknesses of postcodes as a covariate. If an individual moves house his postcode changes. But a key point is while an individual's postcode may change relatively frequently, their lifestyle does not. For example, if an individual takes foreign holidays, if they read the Telegraph, if they enjoy a glass of wine with dinner, that is unlikely to change if they moved house. So while the postcode may change if they move house, the associated lifestyle typically does not.

She also raised the question of whether individual life expectancies should be used for factor purposes, transfer values, and the like. It is a difficult question to answer. It is clearly a relevant question for scheme actuaries and for trustees to consider. I am not going to express a particular view on this, I am afraid.

Dr Pryor raised the issue of the need for users of actuarial information to understand it and of the need for clear communication. We absolutely agree. What we have found is trustees understand profiling-based approaches to longevity. Trustees understand the differences between members. They recognise the different lifestyles in different postcodes. They see the different levels of salary. So they "get" the concepts behind it.

In many cases they find the concept of setting a longevity assumption by profiling, much easier to understand than making assumptions to a standard actuarial table, which, after all, can come across as just about numbers.

Some of the techniques underlying the approach we are using could be viewed as complex. The principles are not. A common response that we see to the profiling approach is: why have not more actuaries been using this before?

I fully support Mr Bodie's comments around the use of simple statistics to aid the communication of the strength of base assumptions, and also to provide an illustration of the allowance for future improvements.

Mr Thornton raised the issue of using a scheme's own experience, pointing to that being a useful source of evidence in the larger schemes. While useful information certainly can be obtained by analysing the experience of large schemes in isolation, the bar is pretty high in terms of the size of data needed for the results to be sufficiently credible.

There can also be some inherent weaknesses in relying on a scheme's own experience, alluded to both by Mr Bodie and by Mr Lockwood. Is the profile of non-pensioners the same as that of pensioners? Indeed, is the profile of today's pensioners the same as the profile of pensioners over the recent period being studied? If not, the assumptions derived from any pensioner experience study may not be appropriate as a starting point. If the client wishes to understand a subsection of the scheme, does the scheme's own experience help with that?

We would expect a profiling approach to be beneficial for most schemes, even the largest, where it certainly would be an additional rather than a sole source of information.

Dr Pryor also commented on the significance of some of these risk factors. Others in the past have asked the question of whether using several risk factors is perhaps over-engineering. Mr Churchill, in a theme I will pick up in a moment, asked about the impact on pension schemes of using this approach.

It is perhaps worth asking ourselves how big an issue longevity is. Based on there being over £1 trillion of funded DB pension liability in the U.K., being just one year out has around a £30 billion impact. On the macro scale, it is a big issue.

But we can also consider individual schemes. In the work we have done we have seen impacts of up to £80 million for individual schemes.

In some cases these have been increases in liability; in some cases decreases. The schemes in question do not view using multiple factors to be over-engineering. They understand the approach and they are pleased that their scheme actuaries, whoever they are, have been able to provide them with refinement to a longevity assumption that was not available before.

A final question some have asked in the past. People have said, "At the moment we are seeing so many other issues for pension schemes. We have seen the impact of equity markets. There is the issue of future improvement. There are a whole host of other risks. Surely base longevity assumptions just pale into insignificance?"

We fully accept the schemes are subject to significant other risks, but that should not distract us from adopting an objective approach to longevity, where possible.

In my view, we would be doing ourselves and our clients a disservice if we fail to do what we can in this area. Actuaries have been criticised in the past for being slow to move on mortality. "But there were other issues to address" would seem to be a weak defence if we fail to act now and longevity bites us again in the future.

Having said that, I am hugely encouraged by all the comments that have been made this evening. They seem to show that longevity is clearly an issue that is being taken with the utmost seriousness by actuaries in general.

**The President:** It remains only for me to express my own thanks and indeed, I am sure, the thanks of all of us here to the authors for a splendid paper and an excellent discussion. I should like to thank Mr Bodie for his closing remarks and all those who have contributed to making this an excellent discussion this evening. Thank you all.

## Written Contribution

Mr S. J. Richards, F.F.A., subsequently wrote: During the debate a question was raised over the handling of fractional years of exposure. In ¶4.3.1 the authors write that they “weight the contribution of each of the membership records according to its exposure to risk in a year, counting as a full observation when the exposed to risk is equal to one.” This suggests that the authors are using a weighted log-likelihood,  $l^w$ , as follows:

$$l^w = w(1 - d) \log(1 - q_x) + wd \log q_x$$

where  $q_x$  denotes the one-year probability of death at age  $x$ ,  $d$  is an indicator variable taking the value 0 upon survival and 1 upon death and  $w$  is the exposure. The description in ¶4.3.1 suggests that  $w$  takes the value 1 where a complete year of exposure applies and some smaller positive value when a complete year is not possible.

If the authors are doing this then they are creating a bias in their model, even if the weights are not quite as described above. They are taking observations which by definition have artificially low observed mortality by virtue of being only observed for part of a year. They are then including them in a model with a smaller weight than proper observations for a complete year (or potential year). All this achieves is reducing the *influence* of the observations with incomplete years, it does not allow for the incompleteness of the years themselves. The effect would be to produce lower mortality rates than should be the case, i.e. the resulting estimates would be biased estimates of the true underlying mortality rates. The extent of any bias would be directly linked to the relative proportions of part-year and whole-year observations. Another point is that the weight  $w$  should not be applied when  $d = 1$ , as a lower number of expected deaths given fractional exposures should take care of itself without weighting.

To illustrate this, consider the extreme situation where the only data you have is of partial years of exposure. Imagine an annualised mortality rate of 0.2 amongst a group of 100 identical individuals. In a complete year we would therefore expect 20 deaths. If we only had half a year’s exposure, then there would be on average only 10 deaths:  $20 * 0.5 = 10$ , assuming a uniform distribution of deaths (UDD) throughout the year. However, if records are weighted according to exposure, the estimated annual mortality rate would be  $10 * 0.5 / 100 * 0.5 = 0.1$ . In their admirable aim of trying to include fractional years of exposure, the authors may have inadvertently created a material source of bias in their model.

There are a number of ways to properly allow for fractional exposure in  $q_x$  models, including UDD, constant force of mortality or the Balducci assumption. However, none of these fits easily into the chosen GLM framework. The cleanest approach is simply to use survival models as they automatically allow for fractional years of exposure. This is because they model the time to an event, not the number of events occurring.

**The Authors responded:** The authors thank Mr Richards for his additional written contribution and would make the following observations.

The weighted log-likelihood function is not of the form which Mr Richards has inferred. To elaborate on ¶4.3.1 it is the contribution of each of the survivor membership record which is

weighted according to its exposure to risk in a year. Under this approach the form of the log-likelihood assumed by our GLM may be written in the form:

$$\sum_i \left[ d_i \log \left( \frac{q_i}{1-q_i} \right) + ETR_i \log(1-q_i) \right]$$

Where  $i$  indexes individuals;  $q_i$  represents the probability of death for individual  $i$  given their individual characteristics and age;  $d_i$  is a 0/1 variable indicating whether the individual died during the period ( $d_i = 1$ ) or survived ( $d_i = 0$ ), and  $ETR_i$  is the 'exposure to risk' measure which for each individual is: the usual part year 'initial exposed to risk' calculation for those who do not die during the period and, 1 for individuals who die during the period of investigation.

Using this approach, any impact of partial exposures would be significantly smaller than the example provided by Stephen Richards would suggest. To illustrate this consider the scenario where we have 1000 people, 100 of whom are partially exposed for half a year and an underlying  $q$  value as high as 0.20. Here our method would give an estimated  $q$  value of 0.19895 c.f. 0.20. We would also note that the dataset used has very modest levels of partial exposures; furthermore, where these exist, they typically occur at young ages where the much lower probabilities of death mean that the impact is smaller still.