Continuous Mortality Investigation

Mortality sub-committee

Working Paper 3

Projecting future mortality: A discussion paper

March 2004

Continuous Mortality Investigation

Mortality sub-committee

Working Paper 3

Projecting future mortality: A discussion paper

1. Introduction

1.1 *CMI Working Papers No.1 and No.2*

For almost 50 years, the CMIB has made projections of future improvements in mortality, so that these could be taken into account in the pricing and valuation of pension and annuity business. With the benefit of hindsight, the improvements seen in practice have quite consistently exceeded the projected improvements. As a result, insurers have, from time to time, had to allocate more capital to support their in-force annuity business, with adverse effects on free reserves and profitability.

The methodology used in the recent past by the CMIB has been based on extrapolation formulae derived from studies of past trends in mortality, not confined to the CMIB's own experience (see Section 1.3). In recent years, however, there has been much discussion in both the biological and demographical literature concerning the mechanisms that determine ageing and longevity, and empirical features such as cohort effects. The Mortality Sub-Committee decided, therefore, that these features should be reviewed and, if necessary, taken into account in any future projections for insurance use.

In December 2002, the Sub-Committee published Working Papers No.1 and No.2. The first of these described preliminary investigations into the existence of a cohort effect in the CMIB data, meaning that the annual rates at which rates of mortality had been falling were consistently high for the cohort born in a few years around 1926, relative to the rates in respect of other cohorts. The Sub-Committee concluded that such an effect could be discerned on a purely empirical basis in the very extensive Assured Lives' data, and also in the less extensive Pensioners' data. This was broadly consistent with conclusions reached by the Government Actuary's Department (GAD) in respect of mortality in England and Wales (GAD, 2001). The review carried out by the GAD has been very valuable in the preparation of this Working Paper.

The Sub-Committee made three possible projections of future mortality incorporating the tailing-off of the cohort effect over a medium, a shorter and a longer time. These were described only as interim projections, until a more thorough investigation could be undertaken over a longer time period. This Working Paper reports the first results of that investigation.

One feature of the interim cohort projections was new, namely the presentation of a range of projections, with no recommendation as to which might be most suitable for any particular purpose or by any particular insurer. This recognised, though in a wholly *ad hoc* way, that any projection of future mortality is uncertain. The current trend in insurance regulation is towards risk management based on stochastic models of risk, and the Sub-Committee felt that it was necessary to address this source of uncertainty explicitly. This will be a key theme of this and subsequent Working Papers.

This Working Paper has been prepared for the Mortality sub-committee of the CMIB by a Working Party consisting of Angus Macdonald, Adrian Gallop, Keith Miller, Stephen Richards, Rajeev Shah and Richard Willets. It has been approved by the sub-committee.

1.2 *The CMIB/GAD Seminar of 6 October 2003*

The CMIB and GAD held a joint seminar in Edinburgh on 6 October 2003 ('the seminar'), to discuss the views and approaches of demographers, statisticians and gerontologists, all of whom have a strong professional interest in the projection of future mortality and its underlying causes. Its three themes were:

(a) Projecting aggregate mortality *versus* modelling individual causes.
(b) Methodology of projection and statistical methods.
(c) Limits on human lifespan and molecular effects on ageing.

The seminar is described in more detail in Appendix A.

1.3 *Current CMIB Methodology*

This section describes the projection basis recommended in CMI Report No.10 (1990), and first used with the "80" Series tables. Let $q_{x,t}$ be the rate of mortality applicable at age x in calendar year t. The base table is the cross-sectional or secular table regarded as being appropriate in calendar year 0, that is $q_{x,0}$. The *reduction factor* $RF(x,t)$ is defined by:

$$RF(x,t) = \frac{q_{x,t}}{q_{x,0}} \tag{1}$$

and the projection model was:

$$RF(x,t) = \alpha(x) + (1 - \alpha(x))0.4^{t/20}. \tag{2}$$

The parameters of the model may be interpreted as follows.

(a) $\alpha(x)q_{x,0}$ is the 'ultimate' rate of mortality at age x, in the distant (in fact infinite) future. The following values were suggested:

$$\alpha(x) = \begin{cases} 0.5 & x < 60 \\ \dfrac{x-10}{100} & 60 \leq x \leq 110 \\ 1 & x > 110. \end{cases}$$

(b) The factor $0.4^{t/20}$ is such that $q_{x,t}$ declines exponentially from its base value $q_{x,0}$ to its ultimate (asymptotic) value $\alpha(x)q_{x,0}$, and in every 20 years it falls by 60% of the remaining distance between its current and ultimate values.

1.4 *Costs of Past Forecast Errors*

It is generally the case that projections made over the last 35 years have underestimated overall mortality improvements. This has been true of official projections and the CMIB's projections. In this section we illustrate the financial impact this would have had on life offices' annuity business.

Table 1: Illustrative annuity pricing assumptions.

	1969	1973	1977	1989	1993	1997
Quadrennium	1967–70	1971–74	1975–78	1987–90	1991–94	1995–98
Interest rate	8.50%	11.90%	10.90%	10.00%	6.40%	6.30%
Table	PEG	PEG	PEG	PMA80C10	PMA80C10	PMA92C20
100 A/E	100.00%	99.00%	97.00%	108.00%	99.00%	132.00%
Projection	PEG	PEG	PEG	"80" Series	"92" Series	"92" Series

Table 2: Estimated joint life annuity prices charged by life offices

Male age	1969	1973	1977	1989	1993	1997
55	10.511	8.355	8.912	10.069	13.993	14.358
60	9.845	7.970	8.470	9.631	13.098	13.442
65	9.053	7.482	7.918	9.031	11.984	12.263

Table 3: Actual cost of joint life annuities sold by life offices

Male age	1969	1973	1977	1989	1993	1997
55	10.659	8.450	9.064	10.262	14.276	14.584
60	9.930	8.051	8.609	9.869	13.485	13.762
65	9.055	7.525	8.040	9.256	12.415	12.682

We estimated the annuity prices for business written over the period 1969 to 1997 using the aggregate CMIB experience for the relevant quadrennium along with the relevant CMIB projection. We used level joint life annuities with a 50% spouse's pension, assuming females to be 3 years younger than males. The annuities were based on gilt yields prevalent at the time. No allowance was made for offices' profit margins and expense loadings. Table 1 sets out the mortality and interest rate assumptions used. Table 2 sets out the annuity values calculated.

The actual cost of the annuities was then re-estimated using the aggregate CMIB experience over 1969–2000 and assuming that mortality followed the medium cohort projection from 2001. The results of these calculations are shown in Table 3.

Table 4 shows the mortality losses as a percentage of premiums charged. Until the 1980s, the mortality losses at age 65 were below 2% of premiums and may have been absorbed by offices' loadings for prudence and profits. Since the 1980s, the mortality losses have steadily increased to some 3.5% as the cohort centred on 1926 started exhibiting much greater mortality improvements than expected. As male mortality has improved faster than female mortality relative to past projections, the mortality losses for single life male annuities will be higher and it is less likely that offices' loadings would have been sufficient to cover them.

Table 4: Estimated loss on sale of joint life annuities by life offices, as a percentage of premiums charged.

Male age	1969	1973	1977	1989	1993	1997
55	1.41%	1.14%	1.71%	1.92%	2.02%	1.57%
60	0.86%	1.02%	1.64%	2.47%	2.95%	2.38%
65	0.02%	0.57%	1.54%	2.49%	3.60%	3.42%

We also estimated the extent of the mortality loss avoided by life offices through the use of projected improvements. Table 5 shows the estimated annuity rates using the same assumptions as in Table 2 but without allowing for any projected improvements.

Table 5: Estimated joint life annuity prices with no allowance for future mortality improvements.

Male age	1969	1973	1977	1989	1993	1997
55	10.435	8.319	8.868	9.982	13.614	14.038
60	9.767	7.930	8.421	9.532	12.720	13.109
65	8.976	7.439	7.867	8.926	11.630	11.941

Table 6 shows the estimated loss that would have arisen if life offices had not made any allowance for projected improvements, as a percentage of the premium charged. Comparison with Table 4 shows the level of mortality losses avoided through the use of the CMIB projected improvements in mortality.

Table 6: Estimated loss on sale of joint life annuities by life offices, as a percentage of premiums charged, had no allowance been made for future mortality improvements.

Male age	1969	1973	1977	1989	1993	1997
55	2.15 %	1.57 %	2.21 %	2.81 %	4.86 %	3.89 %
60	1.67 %	1.53 %	2.23 %	3.54 %	6.01 %	4.98 %
65	0.88 %	1.16 %	2.20 %	3.70 %	6.75 %	6.21 %

1.5 *Reasons for Producing New Projections*

The reasons for producing new projections are as follows.

(a) Appendices A1.4 and A1.6 of CMIB Working Paper No.1 showed that the male Pensioners experience appeared on the whole to be lighter than that of the "92" Series tables projected to 1999. Therefore, the past history of projections being too pessimistic appeared to be recurring. In producing new standard tables based on the 1999–2002 experience, it would be prudent to review the basis for projections as well.

(b) The methodology employed in the past by the CMIB is one of several advocated by demographers and others, see for example Tabeau *et al.* (2001), and also one of the simplest. Considerable progress has been made in the 1990s, for example in statistical methods of forecasting, that was not available to the CMIB when it last considered projection methodology. This is receiving much attention from actuaries as well as demographers, see for example Tuljapurkar & Boe (1998). It is appropriate to take this into account.

(c) The need to give some indication of uncertainty in projections, to allow a more transparent approach to risk management, requires a new approach. This is discussed in more detail in the following two sections.

1.6 *The Need for a Risk Management Approach*

Longevity risk resembles investment risk, in that it is non-diversifiable: it cannot be controlled by the usual insurance mechanism of selling large numbers of policies, because they are not independent in respect of that source of uncertainty: the law of large numbers does not apply. However it is different in that there are no large traded markets in longevity risk, so its price cannot be directly observed, and it cannot easily be hedged, though it could possibly be offset. Rather, the price for this risk is calculated by those who buy it (insurance companies).

In the case of investment risk, methods such as scenario testing and value at risk (VaR) are used to set capital and margin requirements precisely in respect of that part of the risk that cannot be offset or hedged across an appropriate time horizon. VaR (and similar risk measures, such as conditional tail expectation (CTE)) can be viewed as scenario testing with a probabilistic model generating the scenarios. Increasingly, insurers are moving, or being moved, towards the use of such risk management tools, not least because of the imminent IASB 'fair valuation' rules, FSA 'realistic balance sheet' requirements, and convergence of regulatory regimes in banking and insurance.

It is interesting to revisit the early 1950s, and to recall the economic background to the famous papers of Redington (1952), Haynes & Kirton (1953) and others. Interest rates had recently been very low, below the levels assumed in many past pricing bases, and life funds were vulnerable to the long-term course of interest rates. This was exactly the problem of systematic risk, immune to the law of large numbers. Had today's computing power been available then, surely some of the great insights — the "expanding funnel of doubt", immunisation and cash-flow matching — would have led to quantitative measures of risk, perhaps even those we use today. In their absence, the consensus that emerged was the need to shift the balance of life funds away from non-profit business and towards with-profit business, as a practicable form of risk management.

Insurers who today issue long-term guarantees based on future longevity are in a rather similar position to insurers who, 50 years ago, were offering long-term guarantees based on future interest rates. Today, however, we have accessible methods for the measurement of systematic risk. Their use is practically mandatory in the management of investment risk, and the Sub-Committee believes that similar approaches are valuable aids to the management of longevity risk. This does not imply that these methods, or models upon which they may be based, are beyond criticism, rather that making no attempt at all is not an option. In practical terms, this means that:

(a) The Sub-Committee will, in future, always attempt to provide a measure of uncertainty with projections of future mortality rates. The three scenarios given in Working Paper No.1 may be viewed as the first time this objective has been followed, but in this Working Paper we will consider whether or not a less *ad hoc* approach is feasible.

(b) While users of the CMIB's mortality projections may find the measures of uncertainty useful, they are themselves responsible for the approach taken in their own particular circumstances. That is, just as the responsibility for mortality bases has always fallen upon the individual actuary, who might use the CMIB's tables as a benchmark, so does the responsibility for any assumed uncertainty that may be needed for risk management.

1.7 *FSA Requirements (PSB)*

The draft rules for the Integrated Prudential Sourcebook (PSB), detailed in CP195, contain the existing requirement that when setting mathematical reserves a firm must include appropriate margins for adverse deviation of relevant factors. However, the draft rules go further than previous valuation rules in specifying what is meant by "appropriate margins for adverse deviations". In particular it is stated that:

(a) The margin for adverse deviation of a risk should generally be greater than or equal to the market price for that risk.

(b) Where a risk premium is not readily available, or cannot be determined, an external proxy for the risk should be used, such as adjusted industry mortality tables.

(c) Where there is a considerable range of possible outcomes the FSA expects firms to use stochastic techniques to evaluate these risks. In time, for example, longevity risk, where this constitutes a significant risk for the firm, may fall into this category.

It is further stated in the draft rules that in setting rates of mortality, which contain prudent margins for adverse deviations, a firm should take account of *inter alia*:

(a) The credibility of the firm's actual experience as a basis for projecting future experience including:
    (1) whether there are sufficient data; and
    (2) whether the data are reliable and have been appropriately validated.

(b) The availability and reliability of:
    (1) any published tables; and
    (2) any other information as to the industry-wide experience.

(c) Anticipated or possible future trends in experience including (but only where they increase the liability):

(1) anticipated improvements in mortality;
(2) diseases the impact of which may not yet be reflected fully in current experience; and
(3) changes in market segmentation (such as impaired life annuities) which, in the light of developing experience, may require different assumptions for different parts of the policy class.

In addition to the reserving requirements the draft rules for the PSB cover the assessment of the adequacy of a firm's capital resources. This will involve identifying the major risks faced by a firm and quantifying the capital required to cover these risks with a particular confidence level. For insurance risk the factors mentioned as requiring consideration include:

(a) The potential for catastrophic losses.
(b) Determination of the effect of claims experience being more costly than planned by analysing historic claims experience, volatility and trends in experience.
(c) The frequency and size of large claims.
(d) The ability of the firm to withstand catastrophic events, increases in unexpected exposures, latent claims or aggregation of claims.
(e) The risk of variations in mortality experience.

It is stated that the FSA places credence in capital requirements based on models which transform each element in the financial projection into a statistical distribution with a range of possible outcomes (generally incorporating an economic model which is linked into the generation of insurance related assumptions). However it is not mandatory to use such an approach.

These requirements point to the need for the CMIB to generate mortality projections which include probabilities of different scenarios. The FSA's favouring of stochastic techniques would appear to point to an ideal of the CMIB providing models which insurers could use, varying the parameters to reflect their own judgement.

2. Projection Methodologies

2.1 *General*

The general approaches to projecting age specific mortality rates can be categorised in various different ways, for example, as process-based, explanatory, extrapolative or some combination of these.

Process-based methods concentrate on the factors that determine deaths and attempt to model mortality rates from a bio-medical perspective. An example of this is the reliability theory of ageing which is described in Appendix A, Section A.4 (see Gavrilov & Gavrilova (2003)). Such methods are not generally used to make projections, but may be useful in informing extrapolative methods. These methods are only effective to the extent that the processes causing death are understood and can be modelled mathematically.

Explanatory-based methods employ a causal forecasting approach, for example using econometric techniques based on variables such as economic or environmental factors.

However, most potential explanatory links are not understood well enough. Data allowing deaths to be categorised by the risk factors considered and length of exposure may not be readily available. If the explanatory variables themselves are as difficult to predict as the dependent variables (or indeed more so), then the projection's reliability will not be improved by including them in the model. Even so, a partial attempt for projecting mortality improvements might be made using data where a link is known (for example, extrapolating minimum mortality improvements relying solely on changes in lifetime smoking patterns and their emerging effects on lung cancer). Tabeau *et al.* (2001) describes attempts to model Dutch mortality using various explanatory variables.

Extrapolative methods are based on projecting historical trends in mortality into the future. All such methods include some element of subjective judgement, for example in the choice of period over which the trends are to be determined. Simple extrapolative methods are only reliable to the extent that the conditions which led to changing mortality rates in the past will continue to have a similar impact in the future. Advances in medicine or the emergence of new diseases could invalidate the results of an extrapolative projection. Some examples of approaches to extrapolation are discussed below.

A variety of methods is available to carry out projections under each of these general approaches. Trend-based methods involve the projection of historical trends in the variables under consideration into the future. However, the relationship between mortality rates, or other variables, at different ages is often ignored in these methods and thus the results when translated back into age-specific mortality rates may appear implausible. This might happen, for example, when producing aggregate all-cause mortality rates from rates of mortality projected by cause of death or independently projecting forward age-specific mortality rates at different ages. These may produce results which are counter intuitive, although consistent with the assumptions made (for example, mortality rates at older ages may eventually become lower than those at younger ages).

Parametric methods involve fitting a parameterised curve to data for previous years and then projecting trends in these parameters forward. However, the shape of the curve may not continue to describe mortality satisfactorily in the future.

Targeting methods involve assuming a target or set of targets which it is assumed the population for which the projection is being made will approach over time. The targets could be expressed in terms of a set of age-specific mortality rates (either aggregate or by cause of death), specified rates of mortality improvement or expectations of life or in terms of other variables being projected forward, for example. Assumptions are required as to the time at which the target would be reached and the speed of convergence. The target levels may be obtained by analysis of past trends in historic data for the population being projected (in which case this might be regarded as an extension of an extrapolation methodology) or might be obtained from a different population group. Targeting can overcome some of the drawbacks of a purely extrapolative approach, since the targets chosen can take into account any evidence of the possible effects of advances in medical practice, changes in the incidence of disease or the recent emergence of new diseases.

All these methodologies can be used to provide deterministic projections. Also, having fitted a model to historical data most methodologies can then be adapted in some way to provide stochastic projections. Methods which use some form of targeting or setting of parameters can provide stochastic projections by choosing parameters at random from an assumed probability distribution.

Most methodologies can be applied either to aggregate mortality data or to data by cause of death or to some other variable. Projecting mortality by cause of death appears to provide a number of benefits such as providing insights into the ways in which mortality is changing. However, there are problems associated with this approach; as discussed in Section 5 of this paper.

2.2 *Types of Model in Use or Proposed for Use*

Both the current CMIB methodology, and the methodology used by the GAD for projecting mortality in the official national population projections for the U.K. and its constituent countries, are examples of extrapolation with the use of targets. Other methodologies being used in practice or proposed for use include age-period-cohort models, the Lee-Carter method and parametric smoothing models. These are discussed briefly in the following paragraphs.

(a) *Age-Period-Cohort Models.* Age-Period-Cohort (APC) models study variation in demographic rates (for our purposes, mortality rates) along three critical dimensions: age, year (or period) of occurrence, and cohort. The basic form of the model is:

$$\log \mu(x,t) = s_a(x) + s_b(t) + s_c(t-x) \tag{3}$$

where $\mu(x,t)$ is the force of mortality at age x in year t and s_a, s_b and s_c are smooth functions.

There is a fundamental problem with this method, as *cohort + age = period.* Therefore it is not possible to break down changes in demographic rates in terms of the three variables.

Biologically, age effects are clearly an important factor in considering mortality. Period effects are intuitively significant (consider a very bad winter, for example). The role of cohort-specific effects is less obvious. An individual experiences certain critical events (for example infancy, education) alongside peers from their particular cohort, and the after-effects of these situations may remain with that individual for life. However, there do appear to be cohort specific effects in mortality data for the U.K. population as a whole, with those born in the late 1920s and early 1930s enjoying higher mortality improvements than generations born either side.

(b) *The Lee-Carter Method.* There has been particular interest in recent years in the Lee-Carter methodology for projecting mortality rates, first proposed in the early 1990s (Lee & Carter, 1992). This is a bilinear model in the variables x (age) and t (calendar time) of the following form:

$$\log \mu(x,t) = a(x) + b(x)k(t) + \epsilon(x,t) \tag{4}$$

where $\epsilon(x,t)$ is a random error term (or stochastic innovation). The $a(x)$ coefficients describe the average level of the $\log \mu(x,t)$ surface over time. The $b(x)$ coefficients describe the pattern of deviations from the age profile as the parameter k varies. If the $b(x)$ coefficient is particularly high for some ages x, then this means that mortality rates improve faster at these ages than in general. If it were negative at some ages, this would mean that mortality was getting worse. If the $b(x)$ were all equal then mortality rates would change at the same rate at all ages.

The $k(t)$ parameter describes the change in overall mortality. If $k(t)$ falls, then mortality rates fall, and if $k(t)$ rises, then mortality rates rise. The coefficients $b(x)$ determine how this overall change in mortality affects rates at the age in question. If $k(t)$ decreases linearly, then $\mu(x,t)$ decreases exponentially at each age, at a rate that depends on $b(x)$ (unless $b(x)$ is negative, in which case $\mu(x,t)$ increases).

Lee & Carter (1992) suggested using a time series model for $k(t)$, so projections based on a Lee-Carter model share many of the statistical features of time series forecasts that may be familiar to actuaries from economic forecasting. In order to get a unique solution when fitting a Lee-Carter model, constraints must be imposed. For example, if we have a solution $a(x)$, $b(x)$ and $k(t)$, then another solution is $a(x)$, $cb(x)$ and $k(t)/c$, for any non-zero constant c. Usually $b(x)$ is constrained to sum to 1 and $k(t)$ to sum to 0.

Having fitted the rates to this surface, $k(t)$ is projected forward to give mortality rates for future years. The method can be used to make stochastic projections, by modelling $k(t)$ as a suitable time series.

The Lee-Carter method is a very simple model. It is highly structured and has been used very successfully on U.S. data where it has picked up almost all of the variation in the data. It has also been successfully applied elsewhere, see for example Brouhns *et al.* (2002a, 2002b) or Booth & Tickle (2003). However, applying it to U.K. population data has proved more difficult, partly because of the difficulty in dealing with the cohort effects seen in historical data, which the Lee-Carter method in its usual form of projecting by age and calendar year smooths out.

(c) *Parametric Smoothing Models.* Parametric smoothing models are among those most familiar to actuaries, since they have been used for mortality graduations for many years. The Gompertz model and its many generalisations are examples, including the Gompertz-Makeham family used for recent CMIB graduations, and the Heligman-Pollard model. Other examples are spline models, like those used for recent English Life Tables, or the 2-dimensional P-spline model used by Dr Iain Currie to model the CMIB assured lives experience in CMIB Working Paper No. 1.

From a statistical point of view, parametric smoothing models may be regarded as types of regression model. In particular, this defines the approaches that may be used to fit them to data, and also to describe the statistical properties of the resulting fitted model. We will discuss this at length in Section 4.

The crucial question from our point of view is whether or not a fitted model, of any kind, and any measures of uncertainty about its goodness-of-fit, can be extrapolated sensibly beyond the data. The answer may determine the feasibility of making useable probabilistic projections of mortality rates over very long periods.

3. The Cohort Effect

Concurrently with the preparation of this Working Paper, Willets (2004) and Willets *et al.* (2004) have prepared a very extensive study of trends in mortality in the U.K. and elsewhere, and have found convincing evidence of a cohort effect. We refer the reader there for full details. In view of the importance for pension provision, of the cohort enjoying the highest mortality improvents, we conclude that it is prudent for projections to take account of cohort mortality, though this does not necessarily mean that an APC model should be recommended. The interim projections in CMIB Working Paper No. 1 already allow for the cohort effect.

4. Uncertainty

4.1 *Sources of Uncertainty*

Several sources of uncertainty may influence the modelling of rates of mortality, and their projection into the future. Three well-known categories associated specifically with the use of statistical models (Cairns, 2000) are:

(a) *Model uncertainty.* Often a choice of models presents itself, each perhaps equally plausible on *a priori* grounds. An example is the choice of a particular member of the Gompertz-Makeham family for the graduation of a given experience, or the wider choice between the Gompertz-Makeham family and alternative parametric models. If an appropriate family of models has been chosen, collecting more data may reduce model uncertainty.

(b) *Parameter uncertainty.* Even if model uncertainty was absent, and the 'correct' model was known, there would be uncertainty about the choice of parameters suggested by any finite set of observations. This is often capable of being measured by estimating the distribution of the parameter estimates. Given the 'correct' model, collecting more data reduces parameter uncertainty.

(c) *Stochastic uncertainty.* When a model is used for prediction, which is usually the aim in actuarial science, the predicted quantity may be inherently stochastic. For example, suppose a model has been chosen to represent mortality rates in a given population by age and calendar year; it has been parameterised using historic data; and it is to be used to predict the number of deaths next year. Even if we knew the correct model and the correct parameters, the outcome would be uncertain.

In reality, any difference between predicted and actual outcomes may be due to a mixture of all three types of uncertainty. In addition, there are other significant sources of uncertainty, associated with the particular task of modelling and projecting rates of mortality.

(a) Measurement error is a major practical problem in life office work: are the raw data correct? Late reporting of deaths, incorrectly entered dates of birth and other details, maladministered policies and administration backlogs all add an extra dimension of uncertainty before we even consider a statistical model.

(b) Heterogeneity is an issue, perhaps particularly with reference to socio-economic mix. We may observe an aggregate μ_x across all lifestyle groups, and parameterise a model based on the aggregate experience. But if heterogeneity is important, the emerging experience may have a very different shape to the model predictions.

(c) The past may not be a good guide to the future. The profile of annuity buyers today is different from a decade ago. Pension annuities were originally only purchased by a select group, namely self-employed holders of Section 226 policies. Today, a much broader spectrum of the population has a personal pension fund with which to buy an annuity. The experience of the portfolio to date may be of limited applicability to new business pricing.

In the context of projecting future mortality, we focus on model and parameter uncertainty. Stochastic uncertainty depends mainly on the size of the insurer's annuity portfolio, and will not be discussed in this Working Paper.

4.2 *Fitting Data and Making Projections*

Quantifying uncertainty forces us away from *ad hoc* approaches and places the model in centre stage, so we must begin by asking: what exactly do we suppose the model to be?

It is helpful to distinguish between the rôles of the model in the region of the data and in the region of the projection; they may be very different. In this case 'region of the data' means the past, and 'region of the projection' means the future. We want a model to do two things:

(a) To be sensible in the region of the data, it should represent the data well enough, given the usual requirements of parsimony and adherence to the data. This usually involves a trade-off between smoothness and goodness-of-fit.

(b) To be sensible in the region of the projection, it should be constrained to behave in ways that are reasonable, or at least plausible, given the nature of the quantities being modelled.

The following example, of a polynomial regression model, may be helpful[1]. Suppose we want to model how children grow taller as they grow older. At each of ages 1 to 18 (the region of the data), we find the mean and standard deviation of the heights of a sample group of children. We plot these on a graph, and first try to fit the simplest polynomial model, a straight line. If it does not fit very well, we try the next simplest, a quadratic model, then a cubic model and so on, until we have an acceptable balance between goodness-of-fit and simplicity. This may give excellent results in the region of the data. Now suppose we wish to project heights at adult ages (the region of the projection) from these data. A polynomial that fits the data very well may behave arbitrarily badly outside the region of the data; there are no constraints on its behaviour there. In the extreme, a polynomial of order 20 would fit the data perfectly, but would be almost guaranteed to give a nonsensical projection[2].

So a model must be chosen that has enough flexibility to provide a parsimonious fit to the data, yet enough structure that it can carry sensible features into the region of the projection. The longer the period over which projections are required, the more difficult this may be.

[1] The CMIB graduations based on Gompertz-Makeham functions are related, being 'polynomial + exp(polynomial)' regression models

[2] The Gompertz-Makeham family, although it has worked well, is not exempt from this problem. Graduated rates of mortality at very high ages are, for all practical purposes, projections beyond the region of the data, and are often observed to turn downwards at age 100 or over.

4.3 *Confidence Intervals in the Region of the Projection*

Let us suppose for the moment that the measure of uncertainty we seek is a confidence interval around a central projection. This is easy to envisage, and is consistent with risk management methodologies used elsewhere, but how should we interpret it? How do the data, by definition confined to the region of the data, produce probabilistic statements that apply in the region of the projection? The answer depends on the methodology, and in particular whether we use regression models or time series models. Since P-spline models are examples of the former, and Lee-Carter models are examples of the latter, there is an interesting choice to be made. What is rather unclear is whether there is a rational basis for that choice.

(a) Regression models take a family of basis functions, and choose a combination of them that best fits the data according to some criterion. For example, the naive[3] polynomial regression model fits linear combinations of the basis functions $1, x, x^2, x^3, \ldots$, spline models fit linear combinations of the basis splines, and so on. The projection is then the extrapolation of the fitted function beyond the region of the data.

Suppose the best-fitting model uses the first N basis functions. Then the fitted coefficients (or parameters) form an N-dimensional vector, which is a realisation of an N-dimensional estimator, and has an $N \times N$ sampling variance matrix, which is estimated along with the parameters. The probabilistic properties of the parameter estimates define the probability that the parameter vector should lie within any given region of the N-dimensional parameter space; very often an asymptotic result is called upon so we can suppose that the parameter estimate has a multivariate Normal distribution, which makes it tractable.

In particular, we can define a region in N-dimensional space, centred on the parameter estimate, that we suppose contains the true parameter with 95% probability. (We ought to choose a region centred on the true parameter, but the estimate is the best we have got.) As the parameter wanders around in this region, both the fitted and the projected models vary as the regression coefficients change. The confidence intervals for the projection are defined by the regions of the plane visited by the projection as the parameter varies. Thus, the source of the uncertainty in the region of the projection is the variance matrix of the fitted parameters.

Actually computing the confidence intervals for the projection analytically may be impossible, but they can usually be found by simulation, and in fact this has been a feature of recent CMIB graduations. Forfar, McCutcheon & Wilkie (1988) estimated confidence intervals for graduations in the following way:

(1) Fit a member of the Gompertz-Makeham family to an age range where the data seem reliable, and estimate the variance matrix of the fitted parameters.

(2) Assuming the parameter estimate to have a multivariate Normal distribution, specified by the fitted value and the variance matrix, simulate a large number of vectors from this distribution.

(3) Plot or otherwise record the graduation function for each simulated parameter

[3] We say 'naive' because $1, x, x^2, x^3, \ldots$ are not ideal basis functions for polynomial regression; adding a term can change the already-fitted coefficients, for example. Certain systems of orthogonal polynomials have better properties, such as the Chebyshev polynomials used in recent CMIB graduations.

vector. Recall that at extreme ages the graduation is in fact a projection.

(4) Estimate confidence intervals at each age by taking the appropriate percentiles of the simulated graduations. For example if 1,000 such simulations have been made, the 25th and 975th (ordered) values of graduation functions at age x provide an approximate 95% confidence interval for μ_x (or whatever other kind of rate was modelled).

See Forfar, McCutcheon & Wilkie (1988) Section 11 and Figures 15.2, 16.2, 16.3, 17.3 and 17.4 for examples. To statisticians, this procedure is known as 'parametric bootstrapping'.

From the description above, it should be evident that this confidence interval measures parameter uncertainty only. It does not account for the model uncertainty, because it does not consider the possibility that a different member of the Gompertz-Makeham family might have been chosen, or a member of a completely different parametric family. The most we can say about confidence intervals allowing for model uncertainty is that they might be different.

(b) Turning to time series models, many actuaries may be more familiar with the analysis of financial time series than with mortality projections, so we will begin our discussion there. Suppose we have some financial data, for example values of an index on each 1 January for many past years. Fitting a time series model means deciding how much of the irregularity observed in the data is structural, and should be retained in the model, and how much is noise, part of the error term. This takes the place of the trade-off of smoothness against goodness-of-fit in a regression model (to which we return in Section 4.5). 'Best-estimate' projections are then based on the structural part of the model, with the fitted error term providing measures of uncertainty such as confidence intervals.

For example, an autoregressive process of order 1 (an AR(1) process) for the quantity $x(t)$ can be specified as:

$$x(t) = \alpha + \beta x(t-1) + \epsilon(t) \tag{5}$$

where α and β are parameters and $\epsilon(t)$ is the random innovation at time t. The structural part of the model is $\alpha + \beta x(t-1)$, determining how $x(t)$ would behave if there were no random innovations. The innovations $\epsilon(t)$ may require further parameters for their definition, for example if they are assumed to be Normal. If T is the time of the last observation, the 'best estimate' projection is simply the iterated application of $x(t) = \alpha + \beta x(t-1)$, and the standard error of the projection[4] is a function of the unobserved $\epsilon(T+1),\ \epsilon(T+2),\ \ldots$.

An interesting point is that time series data are usually *measurements* and not *estimates*. Each datum in a time series is usually a single observation, whereas in fitting a mortality model we are dealing with estimates at each age derived from samples. There are no standard errors around the individual data values in a financial time series model. If we tried to fit a regression model to time series data, we would get no

[4] In the time series literature the word 'forecast' would be used more often then 'projection', and standard errors and so on would be called 'forecast standard errors'. This is quite a helpful convention that could usefully be adopted for mortality studies, but we will not do so here.

information about parameter uncertainty, and it would degenerate into curve-fitting. We would therefore get no idea of a confidence interval in the region of the projection from such an approach. This suggests that, in a time series framework, the choice of model, and not the size of the experience, is most influential in determining measures of uncertainty.

The projection part of the Lee-Carter model takes a time series approach, but unlike a financial time series the data points are estimates with sampling distributions, so we have both estimation uncertainty from the data and projection uncertainty from the fitted time series. In Lee & Carter (1992) it is suggested that the latter dominates and the former can be ignored, which is very helpful for computation, but Brouhns, Denuit & Vermunt (2002b) show how to incorporate both sources of error. On the assumption that the progression of 'true' mortality rates over calendar time has some smooth structure, the effect of larger samples will be to reduce the variability of the estimates over calendar time, hence reducing the variance of the modelled residuals and the standard errors in the region of the projection.

Like the regression model, the confidence intervals in the region of the projection capture parameter risk only. If we have fitted an ARIMA(0,1,0) model, they tell us nothing about the uncertainty that is present because we could have fitted an alternative ARIMA model, or a member of any other class of model.

(c) Both approaches above, regression and time series models, seemed to lead to estimates of parameter uncertainty only after we had chosen what model to fit. Can we somehow incorporate model uncertainty? This question is discussed at length in Cairns (2000) and we will refer the reader there for details. One approach which is certainly feasible is to fit different models and see how sensitive the results are to model choice; a form of sensitivity analysis. To go beyond this to make probabilistic statements about model risk is very difficult, because that requires us to define a universe of possible models and then to define a probability distribution on the models in that universe. Cairns (2000) suggests a Bayesian approach, in which uninformative priors can be used, but also warns that it has pitfalls: "... when two models give similar fits but make significantly different predictions about the future, a change in the prior model probabilities can have a significant impact on the posterior distribution of the quantity y of interest. ... it may just be the case that it is not possible to combine the results from different models into a single statement because we are unwilling to prescribe specific prior model probabilities."

We are of the view that that is the case here. It is difficult to see how we might define a probability distribution on a universe containing the Gompertz-Makeham family, other parametric models, Lee-Carter models and any other models that might be plausible. Any attempt might lead to some interesting research, but we propose to accept that we can only try to quantify parameter uncertainty, and that model uncertainty might best be dealt with by sensitivity analysis. We cannot avoid choosing a model.

4.4 *What Experience(s) Should We Project?*

Past projections made by the CMIB have been guided by historic trends in the largest and most reliable mortality experiences available, such as the U.K. population and Assured Lives data, and the resulting projections, in the form of two-dimensional mortality tables, have then been applied to the individual experiences underlying each office's annuity portfolios, which are often much smaller. When all that was projected was a trend line, there was no statistical model, and issues of sample size did not arise. But is it legitimate to obtain measures of uncertainty, such as confidence intervals, from a large mortality experience and then to apply these to a different, smaller experience? Actuaries have always known to take care in applying life tables based on one experience to any other experience, and similar care is needed with measures of uncertainty. This question may be posed at several levels.

The argument for collecting data and producing standard tables separately for different classes of business and for males and females is that each might have quite different characteristics, for example because of socio-economic mix. The Graduation Working Party of the Mortality Sub-Committee, which is currently preparing trial graduations based on the 1999–2002 data, intends to use the same methods as were used for the "92" Series tables, namely fitting Gompertz-Makeham (or other) models to separate experiences. Since projection depends on the model, it would be consistent from a statistical point of view to make separate projections as well.

This would not cause problems for the users of CMIB standard tables. Suppose the CMIB had produced projections for a particular class of business, say retirement annuities in payment. An insurance company with a portfolio of such policies would not fall into any statistical error by using these projections, including the measures of uncertainty, as the basis for risk management. This is because its own portfolio is (presumably) part of the CMIB experience which formed the basis of the projection. The model makes the (strong) assumption that all parts of the experience are homogeneous, so any characterisation of parameter uncertainty based on modelling the entire experience carries over to subsets of it.

We might try to argue along similar lines, and say that we could make projections based on the U.K. population, and that since the lives in any CMIB experience are (presumably) part of the U.K. population, those projections are applicable. The flaw in this argument is that CMIB experiences are not assumed to be homogeneous parts of the experience of the U.K. population, in fact they are rather well known not to be.

In the next section, we explore ways in which projections for CMIB experiences might be able to draw upon information derived from other, larger experiences. In essence we propose versions of the technique known as 'graduation by reference to a standard table', long used to graduate sparse experiences, but in the setting of statistical models.

4.5 *Examples of Uncertainty in Projections*

To avoid repetition, from now on we assume that 'making a projection' necessarily includes measures of uncertainty.

Figure 1 shows two mortality experiences. The upper plot shows crude μ_{65} from the CMIB Assured Lives experience, from 1951 to 1999. Also shown is a fitted model and the associated 95% confidence limits, and a projection from 2000 to 2049, with its associated

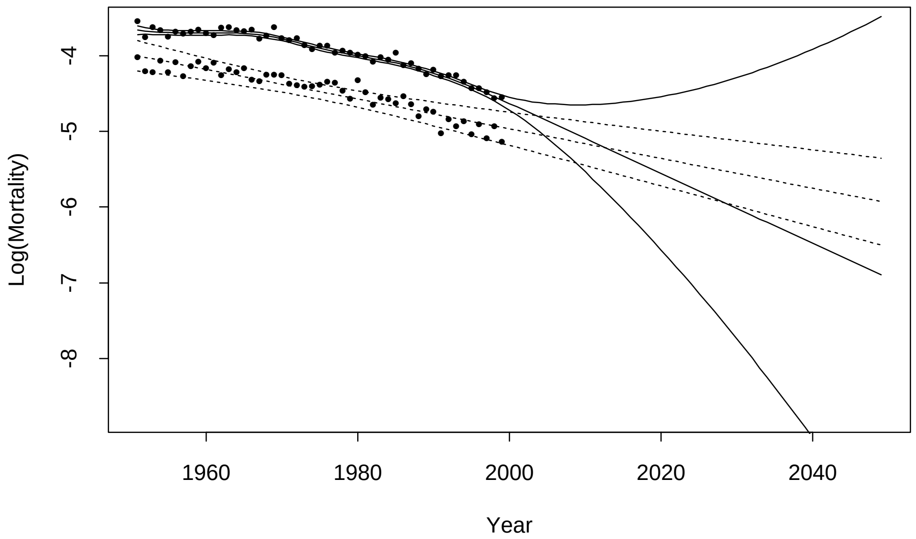

Figure 1: Fitted and projected models of a larger (top) and smaller (bottom) mortality experience. P-spline model with smoothing parameter chosen separately for each experience. 95% confidence intervals are shown.

95% confidence limits. The lower plot shows μ_{60} from the same data, but with the actual deaths and exposures divided by 100, to simulate a much smaller experience. This does not properly represent the dispersion we would expect to see in a smaller experience, but it allows us to illustrate the consequences of fitting models to estimates with very different standard errors.

The form of the models fitted in this example is a one-dimensional P-spline, similar in nature to the two-dimensional P-spline models used to smooth the CMIB experiences in Working Paper No. 1. The degree of smoothing is determined by the choice of smoothing parameter and the number of regression splines, and this choice is made to optimise a certain penalty function that trades off smoothness against goodness-of-fit.

The most obvious, and at first sight surprising, feature of Figure 1 is that the *smaller* experience has much narrower confidence intervals, and that the confidence intervals of the larger experience are very wide indeed. The more data we have, the less certain we appear to be about the projection. How can this be so?

(a) Notice that the fit to the smaller experience is not very good at the extremes, but this is because the estimates have relatively large standard errors so smoothness is preferred to goodness-of-fit. In fact the fitted function is linear, so essentially only two parameters are needed, slope and intercept. The confidence intervals in the region of the projection reflect the uncertainty associated with the choice of these parameters, which can be summed up by the question, “how many other straight lines could be plausibly fitted to these data?”. The answer is: “not very many”. With such a simple model, there is not much 'wiggle room' when it comes to fitting it, so the projection lies in quite a narrow funnel of doubt.

(b) The estimates in the larger experience have much smaller standard errors, so goodness-of-fit is given greater weight. This is achieved by choosing a smaller smoothing parameter. Many more parameters are fitted, and the uncertainty associated with each of them adds to the uncertainty of where the projection might go, leading to wide confidence intervals as soon as we leave the region of the data.

(c) For the technically minded, the variance matrix of the fitted parameter vector $\hat{\theta}$ in the P-spline model takes the form:

$$\mathrm{Var}[\hat{\theta}] = (B'WB + \lambda P)^{-1} \tag{6}$$

where B is the regression matrix, W is a (diagonal) matrix of weights, P is a penalty matrix and λ is a smoothing parameter. λ is chosen by optimising the Bayesian Information Criterion, which means that it is chosen by reference to the data. Large values of λ lead to smoothness rather than goodness-of-fit, but at the same time, from the equation above, to a 'small' variance for $\hat{\theta}$, which is the reason for the narrower confidence intervals. Small values of λ lead to a 'large' variance for $\hat{\theta}$, and wide confidence intervals for the projection. An alternative approach would be to propose some prior distribution for λ and adopt a Bayesian model. In that framework:

$$\begin{aligned}
\mathrm{Var}[\hat{\theta}] &= \mathrm{Var}_\lambda[\mathrm{E}[\hat{\theta}|\lambda]] + \mathrm{E}_\lambda[\mathrm{Var}[\hat{\theta}|\lambda]] \\
&= \mathrm{Var}_\lambda[\mathrm{E}[\hat{\theta}|\lambda]] + \mathrm{E}_\lambda[(B'WB + \lambda P)^{-1}] \\
&\approx \mathrm{Var}_\lambda[\mathrm{E}[\hat{\theta}|\lambda]] + (B'WB + \lambda P)^{-1}
\end{aligned}$$

where the subscript λ on the expectation and variance operators indicates that they are taken with respect to the prior distribution of λ. The first term $\mathrm{Var}_\lambda[\mathrm{E}[\hat{\theta}|\lambda]]$ would allow for model uncertainty, and it could be very large, as in the graduation of μ_{60} in Figure 1.

Another feature of Figure 1 is that μ_{65} quite quickly falls below μ_{60}. This would usually be regarded as anomalous, but such anomalies are very likely to appear if we project different experiences separately.

Figure 2 shows what happens if we take the model that is optimal for the larger experience and also fit it to the smaller experience. Now, more in line with intuition, the smaller experience has the larger confidence intervals around the projection. μ_{65} still falls below μ_{60}, though at a much later duration; these are still separate projections.

The contrast between Figures 1 and 2 emphasises the huge effect that model choice can have on the statistical properties of the projection.

For convenience, call the larger experience in the example above 'experience A' and the smaller one 'experience B'. Because we modelled each quite separately, we can learn nothing at all about experience A from the model of experience B, and *vice versa.* If we believe that experiences A and B share some common features that mean that they do contain information about each other, this is in a sense wasteful. This information can be harnessed by proposing a *joint* model for both experiences. A naive approach might be to

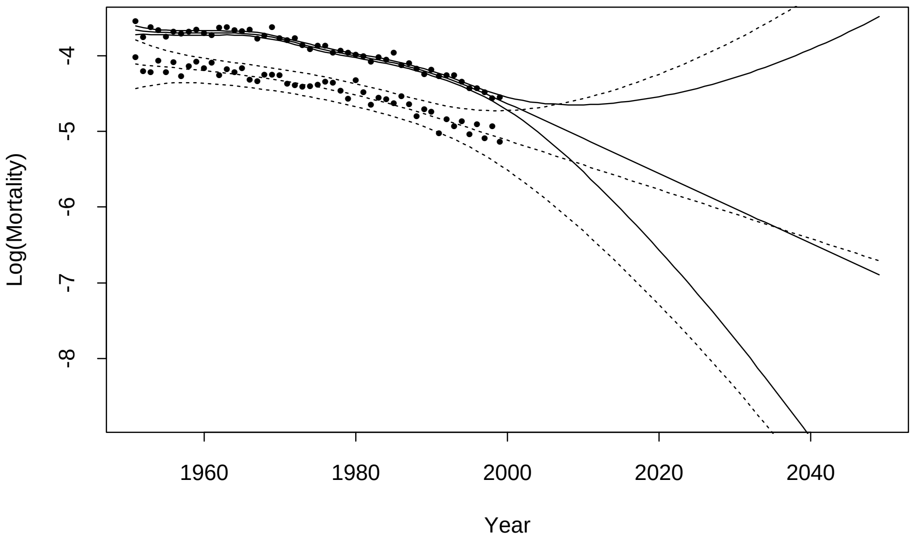

Figure 2: Fitted and projected models of a larger (top) and smaller (bottom) mortality experience. P-spline model with smoothing parameter chosen to favour goodness-of-fit. 95% confidence intervals are shown.

combine them, and try to fit a single model to the aggregated data, but it is obvious that this is inappropriate. A more sophisticated approach would be to move to a richer family of models, in which rates or forces of mortality were parameterised by age and experience instead of by age alone. If we now made mortality projections based on the joint model, we might hope that the uncertainty would be less than in projections based on the separate models for A and B, because we have increased the volume of data. This might fail to work, because the joint model is more complex and this could increase uncertainty, but when the subject of the modelling is human mortality it might be reasonable to hope that it would succeed[5].

Figure 3 shows the result of modelling experiences A and B jointly. The model is very simple, namely:

$$\log \mu_{65}(t) = a + \log \mu_{60}(t) \tag{7}$$

where t is the calendar year and a is a constant; in other words, mortality rates at different ages are parallel across time. (This will automatically prevent μ_{65} from falling below μ_{60}, another useful feature of a suitably chosen joint model.) The two experiences are fitted simultaneously, with appropriate allowance being made for their relative sizes (that is, more weight is given to the larger experience). It can be seen by how much

[5] Note that modelling each experience, gender and select duration separately, as the CMI has usually done, has worked well because the data are so extensive, even after being so subdivided, that good-fitting models can still be found, and goodness-of-fit has been the criterion. Projections are more demanding.

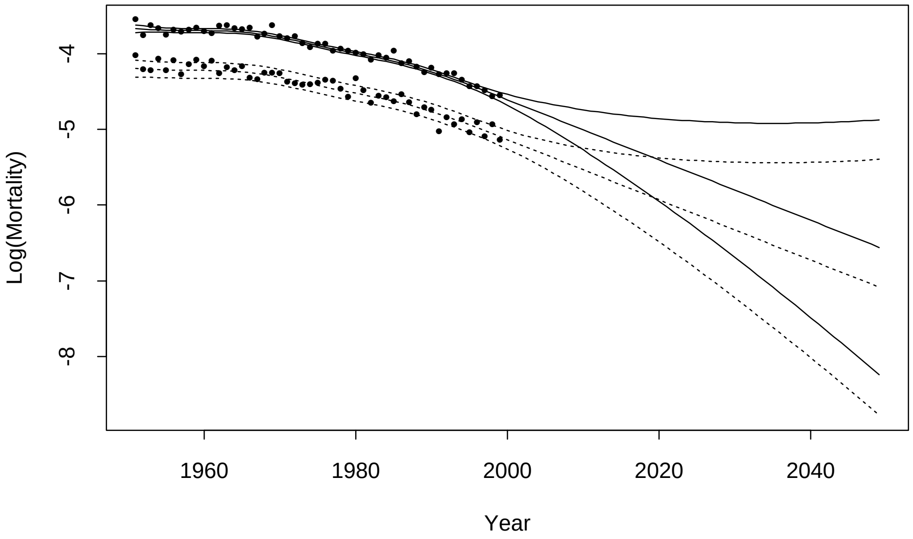

Figure 3: Fitted and projected joint model $\log \mu_{65}(t) = a + \log \mu_{60}(t)$ of a larger (top) and smaller (bottom) mortality experience. P-spline model with smoothing parameter chosen to favour goodness-of-fit. 95% confidence intervals are shown.

the wide confidence intervals in Figure 2 have been reduced. Figure 4 shows the ratio of the standard deviations of the fitted and projected mortality rates (smaller experience divided by larger experience). It is high in the region of the data, and falls to 1 in the region of the projection.

Note that the 2-dimensional P-spline model fitted to the entire surface of estimates $\hat{\mu}(x,t)$ as a function of age and calendar year is a joint model of all the individual experiences $\mu_x(t)$ at each age x, so the somewhat startling contrast shown in Figure 1 is unlikely to be a feature of a fully-implemented model; we are displaying very simple examples here.

A fully joint model requires us to fit larger and smaller experiences simultaneously. A simpler alternative, corresponding to the use of an offset in a generalised linear model, would be to propose some parametric relationship:

$$\log \mu_{60}(t) = f(\theta, \log \hat{\mu}_{65}(t)) \tag{8}$$

where $\hat{\mu}_{65}(t)$ are the previously graduated estimates and θ is a parameter of suitable dimension. This is very close to the idea of graduation by reference to a standard table. We assume that this relationship holds in the region of the projection as well as the region of the data. We model experience A and obtain the projection and its standard errors. We estimate θ and its standard errors. Then (informally) standard errors for experience B are obtained as:

$$(\text{standard error B})^2 = (\text{standard error A})^2 + (\text{standard error } \theta)^2. \tag{9}$$

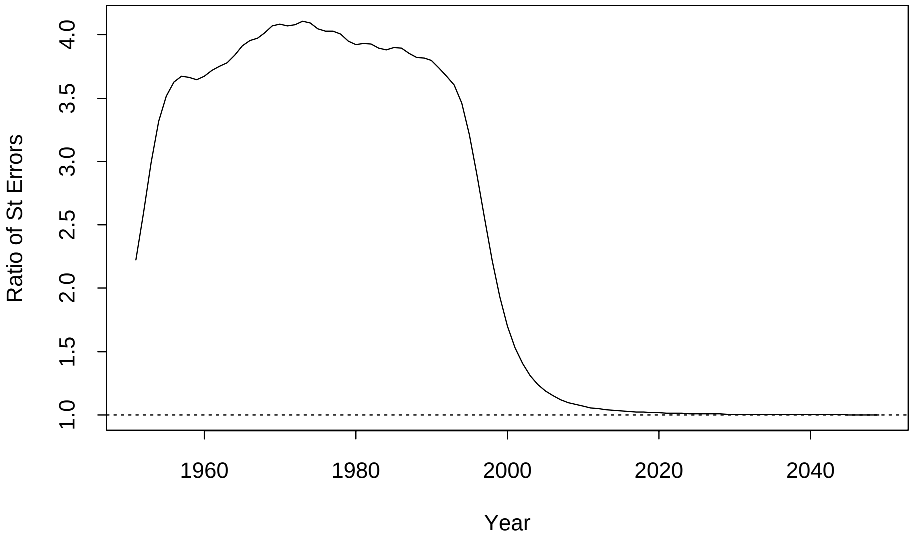

Figure 4: Ratio of standard errors (smaller experience divided by larger experience) of fitted and projected joint model of a larger and smaller mortality experience.

Note that neither of Equations (7) or (8) incorporates cohort effects. The co-ordinates are age and calendar time, and it is features in these two dimensions that will be carried over into the region of the projection. (The same is true of the Lee-Carter model.)

A possible justification for basing projections of CMIB experiences on (say) the U.K. population, therefore runs as follows.

(a) We assume that there is a reasonable joint model for the CMIB experience and the U.K. population experience, or we may treat the U.K. experience as an offset, as above. This model may be impossible to write down explicitly, but we assume that the similarities of different mortality experiences make it plausible that a suitably simple relationship exists.

(b) We then assume that any measures of uncertainty associated with projections based on the joint model or offset model are approximately the same as those associated with projections based on the U.K. population alone. These may therefore be applied to the CMIB experience.

The reader will readily see that some very large assumptions are made above, and may question their statistical validity. We have, however, attempted to make it clear what assumptions may be involved, first, in attaching any meaning to measures of uncertainty, and second, in basing measures of uncertainty on a population other than that being studied.

5. The Choice Between Aggregate or Cause-Specific Mortality Projections

A choice must be made between projecting aggregate mortality (as in the past) or projecting mortality from individual causes separately. In an aggregate approach we project the aggregate mortality directly, whereas in a cause-specific approach we project mortality from different causes or groups of causes and then derive the aggregate mortality rates by combining the cause-specific mortality rates. Both approaches could use any suitable projection methodology (see Section 2).

An advantage of the cause-specific approach is that full account can be taken of information on behavioural and environmental changes as well as expert medical knowledge when projecting mortality from specific causes. This idea underlies the work of the Actuarial Panel on Medical Advances (APMA), for example. However, the aggregate approach is less demanding of the data and is not affected by the unreliability of historic cause of death statistics at the older ages. Tabeau *et al.* (2000) is a useful summary of state-of-the-art developments in cause-specific models.

While the cause-specific approach is obviously needed when investigating the effects of improvements in mortality from specific causes, there are a number of difficulties with using such approaches to project aggregate mortality:

(a) Deaths from specific causes are not always independent and the complex inter-relationships are not always well understood. The same risk factor can affect several causes; for example smoking affects both lung cancer and heart disease; or, the $\epsilon 4$ allele of the APOE gene affects both heart disease and Alzheimer's disease. States of health are very complex particularly at older ages. If these complex inter-relationships were incorrectly modelled, the projected aggregate mortality could be seriously misestimated.

(b) There is limited understanding of how various risk factors affect causes of death, making them difficult to model even at the population level. Even smoking prevalence is not always a good predictor of mortality (for example, in Japan, changing smoking habits have not coincided with changes in mortality in the same way as elsewhere).

(c) Models of several decrements operating simultaneously are known to statisticians as 'competing risks' models, and they have been studied extensively. It is relatively straightforward to estimate the parameters of such a model if the aim is only to *describe* the data, but there are intrinsic problems if the aim is to *model changes in the underlying causes*, because that requires the relationships among them to be modelled explicitly. These relationships are impossible to identify from the data themselves without making assumptions about what they are, known to statisticians as the problem of unidentifiability. This is described in more detail in Appendix B, because it is such a fundamental problem, and because it may be useful to clarify the traditional actuarial treatment of multiple decrement tables.

(d) The proportion of deaths due to a particular cause (the cause structure) shifts over time as a cause appears, peaks and then disappears. If we do not die from something, we must die from something else. The shifts are a result of changing medical/research effort as the relative importance of the causes change. This is why cancer became a major cause of death in developed countries during the 20^{th} century; it was 'unmasked' as infectious diseases were beaten back by medical advances. Therefore, the aggregate projected mortality improvements arising from cause specific approaches will have a tendency to undershoot historic aggregate improvement rates. Combined with limited understanding of the risk factor dynamics, this makes it very hard to model cause specific mortality even at the population level.

We illustrate this with a simple numerical example. Suppose that in 1990, $\mu_{60} = 0.01$, with two causes of death — heart disease and cancer — each accounting for half of the deaths. Between 1990 and 2000, the mortality rate from cancer stays level whereas the mortality rate from heart disease reduces by 5% pa. Therefore μ_{60} in 2000 is:

$$(0.5 \times 0.01) + (0.5 \times 0.01(1 - 0.05)^{10}) = 0.00799. \tag{10}$$

The average rate of improvement in μ_{60} between 1990 to 2000 has been 2.21% pa. Then, assuming mortality from heart disease continues to improve at the same rate, by 2010 μ_{60} is:

$$(0.5 \times 0.01) + (0.5 \times 0.01(1 - 0.05)^{20}) = 0.00679. \tag{11}$$

The average rate of improvement in μ_{60} between 2000 and 2010 has been 1.62% pa. The rate of improvement has reduced because the cause showing the highest past rates of improvement (heart disease) has become less common as a result. Arguably, aggregate projection methodologies are more appropriate as medical research resources, in this example, might be expected to shift from heart disease to cancer as their relative importance changes.

(e) The death of a very old person may have multiple causes, but only one may be recorded on the death certificate, or an incorrect or general cause may be recorded, leading to significant misclassification. For general or ill-defined causes of mortality, there is no objective method of projecting improvements. Also, changing methods of diagnosis and of classification of causes of death reduce still further the reliability of historic cause-specific mortality rates. Therefore, historic cause-specific mortality rates are not as reliable for older ages, reducing the credibility of projections based on them.

(f) There may be causes of mortality at extreme old ages that have not been identified yet as other, known, causes have resulted in deaths at earlier ages. If only the known causes of death are projected, future aggregate mortality would be underestimated. This problem is aggravated by the sparsity of data at older ages.

Given the above difficulties with using a cause-specific approach, particularly projections of mortality at older ages where the CMIB projections are focused, such an approach appears less suitable for the CMIB projections. Further, as a projection of aggregate mortality can still be informed by trends in cause-specific mortality, we believe that this is more suitable for the CMIB projections. This is in agreement with the conclusion reached by the GAD (GAD, 2001).

That does not mean that possible differences in the projected mortality of men and women, and possibly differences between other broad classes, should be neglected. It will be necessary to carry out exploratory modelling to determine if separate projections may be justified.

6. Conclusions and Questions

6.1 *The Status of This Working Paper: An Invitation to Comment*

This Working Paper is intended to be a consultation document, to stimulate thinking within the profession and to invite discussion and responses. It should be clear that the Working Party has not arrived at a single defensible methodology for projecting mortality, for the specific purpose of actuarial risk management. In our view considerably more work needs to be done before any methodology can be recommended for long-term use, but we feel it is appropriate to expose the results of our work to date so that that further work can benefit from the views of others. In the meantime, the interim projections in CMIB Working Paper No. 1 may continue to be used, updated to apply to new base tables based on the 1999–2002 quadrennium if necessary. In the following sections we report our conclusions and ask some questions.

6.2 *Current Practice*

We had available to us the base tables and projections published in the past by the CMIB, and the cohort projections published in Working Paper No. 1. However, we did not have any specific information about how projections of future mortality were used in practice. It would be useful to gather information on current practice. **Question: what base tables and projections do offices use now?**

6.3 *Projecting Aggregate or Cause-Specific Mortality?*

In Section 5 we concluded, as did the GAD, that it was probably impractical to project cause-specific mortality over very long terms, and that aggregate mortality should be projected. That said, possible differences between mens' and womens' projected mortality should be investigated, insofar as the data permit. **Question: what level of aggregation is appropriate in projecting future mortality?**

6.4 *Cohorts or Calendar Years?*

In Section 3 we referred to work that is reported elsewhere, (CMIB Working Paper No. 1; Willets (2004); Willets *et al.* (2004)) and supported the conclusions drawn there that the cohort effect is a prominent feature in the U.K., particularly affecting the recently retired cohort, and it should be reflected in projections. **Question: should we continue to project cohorts?**

6.5 *Measures of Uncertainty I: Are They Needed?*

We found it hard to conceive that projections without associated measures of uncertainty would be acceptable for use in actuarial risk management in future. The FSA is evidently of the same mind. The fact that such techniques are now quite familiar in respect of investment risk could be helpful, but it could also be misleading. **Question: Do we need quantitative measures of uncertainty associated with any projections, and if so, what form should they take?**

6.6 *Measures of Uncertainty II: Interpretation*

We have discussed uncertainty in terms of a probability distribution of future rates of mortality — an 'expanding funnel of doubt' — and such convenient features as standard errors and confidence intervals. These have the merit of being widely understood, but we did not find their interpretation straightforward, and this is perhaps the aspect of projection that we found most difficult. We found it helpful, indeed essential, to locate the discussion in the context of well-specified statistical models (although we treated particular models only in the most general terms; reaching sharper conclusions about model choice should be an aim of further research).

Two quite general classes of model were discussed in Section 4.3: regression models and time series models. We found that regression models, chosen for their nice properties in the region of the data, could behave less well in the region of the projection. Time series models (of which the Lee-Carter model is an example) are specified with projection in mind and may have more predictable behaviour. However, and this is critical, in the absence of a sound approach to quantifying model uncertainty, confidence intervals for the projection reflect parameter uncertainty alone, and this can be greatly reduced by choosing a highly structured model. In other words, the more weight we give to prior belief in the choice of model, the less uncertainty surrounds the projection. We do not believe that this necessarily means we can actually have more confidence in the projections. **Question: are distributions or percentiles of future rates of mortality, derived from statistical models of past rates of mortality, sufficiently meaningful to be used in practice?**

6.7 *Measures of Uncertainty III: Choice of Population*

In Section 4.4 we asked what basis there might be for projecting the future mortality of a large population such as a national population, and applying these projections to other, smaller, populations. This question did not arise in the past, when making deterministic projections. We suggested that well-specified statistical approaches do exist, akin to 'graduation by reference to a standard table', that do justify the use of projections based on large populations, as an approximation at least. **Question: should projections and any measures of uncertainty be based on the largest available appropriate populations?**

6.8 *Potential Methodologies*

We identified regression models, time series models and age-period-cohort models as the major approaches used in the literature. The first of these was represented by P-spline models, and the second by Lee-Carter models, but that was not meant to exclude other models. A specific problem, which is also a priority for future research, is that of adapting existing regression or time series models to project cohorts instead of calendar years. **Question: is there, at this stage, any clearly preferred methodology?**

6.9 *The Financial Consequences of New Projections*

If we are unable to quantify model uncertainty, the choice of methodology, and then the choice of a particular model, may greatly influence the results. The financial consequences of that choice, for the industry and for pensioners, could be substantial and lasting. To some extent these have already begun to emerge, following the publication of the interim cohort projections in CMIB Working Paper No. 1. We believe it is essential that the profession, the FSA and other interested parties fully recognize the extent to which probability statements about mortality projections may, unavoidably, be based on untestable assumptions and model choices. **Question: what may be the financial consequences of allowing for uncertainty in projecting future mortality?**

Acknowledgements

The Working Party is very grateful for much help and advice from Dr Iain Currie, in the Department of Actuarial Mathematics and Statistics at Heriot-Watt University. Dr Currie provided all the figures in the main part of the paper (and Figure 5 in Appendix A).

Section 2 draws freely upon material published in the report by the Government Actuary's Department (GAD, 2001) and we are grateful to the GAD for permission to use it.

References

Alderson, M.R. & Ashwood, F.L. (1985). Projection of mortality rates for the elderly. *Population Trends* **42**, 22–29.

Booth, H. & Tickle, L. (2003). The future aged: New projections of Australia's elderly population. *Macquarie University Research Paper No. 2003/1.*

Brouhns, N., Denuit, M. & Vermunt, J.K. (2002a). A Poisson log-bilinear regression approach to the construction of projected life tables. *Insurance: Mathematics & Economics* **31**, 373–393.

Brouhns, N., Denuit, M. & Vermunt, J.K. (2002b). Measuring the longevity risk in mortality projections. *Mitteilungen der Schweizerische Aktuarvereinigung* **2002**, 105–130.

Cairns, A.J.G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* **27**, 313–330.

Carriere, J.F. (1994). Dependent decrement theory (with discussion). *Transactions of the Society of Actuaries* **46**, 45–74.

CMIB (1990). *Continuous mortality investigation reports No. 10.* The Faculty of Actuaries and Institute of Actuaries.

CMIB (2002). *Working Paper No. 1.* The Faculty of Actuaries and Institute of Actuaries.

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

GAD (2001). *National population projections: Review of methodology for projecting mortality.* NSQR Series No.8, Office of National Statistics, London.

Gavrilov, L. & Gavrilova, N. (2003). *The quest for a general theory of aging and longevity.* Science SAGE KE.

HAYNES, A.T. & KIRTON, R.J. (1953). The financial structure of a life office (with discussion). *Transactions of the Faculty of Actuaries* **21**, 141–218.

LEE, R.D. & CARTER, L. (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association* **87**, 659–671.

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

REDINGTON, F.M. (1952). A review of the principles of life office valuation (with discussion). *Journal of the Institute of Actuaries* **78**, 286–340.

TABEAU, E., VAN DEN BERG JETHS, A. & HEATHCOTE, C. (EDS.) (2001). *Forecasting mortality in developed countries: Insights from a statistical, demographical and epidemiological perspective.* Kluwer Academic Publishers.

TULJAPURKAR, S. & BOE, C. (1998). Mortality change and forecasting: How much and how little do we know? *North American Actuarial Journal*, **2(4)**, 13–47.

WILLETS, R. (2004). The cohort effect: Insights and explanations. *To appear in British Actuarial Journal.*

WILLETS, R.C., GALLOP, A.P., LEANDRO, P.A., LU, J.L.C., MACDONALD, A.S., MILLER, K.A., RICHARDS, S.J., ROBJOHNS, N., RYAN, J.P. & WATERS, H.R. (2004). Longevity in the 21st Century. *To appear in British Actuarial Journal.*

Appendices

A. Summary of the Seminar on 6 October 2003

A.1 *Introduction*

A seminar entitled "Projecting Future Mortality", organised by the CMIB and the GAD, was held on 6 October 2003 in Edinburgh. The seminar was followed by the inaugural Lecture to the Faculty of Actuaries by Professor Tom Kirkwood entitled "Expectations of Life".

The seminar was organized to ensure that the work of the CMIB and the GAD on mortality projections is informed about current research and development in the wider academic world. The project that triggered this requirement is the CMIB's intention to publish tables of mortality based on the mortality experience of the 1999–2002 quadrennium. The seminar brought together experts from related disciplines to provide insight into the following questions:

(a) Should aggregate mortality be projected, or is it necessary to model individual causes of mortality?

(b) What methodology should be chosen to project mortality?

(c) How should a range of possible projections be constructed and communicated?

(d) What, if any, may be the limits of the human lifespan?

The seminar was split into three sessions: "Projecting Aggregate Mortality and Modelling Individual Causes"; "Methodology of Projection and Statistical Methods"; and "Limits on Human Lifespan and Molecular effects on Ageing". Two speakers were invited to present their research in each of these sessions; unfortunately, one had to withdraw at short notice.

A.2 *Projecting Aggregate Mortality and Modelling Individual Causes*

The first speaker, Professor Shripad Tuljapurkar, observed that though mortality improvements were volatile from year to year, there seemed to be a simple pattern of general decline in aggregate mortality for highly industrialised countries. However, mortality improvements could still be reversed as had been seen in Russia. Professor Tuljapurkar advised anyone undertaking mortality projections to take uncertainty seriously. Information on the uncertainty of the projection methodology used could be provided by giving percentiles or other indications of a probability distribution along with a central projection.

He explained that the cause of death structure of mortality was difficult to understand and predict due to poor understanding of causal relationships in the driving forces. Causes of death in the G7 countries varied. He noted the persistent differentials associated with other factors such as social class and income. He suggested that adult mortality should be forecast separately from infant/child mortality as the distribution of deaths by age is heavily influenced by infant mortality.

Professor Tuljapurkar believed that projecting aggregate mortality rather than mortality by cause of death or some other disaggregated approach would produce the best results. He set out the problems with projecting mortality by cause of death:

(a) Mortality risk factors have multiple effects. States of health are very complex making it difficult to estimate the dependence between causes.
(b) There are statistical difficulties with determining dependencies between the causes of death. The various causes of death are usually treated as independent but this is often not the case.
(c) The proportion of deaths due to a particular cause (the cause structure) shifts over time as a cause appears, peaks and then disappears.
(d) There is limited understanding of how various risk factors affect causes of death, making them difficult to model at the population level. Even smoking prevalence is not always a good predictor of mortality (for example in Japan, where changing smoking habits have not coincided with changes in mortality in the same way as elsewhere).
(e) Inaccuracy in the assignment of causes of death reduces the reliability of projections based upon them.

Professor Tuljapurkar also noted that it is much easier to decompose projected mortality into elements associated with particular causes of death (as deviations from the aggregate) than to reproduce the aggregate mortality experience by combining several cause-specific models, and that this is a significant failing because the regularity of aggregate mortality is one of the few major features of the data.

The second speaker, Professor Nico Keilman, started with the comment that the GAD review of methodology for projecting mortality had been sceptical of the benefits of projecting mortality by cause of death for many of the reasons mentioned by Professor Tuljapurkar. However, he believed that empirical evidence demonstrating the superiority of aggregate projection methods was lacking.

Professor Keilman had measured the performance of the UK population projections made by the GAD during the period 1970 to 1998 against the observed outcome over 1970 to 2000. Although not the most useful measure, he used crude death rates as he did not have data on expectations of life. This meant that trends at individual ages or by cohort could not be distinguished.

The general approach to population projections adopted by the GAD used age-specific reductions in death rates for broad groups by age and sex. Some cause of death analysis had been used to inform assumptions used in the 1976 projections. A paper by Alderson & Ashwood (1985) looked at mortality trends at older ages for deaths by heart disease, lung cancer and respiratory diseases. This work had informed the assumptions used in the 1985, 1987 and 1989 population projections. Later projections reverted to a more generalised approach.

Professor Keilman described a model he used to calculate forecast error in the GAD's mortality projections. This was similar to an Age-Period-Cohort model and calculated the absolute error as the sum of the duration, period and forecast effects. The duration effect was modelled using a quadratic function based on the lead time, defined as the period since the base year of the projection. The 1985 projection was used as the reference forecast, and the period from each projection base year to 2000 was used as the reference period.

Measured by means of the absolute errors in the crude death rate, the 1973–83 projections produced relatively inaccurate results. The 1985 forecast had been more accurate. This indicated that cause of death analysis might have made the 1985 forecast better than its immediate predecessors. However, the most accurate forecasts were those of 1970–1972, which did not allow for any cause of death analysis. Thus it was difficult to draw firm conclusions on the matter.

A.3 *Methodology of Projection and Statistical Methods*

Dr Iain Currie presented his work on graduating the CMIB's mortality experiences. He started by giving three health warnings regarding all projections. Projections are affected by the choice of the graduation function, by the forecast/extrapolation function and by random variations. He demonstrated that while the choice of the forecast function made little difference in the short term, the difference could be large at longer durations. The most important point, perhaps, was that no amount of past data absolves the user from the need to choose a projection model.

Commenting on the Lee Carter model, Dr Currie said that it was a very simple model but picked up almost all of the variation in mortality rates. This model is highly structured and has been used very successfully on US data.

Dr Currie gave a brief description of the penalised 2-dimensional splines model he used to model the CMIB's assured lives experience. One-dimensional spline models of mortality by age are familiar to actuaries. This kind of model gives a smooth graduation that is sensitive to local variation in the observed mortality rates. Dr Currie's model extends this idea to 2 dimensions. He uses 2-dimensional splines that give a smooth graduation of the mortality surface that is sensitive to local variation in the observed mortality rates in 2-dimensions. One attractive feature of these models is that cohort effects, although not explicity modelled, will be found if they are a feature of the 2-dimensional surface. The model is expressed as a regression model and roughness in the fitted surface is penalised so that the final graduation is smooth. The smooth surface can be projected and the form of the penalty determines the form of the forecast. For example a quadratic penalty give projections that are essentially linear in time; see Figure 5.

Dr Currie also showed how his model would have fared if the actual experience from 1947 to to 1974 had been used to project mortality rates forward. A comparison of this projection with the experience since 1974 showed that actual experience had been within the 95% confidence intervals. However, the confidence intervals were very wide, especially for relatively younger ages. Figure 5 below shows the 95% confidence intervals at age 65 for Dr Currie's projections to 2050 based on actual experience to 2000.

A.4 *Limits on Human Lifespan and Molecular Effects on Ageing*

Professor Leonid Gavrilov started his talk (which was given jointly with Professor Natalia Gavrilova) by summarising the questions of actuarial significance regarding future improvements in mortality:

(a) How far could rates of mortality decline? He considered a decline to zero as implausible.
(b) Are there any biological limits to mortality decline?
(c) Were there any indications for such biological limits in past data?
(d) Are there any indications that biological limits may apply in the future?

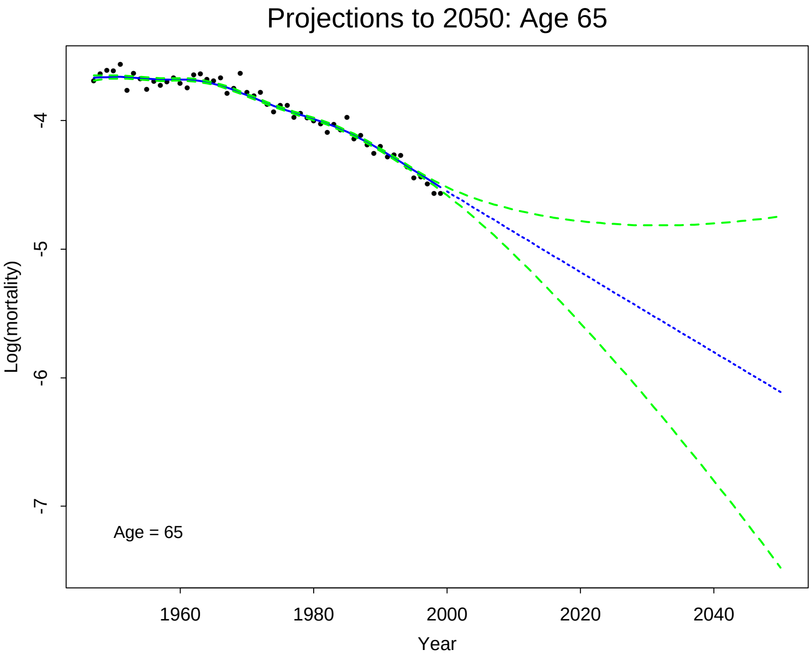

Figure 5: Projection, and 95% confidence intervals, of male Assured Lives $\log \mu_{65}$ to 2050. Based on 2-dimensional P-spline model fitted to the surface $\mu(x,t)$ of mortality rates by age and calendar year.

Borrowing an idea from engineering, he described such biological limits in the context of the reliability of the human body. Reliability theory treats complex, multi-cellular organisms as being similar to a machine built from many low quality components with no pre-testing but with some redundancy amongst those components. Each component can fail with the duration to failure being a random variable. Ageing of the organism is defined as the increasing chance of failure over time. In this model, ageing emerges as a consequence of reducing redundancy amongst the component cells of the organism. This theory also predicts the deceleration observed in late life mortality rates, which seem to stabilise once component redundancy has been exhausted.

Next, Professor Gavrilov described some work he had done using the Gompertz-Makeham model. This model splits mortality into age related and base level factors. He had investigated the changes in these factors using the mortality experience of Scandinavia from 1910 to 1970. The results demonstrated that almost all of the improvements in mortality seen during that period were due to reductions in the base level of mortality rather than reductions in the age related factor.

Professor Gavrilov also highlighted the historical change in survival rates at the oldest ages by reference to the dramatic increase in survival of French and Japanese 90 year olds to age 100. The 10-year survival rate at age 90 had been about 1% from 1900–1960 for both sexes before increasing significantly to 5% for females and 2.5% for males by 2000.

Professor Gavrilov then discussed the ideas and findings of Bruce Ames concerning molecular effects on ageing. Ames had found that the rate of cellular mutation damage could be drastically reduced by the very simple measure of reducing deficiencies in vitamins and other micronutrients. These deficiencies caused greater mutation damage than radiation, industrial pollution and most other hazards. As micronutrient deficiency is common even in the modern wealthy populations, large improvements in mortality at the oldest ages may be achieved by eliminating them.

A further finding of Bruce Ames is one of the mechanisms of DNA damage which is the action of hypochlorous acid (HOCl) produced as a by-product of tissue inflammation. Medical treatments to reduce inflammation may be unintentionally reducing DNA damage in older age groups and this could be part of the reason for the remarkable improvements in mortality at these ages.

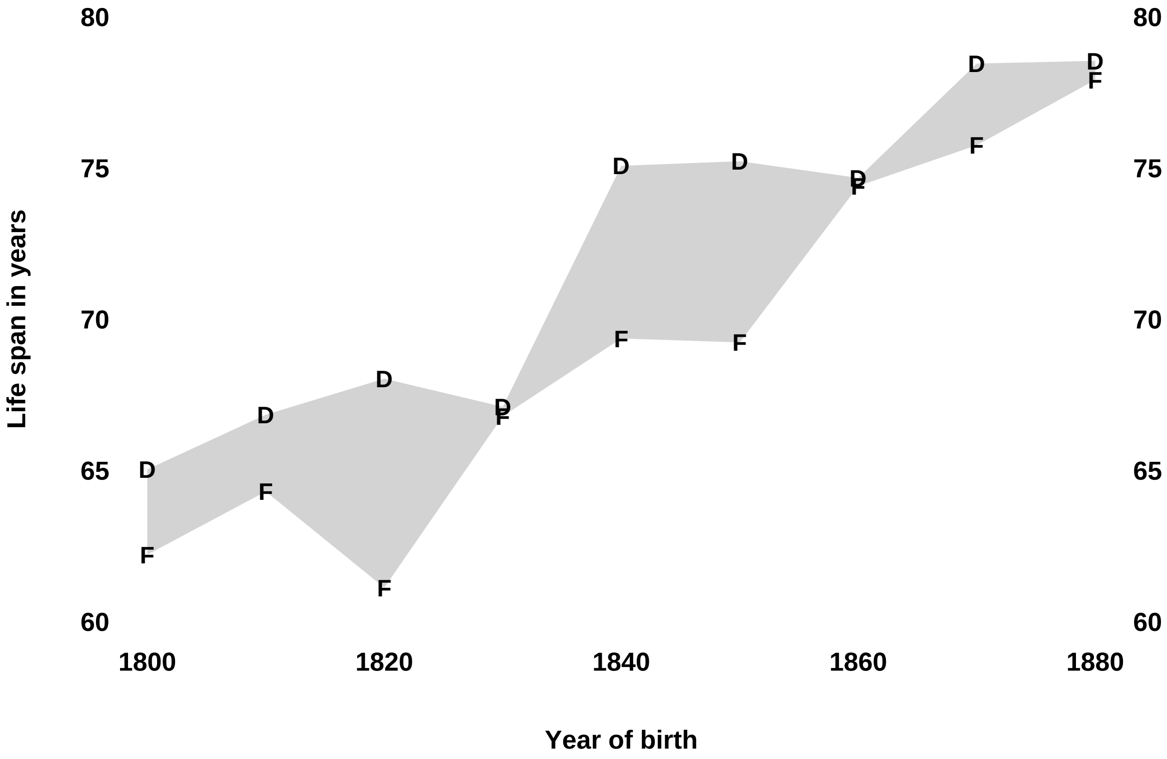

Figure 6: Mean lifetimes in years of females in Professors Gavrilov and Gavrilova's database of the European aristocracy. Mean lifetimes for those born in December (D) and February (F) are plotted against year of birth. The shaded area represents the shortfall in life-span for those born in February, compared with those born in December.

Professor Gavrilov illustrated how lifespan limits can be overcome by new technology using as an example aircraft speed which was initially thought to be limited by the sound barrier. His preliminary conclusions on biological limits to human lifespan were:

(a) There was evidence for 'biological' limits to lifespan in the past data but these may prove to be responsive to recent medical advancements and technological progress.

(b) There is no convincing evidence for thinking that a limit to lifespan exists now.

He then presented some results from his work with Professor Natalia Gavrilova on a genealogical database they had created of female members of the European aristocracy. This covered a 200 year period. Factors found to affect lifespan during that time included month or season of birth and the age of the father at conception. Children of very young or very old fathers had shorter lifespans than average. Noting the interest of actuaries in life expectancy at retirement age, Professor Gavrilov presented results for older women based on this data set: the 'old father' effect had disappeared, yet the 'young father' effect had increased. Figure 6 below shows how the season of birth affected the mean lifespan.

Professor Gavrilov concluded his talk by stressing the role of reliability theory in explaining ageing. Redundancy is a key notion for understanding ageing. Systems which contain redundancy in a number of irreplaceable elements deteriorate over time, even if they are built of non-ageing components. The higher the level of built-in redundancy, the greater the system's failure rate with increased age.

The final speaker, Dr Aubrey de Grey, explained why the ageing process could be reversed in the next 20–30 years and what this meant for those projecting population trends. He started his talk by providing some definitions of ageing:

(a) Senescence: the process that progressively reduces an organism's remaining life expectancy.

(b) Negligible senescence: absence, in a population, of a degree of senescence sufficient to be statistically detectable by examining its age distribution.

(c) Engineered negligible senescence: the biotechnological conversion of a population that exhibits statistically detectable senescence into one that does not.

Dr de Grey explained why technological advances could rapidly produce "engineered negligible senescence" and described a scenario in which this could happen. He felt that the important components of human ageing are effectively all known as no new component had been found since 1982. These components can be described as seven basic types of accumulating damage. Progress in reversing or obviating these types of damage is further advanced for some than others, but such therapies are clearly foreseeable for all seven; three are in clinical trials. As soon as all these methods could be used to reverse the ageing process of laboratory animals, Dr de Grey felt that the funding for research in this area would increase dramatically, quickly leading to clinical trials on humans.

Initial research may not necessarily fully reverse ageing. However, the resulting extended lifespan would allow time for people to further benefit from the continuing research, which would again extend lifespans.

Dr de Grey considered the chance of mortality at old ages increasing from causes that were as yet unknown to be low. This was based on the theory that humans aged x are unlikely to suffer from anything that monkeys aged $x/2$ do not already. The long lead times provided by the increase in lifespan resulting from the initial research would allow time to develop cures for these, as yet unknown, ills.

B. Competing Risks Models

Here we describe an intrinsic and unavoidable problem in attempting to analyse mortality by cause of death. It is well-known to statisticians as the problem of unidentifiability, and it crops up in many other areas, including age-period-cohort models.

It is related to the question, going back to Bernoulli, of how aggregate mortality would be changed if one of its causes were eliminated (he considered smallpox). Actuarial textbooks even up to the present day may give the impression that this problem has a solution. In the actuarial terminology, we start with a multiple-decrement table (one decrement per cause). By definition this shows *dependent* rates of mortality, meaning those rates that are observed in the presence of all decrements operating at once. From these we obtain a set of *independent* rates of mortality, meaning those that would be observed in respect of each decrement operating on its own. We make whatever changes we like to these (eliminating smallpox for example) and then recombine them to get a new multiple decrement table.

The first point to discuss is the unfortunate actuarial terminology. We talk of 'dependent' and 'independent' decrements, which suggests that random variables are involved somewhere, but where are they? Relevant random variables can be defined (see below) and indeed their dependence or independence is important, but this proper statistical usage of these terms is *not* the same as the actuarial usage. This is a source of potential confusion. Statisticians would refer to 'gross' and 'net' decrements instead of 'dependent' and 'independent' decrements, and this is much better.

Taking the simplest case of two decrements labelled α and β, the relevant random variables are T^α and T^β, defined as the times from birth until suffering decrements α and β respectively. They may be dependent (in the statistical sense); this gives precise meaning to the notion of dependent decrements. Let us assume that they are dependent; for example, decrement α might be death from heart disease and decrement β death from lung cancer. Of course we can only ever observe the minimum $\min[T^\alpha, T^\beta]$, and also which of decrements α or β occurred first. This is the crux of the problem: the *model* (called the 'competing risks model') is formulated in terms of *two* random variables, which are fully specified by their joint distribution:

$$F(t^\alpha, t^\beta) = \mathrm{P}(T^\alpha \leq t^\alpha, T^\beta \leq t^\beta) \tag{12}$$

which is what we must estimate. But we cannot observe samples of the bivariate random variable (T^α, T^β), because if we observe T^α we do not observe T^β and *vice versa.* This makes the distribution $F(t^\alpha, t^\beta)$ impossible to estimate, as follows.

First, it is easy to show (or see) that gross forces of mortality are additive:

$$(a\mu)_x = (a\mu)_x^\alpha + (a\mu)_x^\beta \tag{13}$$

regardless of any dependencies between the random variables T^α and T^β. Hence we have:

$$\exp\left(-\int_0^x (a\mu)_t dt\right) = \exp\left(-\int_0^x (a\mu)_t^\alpha dt\right) \exp\left(-\int_0^x (a\mu)_t^\beta dt\right). \tag{14}$$

On the left side is the survival function of the observable minimum, $\min[T^\alpha, T^\beta]$, while the probability that a decrement at age x is α is just $(a\mu)_x^\alpha/(a\mu)_x$. In other words, the distribution of *everything that can be observed* is determined by the gross forces of decrement.

Now define a pair of independent random lifetimes T^δ and T^κ by specifying their respective forces of mortality as follows:

$$(a\mu)_x^\delta = (a\mu)_x^\alpha \qquad \text{and} \qquad (a\mu)_x^\kappa = (a\mu)_x^\beta.$$

Then $F(t^\alpha, t^\beta) \neq F(t^\delta, t^\kappa)$ (since we have assumed T^α and T^β to be dependent) but Equation (14) still holds. It follows that we can never estimate the joint distribution from this kind of data; in fact we cannot even answer the simpler question of whether or not the underlying lifetimes are independent. Collecting more and more data makes no difference, it is impossible. See Macdonald (1996) for a more detailed account.

The only way round this is to make some assumption about the dependencies between the random variables T^α, T^β and any others that may be involved. Sometimes this can be done by assuming a parametric model; if chosen cleverly enough, this may allow the gross forces to identify a unique distribution $F(t^\alpha, t^\beta)$. A popular approach recently has been to use *copulas* to define a dependency structure, see Carriere (1994) for example.

If we are thinking about mortality projections over very long terms, allowing for changes in particular underlying causes, then the issue of dependencies could be important. It is hard to see any way other than some hypothesised dependencies (which could include the special case of independence) but that would make the whole exercise depend on assumptions that the data could never support.