ISSN 2044-3145

Continuous Mortality Investigation

Institute and Faculty of Actuaries

Mortality Projections Committee

WORKING PAPER 91

CMI Mortality Projections Model consultation – technical paper

August 2016

NOTE: This document is being made available publicly and its use is not subject to CMI's Terms and Conditions for Subscribers or Terms and Conditions for Academics and CMI Committee Members.

Contents

1. Introduction

The CMI Mortality Projections Committee has been critically reviewing the CMI Mortality Projections Model ("the Model") and proposes a number of changes. Subject to consultation, these changes would be made in the next version of the Model, CMI_2016, which is planned to be released in March 2017. The proposed changes to the Model were described in Working Paper 90, released in June 2016. An updated version was issued on 19 July 2016, incorporating changes relating to the timing of life expectancies calculated using the proposed model. Section 14 of Working Paper 90 sets out a series of consultation questions. We originally requested responses by 9 September 2016. **We have extended the deadline for responses to 30 September 2016.**

This paper, referred to in Working Paper 90 as the "Technical Working Paper", is intended to be read in conjunction with Working Paper 90 and contains supplementary information.

This paper falls into three broad parts.

The first part of the paper relates to the proposals made in Working Paper 90.

- Section 2 considers the data used to calibrate the Model. We discuss concerns with exposure data, particularly at high ages, and illustrate the method used to adjust exposures.
- Sections 3 and 4 concern the APCI model that we propose to use to determine the initial rates of mortality improvements, and their components. Section 3 contains technical detail of the algorithm used to fit the APCI model. Section 4 sets out how we have determined the fitted parameters and the derived mortality improvements, and considers how the hyperparameters affect the amount of smoothing.
- Section 5 looks at aspects of the projection of mortality improvements, including the tapering of the long-term rate by age, and the difficulty of estimating direction of travel.
- Sections 6 and 7 show how projected life expectancy varies for different assumptions. Section 6 considers the sensitivity of the proposed model to a wide range of assumptions based on data to 31 December 2015. Section 7 places these results in a wider context by comparing them to CMI_2014 and CMI_2015.

The second part of the paper, in Sections 8 to 11, discusses options that the Committee considered as part of its review of the Model but decided not to include in its proposals. These include convergence functions (Section 8), alternative models for initial improvements (Section 9), an "integrated approach" (Section 10) and discussion of why the predictive power of a model was not considered to be an important factor in model choice (Section 11).

Finally, Section 12 describes the software that accompanies this working paper. This software is intended to allow interested parties to replicate the results in Working Paper 90 and this paper, and to consider the impact of particular parameter choices.

2. Exposure adjustment

This section considers the adjustment to exposure data, described in Section 5.9 of Working Paper 90, in more detail.

2.1. Background

The Model currently derives initial rates of improvement from population data for England and Wales from the Office for National Statistics ("ONS"). Actual numbers of registered deaths are divided by mid-year population estimates to derive historical central mortality rates at each age.

It has been observed for some time that the raw mortality improvements derived from these rates contain some unusual features for certain years of birth.

The most obvious of these is around the 1919 and 1920 birth cohorts where plots of historical mortality improvements by age and calendar year highlight a strong diagonal pattern for these lives, seemingly indicating very strong positive or negative mortality improvements compared to those born the year before, or the year after. Similar patterns appear to exist for some other cohorts, for example the 1947 cohort. Some of these year of birth patterns are only noticeable for limited time periods.

The Committee highlighted these issues in its June 2014 presentation to the Staple Inn Actuarial Society[1], observing that they also appear in many other (particularly European) datasets and commenting how they directly contribute to the overdispersion seen when fitting the Model to the historical data.

At that presentation we referred to Cairns et al (2014)[2] in which the authors:

1) Highlighted the concept of how errors in population estimates in census years continue without exposure to decrements and therefore become progressively more significant in subsequent population estimates, particularly at high ages, until the next census year (i.e. the "Phantoms Never Die" reference).
2) Described how the ONS "backfills" over the intercensal period to adjust for discrepancies that they find between the population measured by the census and its previous estimate for the population in the census year.
3) Described a specific issue with the 2001 Census (which might also have affected other censuses) that would cause errors in the mid-year population estimate for that year, and calculated (using the distribution of births) the effect that the error would have had on the population estimates in 2001.
4) Described in general why the mid-year population estimate aged x in year t is not necessarily a good proxy for the central exposure aged x in year t, and derived a method to adjust for this based on the distribution of births in each year.
5) Proposed a set of graphical diagnostics to help identify potential anomalies in any population and death data.
6) Developed an objective technique to try and correct for apparent anomalies in any population data, with or without detail on the underlying distribution of births.

The issues highlighted by Cairns et al appeared to tie in with some of the suspected anomalies that we saw in the ONS data which we used for the Model, and certainly suggested that these strong patterns for individual years are unlikely to be true reflections of the experience. They appear more likely to be effects of the way that population estimates are derived and the implicit assumption when estimating historical population mortality rates that births are evenly-distributed throughout the year.

[1] https://www.actuaries.org.uk/documents/projecting-future-mortality-trends
[2] That paper has since been updated and will be published as Cairns et al (2016).

2.2. CMI_2014

Following the publication of Cairns et al (2014), which gave strong support for our suspicion that some of the historical patterns seen in mortality improvements were artefacts rather than true reflections of experience, it seemed appropriate that we try to allow for these issues in the data used in the Model.

The generic method proposed in (6) above was not particularly simple to implement, or to describe and make available to users, and so for CMI_2014 (and again for CMI_2015) we adopted a simple and transparent method of:

a) fitting the penalised spline (p-spline) model to our data as normal;

b) identifying cells with an absolute deviance residual of more than 3.891 (i.e. significant at the 0.01% level);

c) adjusting the exposure in those cells so that the raw mortality rates match those fitted in a); and

d) re-fitting our p-spline model using these adjusted exposures.

This appeared to reduce some of the anomalies that were observed in the population data, as well as reducing the overdispersion in the final model-fitting. It did, though, feel slightly circular (although it generally only affects cells where we have good reason to doubt the original data) and it potentially misses some of the other issues in other parts of our data. Additionally the method was based on our continued use of the p-spline model, and so the Committee wanted to investigate alternate approaches to adjusting for these data issues.

2.3. Investigatory work

In our presentations in Edinburgh and London in 2015[3], the Committee presented the results of trying to allow explicitly for the issues highlighted in Cairns et al (2014) – i.e. points 3) and 4) in Section 2.1 – by making the adjustments in the way that the authors described.

Although this showed promising results for many historical periods, there were still strong artefacts remaining, particularly for the 1919 and 1920 cohorts from around 2005 onwards. This period also runs in to when these cohorts join the 90+ age group, for which it is suspected there are other issues with the data.

Considering also that some Subscribers had expressed a preference to be able to run the Model against alternative datasets the Committee decided instead to retain a generic approach which requires no detailed knowledge of the issues underlying the dataset being used.

2.4. Proposal

Our proposed model, as with CMI_2014 and CMI_2015, again uses a pragmatic and simple approach to target and "clean" the most extreme-looking cells in any dataset.

The principle of our proposed method is that we expect mortality rates to vary smoothly with age. While there may be some inaccuracy in the deaths data, for example the age recorded at death, we expect deaths data to be much more reliable then exposure data; so any outliers from the assumption of smooth mortality rates suggest a problem with the exposure data.

We start with ONS data consisting of registered deaths $D_{x,t}$ and exposure estimates $E_{x,t}$ for a range of ages x and years t. For each specific combination (X, T) of age and year we want to decide whether to use the existing exposure $E_{X,T}$ or to adjust it.

[3] https://www.actuaries.org.uk/documents/future-cmi-mortality-projections-model

We assume that the smoothed mortality rate $m_{x,T}$ in the age range $[X-n, X+n]$ in year T is exponential (i.e. follows Gompertz's law) and so can be expressed as:

$$\log m_{x,T} = a_{X,T} + b_{X,T}\, x$$

for some parameters $a_{X,T}$ and $b_{X,T}$. We fit these parameters using least squares regression over that age range, to minimise the expression:

$$\sum_{x\in[X-n,X+n]} \left(a_{X,T} + b_{X,T}x - \log\left(\frac{D_{x,T}}{E_{x,T}}\right)\right)^2$$

The approach taken means that our estimate of the smoothed mortality rate $m_{x,T}$ for the specific point that we are considering is given by:

$$\log m_{x,T} = \frac{1}{2n+1}\sum_{x\in[X-n,X+n]} \log\left(\frac{D_{x,T}}{E_{x,T}}\right)$$

We then calculate the deviance residual $r_{X,T}$ (described in Section 3.1) as:

$$r_{X,T} = \text{sign}\left(D_{X,T} - E_{X,T}m_{X,T}\right)\sqrt{2\left(D_{X,T}\log\left(\frac{D_{X,T}}{E_{X,T}m_{X,T}}\right) - \left(D_{X,T} - E_{X,T}m_{X,T}\right)\right)}$$

If our assumption that smoothed mortality is exponential in the age range $[X-n, X+n]$ in year T holds, then we would expect the deviance residual to be Normally-distributed with a mean of zero and a variance of one. If this is not the case, then this suggests a potential problem with exposure data.

We write Φ for the cumulative distribution function of the standard Normal distribution, and specify a probability threshold, p. Then:

- if $|r_{X,T}| \le \Phi^{-1}(1-\frac{p}{2})$ we use the unadjusted exposure $E_{X,T}$.
- However if $|r_{X,T}| > \Phi^{-1}(1-\frac{p}{2})$ we use the adjusted exposure $E'_{X,T} = \frac{D_{X,T}}{m_{X,T}}$.

For ages at and near the edges of the data we need to use a lower value of n, e.g. for ages 21 and 99 we use $n = 1$, and for ages 20 and 100 we make no adjustment.

2.5. Parameterisation and impact

Our proposed approach to adjusting the exposure requires two parameters:

- n, which defines the age range $[X-n, X+n]$ used to determine the smoothed mortality rate; and
- p, the probability threshold used to decide whether or not to adjust an exposure.

There is an element of subjectivity in the choice of values for these parameters; we propose to use $n = 2$ and $p = 1\%$.

We initially considered various combinations of parameter values, and the impacts of some of these are shown in Charts 2A, 2B and 2C below as resulting adjustments to the exposures, and as crude mortality improvements.

A value of 1% for the probability threshold, p, seemed reasonable given that the method is effectively applied to individual calendar years on their own, and the ONS dataset used by the CMI has 81 rates in each year.

Together with a value for n of 2 this combination seemed to meet quite well our intention of targeting the known areas of doubt (i.e. the diagonal patterns around the 1919 and 1920 cohorts, as well as to an extent around the

1947 cohort) without too much effect in the younger and very older areas of the charts which are expected to be naturally more noisy anyway.

Applying the adjustment in this way, to effectively smooth out some of the outliers in the data before fitting our proposed model, obviously makes for a closer fit to the adjusted data and so improves the deviance of the fitted model to the adjusted data.

In this section we illustrate our method using data for ages 20-100 and years 1975-2015 for England & Wales males. The deviances for various combinations of n and p are illustrated in Table 2.1; these compare to a deviance of 9,584 when fitted to the original unadjusted data.

Table 2.1: Deviance for different parameter values (England & Wales males, 20-100, 1975-2015)

	$p = 0.01\%$	$p = 0.1\%$	$p = 1\%$	$p = 10\%$
$n = 2$	7,934	7,720	7,393	6,289
$n = 3$			7,049	
$n = 4$			6,892	

In practice the final impact of the adjustments is actually very minor. The only parts of our fitted historical model that feed into our projections are the initial rates for the final year of the calibration data. Chart 2A illustrates the limited variation in the fitted initial rates from our proposed model for the same combinations of n and p.

Chart 2A: Initial rates for 2015 for various combinations of parameters n and p (England & Wales males, 20-100, 1975-2015)

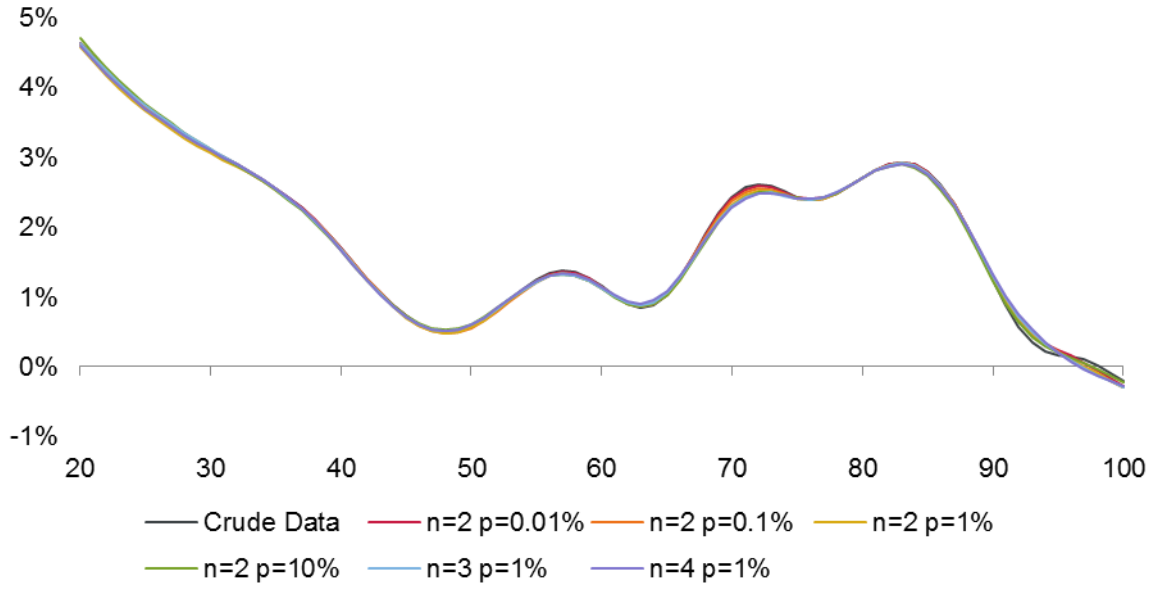


Table 2.2 shows life expectancies at age 65, calculated as at end 31 December 2015, using S2PMA and projected using our proposed model calibrated with the various combinations of n and p and with an illustrative long-term rate of 1.5% p.a. These compare to the base level (i.e. with no data adjustments) of 22.35.

Table 2.2: Life expectancy at age 65 for different parameter values (England & Wales males, 20-100, 1975-2015)

	$p = 0.01\%$	$p = 0.1\%$	$p = 1\%$	$p = 10\%$
$n = 2$	22.36	22.35	22.33	22.32
$n = 3$			22.33	
$n = 4$			22.33	

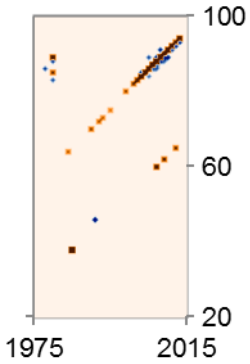
Table 2.2 confirms that the exposure adjustments have very little impact on life expectancies at age 65. In particular the choice of n has a negligible impact. Similarly, life expectancy values for other sample ages (not shown) also show low variation.

In Section 6 we consider the sensitivity of life expectancies to various parameters and model choices, and we consider there the impact of adjusting exposures or using unadjusted values.

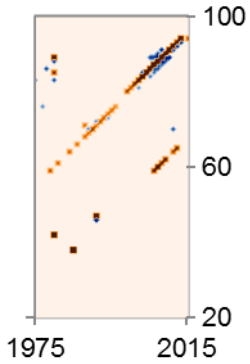
Chart 2B shows which age/year cells have their exposures adjusted, for different choices of n and p; and Chart 2C shows the resulting crude mortality improvements.

Chart 2B: Adjustments made to exposures for different combinations of parameters n and p (England & Wales males, 20-100, 1975-2015)

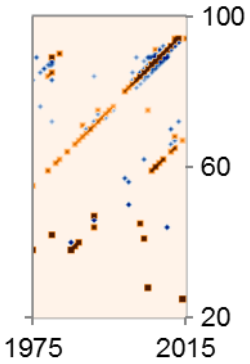
$n = 2;\ p = 0.01\%$



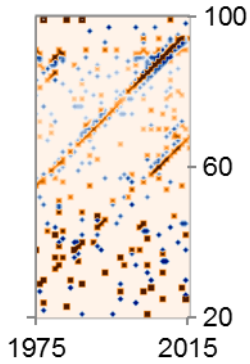
$n = 2;\ p = 0.1\%$



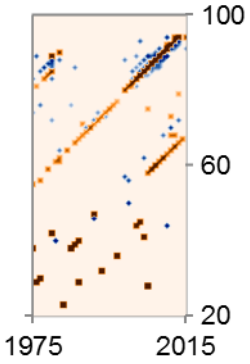
$n = 2;\ p = 1\%$



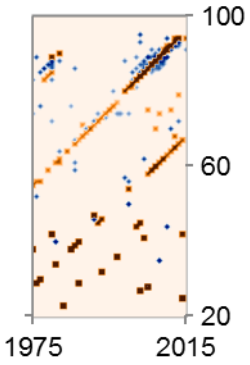
$n = 2;\ p = 10\%$



$n = 3;\ p = 1\%$



$n = 4;\ p = 1\%$



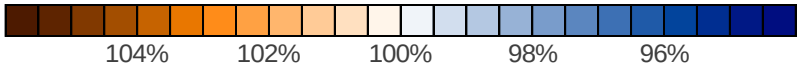
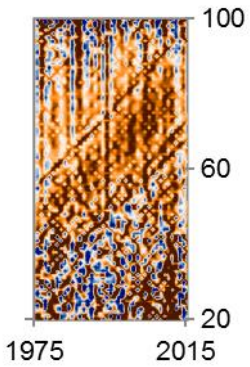
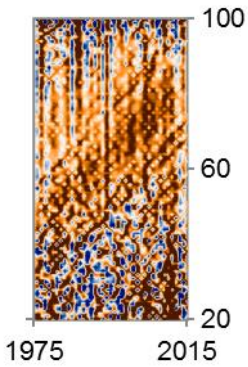


Chart 2C: Crude mortality improvements for different combinations of parameters n and p (England & Wales males, 20-100, 1975-2015)

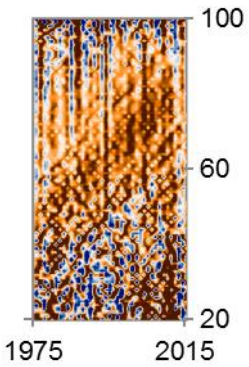
$n = 2;\ p = 0.01\%$



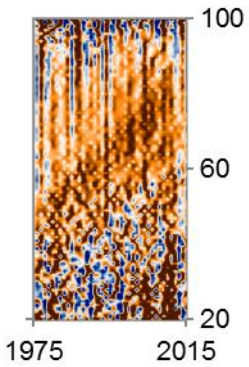
$n = 2;\ p = 0.1\%$



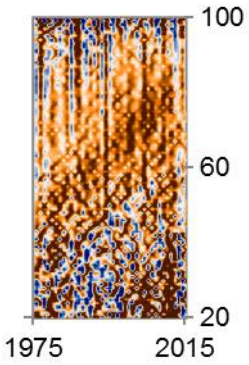
$n = 2;\ p = 1\%$



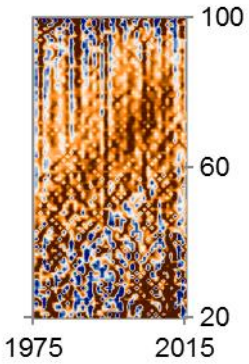
$n = 2;\ p = 10\%$

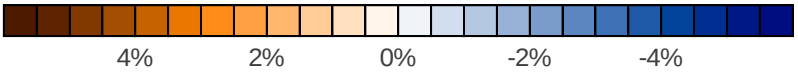


$n = 3;\ p = 1\%$



$n = 4;\ p = 1\%$





2.6. Discussion

The key advantages of the proposed method are that:

- It is based solely on the assumption of a smooth progression of mortality rates by age. There is no assumption of smoothness over time so the method is unaffected by, and should not inadvertently remove, the impact of annual noise.
- It is not dependent on the form of the model used subsequently. In particular, it is helpful to move away from the previous use of a p-spline model for exposure adjustment as we propose not to use such a model for fitting mortality rates.
- It is not affected by the number of years included within the dataset, and only the highest and lowest n ages are affected by the age range used.
- It is quick and easy to apply, and can be replicated by users with spreadsheet formulae.
- It can be applied to any dataset with no knowledge required of the particular data issues.
- If the individual user wishes, the strength of the adjustment can be increased or reduced by adjusting the smoothness parameter (n) and the probability threshold (p).

We note that:

- The simple "locally Gompertz" assumption is not necessarily appropriate at all ages, particularly the youngest ages for males. However the probability threshold used means that the method is unlikely to result in many adjustments at the extreme ages.
- The method should not be seen as an attempt to "correct" the exposure estimates, and hence no attempt has been made to redistribute the adjustments to exposures across other ages. We are simply adjusting the cells with the most extreme and questionable values in order to improve our subsequent model-fitting.
- In fact because we are only targeting, and then adjusting, the most extreme cells, we are aware that the method will result in some discontinuities in the adjusted exposures.
- As the method does not compare an adjusted exposure to that for the same cohort in adjacent years, this can give rise to "false positives" where the adjusted exposure does not seem plausible. An example of this is shown in Table 2.3 where the adjusted exposure for age 25 in 2014 is significantly different to surrounding figures. This occurs due to the unusually low number of deaths for that age and year.

Table 2.3: Adjusted exposure data and deaths for England & Wales males

Age	Adjusted exposure			Deaths		
	2013	2014	2015	2013	2014	2015
24	389,408	400,763	407,895	203	208	204
25	391,983	323,341	403,085	227	182	226
26	380,756	395,587	395,601	226	267	224

The proposed approach is seen simply as an alternative to accurately adjusting the exposures to allow for some of the issues within the data, such as those raised by Cairns et al.

The Committee is still keen to better understand the underlying issues with the ONS dataset and notes that the CMI High Age Mortality Working Party is investigating some of the issues with the data at the highest ages. It is likely that the Committee will revisit this topic to consider it further once their work is complete.

3. APCI model – fitting algorithm

Section 7 of Working Paper 90 describes the Age-Period-Cohort Improvement (APCI) model that we propose to use to calculate initial mortality improvements. The APCI model is defined by:

$$\log m_{x,t} = \alpha_x + \beta_x(t - \bar{t}) + \kappa_t + \gamma_{t-x}$$

where:

x	is age at last birthday
t	is time; i.e. calendar year
$\bar{t}$	is the mean of the years within the calendar year range that is used to fit the model; e.g. if we calibrate to years 1975 to 2015, then $\bar{t}$ is 1995
c	is cohort, with $c = t - x$. Note that this does not correspond exactly to birth year.
α_x	are parameter values for terms by age relating to mortality rates
β_x	are parameter values for fitted terms by age relating to mortality improvements
κ_t	are parameter values for terms by period (i.e. calendar year)
γ_{t-x}	are parameter values for terms by cohort (i.e. birth year)

This section provides full algorithmic detail for the process of fitting the APCI model.

In Section 3.1 we describe deviance, a component of the objective function for the model, described in 3.2. In Section 3.3 we describe the derivation of mortality improvements and the "direction of travel". In Sections 3.4 and 3.5 we describe the iterative fitting process of Newton's method for a general case and for the APCI model. Identifiability is covered in Sections 3.6 and 3.7, and overdispersion in Section 3.8.

3.1. Deviance

This section defines and discusses the deviance statistic that we use to determine the goodness of fit of the APCI model.

If a particular age and year "cell" has exposure $E_{x,t}$ then the expected number of deaths is $E_{x,t}m_{x,t}$. If the actual number of deaths is $D_{x,t}$ then under a Poisson assumption the log-likelihood for that cell is:

$$\mathrm{LL}_{x,t} = D_{x,t}\log E_{x,t}m_{x,t} - E_{x,t}m_{x,t} - \log(D_{x,t}!)$$

and the log-likelihood over the whole of the data is:

$$\mathrm{LL} = \sum_{x,t}\mathrm{LL}_{x,t} = \sum_{x,t}\left(D_{x,t}\log E_{x,t}m_{x,t} - E_{x,t}m_{x,t} - \log(D_{x,t}!)\right)$$

The log-likelihood reaches a maximum for the "saturated model" with a parameter for every observation, so that $D_{x,t} = E_{x,t}m_{x,t}$ for each age and year. In this case we have for one cell:

$$\mathrm{LL}_{x,t}^{sat} = D_{x,t}\log D_{x,t} - D_{x,t} - \log(D_{x,t}!)$$

and for the whole of the data:

$$\mathrm{LL}^{sat} = \sum_{x,t}\mathrm{LL}_{x,t}^{sat} = \sum_{x,t}\left(D_{x,t}\log D_{x,t} - D_{x,t} - \log(D_{x,t}!)\right)$$

Note that in the case where $D_{x,t} = 0$ we have $\mathrm{LL}_{x,t}^{sat} = 0$.

The deviance is defined as twice the difference between the actual log-likelihood and the log-likelihood for the saturated model; i.e.:

$$\mathrm{Deviance}_{x,t} = 2\left(\mathrm{LL}_{x,t}^{sat} - \mathrm{LL}_{x,t}\right) = 2(D_{x,t}\log D_{x,t} - D_{x,t} - D_{x,t}\log E_{x,t}m_{x,t} + E_{x,t}m_{x,t})$$

and:

$$\mathrm{Deviance} = \sum_{x,t}\mathrm{Deviance}_{x,t} = 2\sum_{x,t}\left(D_{x,t}\log D_{x,t} - D_{x,t} - D_{x,t}\log E_{x,t}m_{x,t} + E_{x,t}m_{x,t}\right)$$

The deviance can be expressed as the sum of squares of deviance residuals:

$$\text{Deviance} = \sum_{x,t} \text{DevianceResidual}_{x,t}^2$$

Where:

$$\text{DevianceResidual}_{x,t} = \text{sign}\left(D_{x,t} - E_{x,t}m_{x,t}\right)\sqrt{2\left(D_{x,t}\log D_{x,t} - D_{x,t} - D_{x,t}\log E_{x,t}m_{x,t} + E_{x,t}m_{x,t}\right)}$$

As deviance has a linear relationship with log-likelihood, minimising the deviance is equivalent to maximising the log-likelihood (i.e. it will give the same fitted parameters).

3.2. The objective function

For the purpose of the Model, we want to extract the underlying trends in mortality improvements and smooth out short-term fluctuations (e.g. due to winter temperatures and infectious diseases) and artefacts of the data. To achieve this we define an objective function that is a combination of the deviance (as a measure of goodness of fit) and penalty functions (as a measure of the smoothness of each set of parameters).

We have:

$$\text{Objective} = \text{Deviance} + \text{Penalty}(\alpha_x) + \text{Penalty}(\beta_x) + \text{Penalty}(\kappa_t) + \text{Penalty}(\gamma_{t-x})$$

where the penalties are as described in Section 7.4 of Working Paper 90:

$$\text{Penalty}(\alpha_x) = \lambda_\alpha \sum_x (\nabla_x^3 \alpha_x)^2 = \lambda_\alpha \sum_x (\alpha_x - 3\alpha_{x-1} + 3\alpha_{x-2} - \alpha_{x-3})^2$$

$$\text{Penalty}(\beta_x) = \lambda_\beta \sum_x (\nabla_x^3 \beta_x)^2 = \lambda_\beta \sum_x (\beta_x - 3\beta_{x-1} + 3\beta_{x-2} - \beta_{x-3})^2$$

$$\text{Penalty}(\kappa_t) = \lambda_\kappa \sum_t (\nabla_t^2 \kappa_t)^2 = \lambda_\kappa \sum_t (\kappa_t - 2\kappa_{t-1} + \kappa_{t-2})^2$$

$$\text{Penalty}(\gamma_c) = \lambda_\gamma \sum_c (\nabla_c^3 \gamma_c)^2 = \lambda_\gamma \sum_c (\gamma_c - 3\gamma_{c-1} + 3\gamma_{c-2} - \gamma_{c-3})^2$$

and the hyperparameters λ_α, λ_β, λ_κ and λ_γ can be used to control the smoothness of the parameters to which they relate.

3.3. Mortality improvements and direction of travel

Under our definition of mortality improvements, described in Section 6 of Working Paper 90:

$$MI_{x,t} = \log m_{x,t-1} - \log m_{x,t}$$

so:

$$MI_{x,t} = -\beta_x + \kappa_{t-1} - \kappa_t \ + \gamma_{t-1-x} - \gamma_{t-x}$$

These aggregate improvement are then decomposed into age, period and cohort components:

$$MI_{x,t} = MI_{x,t}^{Age} + MI_{x,t}^{Period} + MI_{x,t}^{Cohort}$$

where:

$$MI_{x,t}^{Age} = -\beta_x$$

$$MI_{x,t}^{Period} = \kappa_{t-1} - \kappa_t$$

$$MI_{x,t}^{Cohort} = \gamma_{t-x-1} - \gamma_{t-x}$$

"Direction of travel" is defined as:

$$DoT_{x,t} = \ MI_{x,t}^{Period} - MI_{x,t-1}^{Period}$$

i.e.:

$$DoT_{x,t} = \ -\kappa_t \ + 2\kappa_{t-1} - \kappa_{t-2}$$

3.4. Newton's method for a general function

We minimise our objective function by using Newton's method. This is an iterative approach: we repeatedly adjust the parameters to improve the objective function, and stop when the objective function stabilises.

We first consider the generic case, where we have a function f of multiple parameters $(\phi_1, \dots \phi_n)$ that we want to minimise.

A necessary condition for $(\phi_1, \dots \phi_n)$ to be a minimum of f is that $\frac{\partial f}{\partial \phi_i}(\phi_1, \dots \phi_n) = 0$ for all $i = 1 \dots n$.

The general form of a first-order multivariate Taylor series approximation to a function g is:

$$g(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n) \approx g(\phi_1, \dots \phi_n) + \Delta\phi_1 \frac{\partial g}{\partial \phi_1}(\phi_1, \dots \phi_n) + \cdots \Delta\phi_n \frac{\partial g}{\partial \phi_n}(\phi_1, \dots \phi_n)$$

Substituting $\frac{\partial f}{\partial \phi_1}$ to $\frac{\partial f}{\partial \phi_n}$ for g in turn gives the n Taylor series approximations:

$$\frac{\partial f}{\partial \phi_1}(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n) \approx \frac{\partial f}{\partial \phi_1}(\phi_1, \dots \phi_n) + \Delta\phi_1 \frac{\partial^2 f}{\partial \phi_1 \partial \phi_1}(\phi_1, \dots \phi_n) + \cdots \Delta\phi_n \frac{\partial^2 f}{\partial \phi_1 \partial \phi_n}(\phi_1, \dots \phi_n)$$

...

$$\frac{\partial f}{\partial \phi_n}(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n) \approx \frac{\partial f}{\partial \phi_n}(\phi_1, \dots \phi_n) + \Delta\phi_1 \frac{\partial^2 f}{\partial \phi_n \partial \phi_1}(\phi_1, \dots \phi_n) + \cdots \Delta\phi_n \frac{\partial^2 f}{\partial \phi_n \partial \phi_n}(\phi_1, \dots \phi_n)$$

These can be expressed in matrix form as:

$$\begin{bmatrix} \frac{\partial f}{\partial \phi_1}(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n) \\ \vdots \\ \frac{\partial f}{\partial \phi_n}(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n) \end{bmatrix} \approx \begin{bmatrix} \frac{\partial f}{\partial \phi_1}(\phi_1, \dots \phi_n) \\ \vdots \\ \frac{\partial f}{\partial \phi_n}(\phi_1, \dots \phi_n) \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 f}{\partial \phi_1 \partial \phi_1}(\phi_1, \dots \phi_n) & \dots & \frac{\partial^2 f}{\partial \phi_1 \partial \phi_n}(\phi_1, \dots \phi_n) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \phi_n \partial \phi_1}(\phi_1, \dots \phi_n) & \dots & \frac{\partial^2 f}{\partial \phi_n \partial \phi_n}(\phi_1, \dots \phi_n) \end{bmatrix} \begin{bmatrix} \Delta\phi_1 \\ \vdots \\ \Delta\phi_n \end{bmatrix}$$

Now if all of $\Delta\phi_1, \dots \Delta\phi_n$ satisfy:

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \approx \begin{bmatrix} \frac{\partial f}{\partial \phi_1}(\phi_1, \dots \phi_n) \\ \vdots \\ \frac{\partial f}{\partial \phi_n}(\phi_1, \dots \phi_n) \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 f}{\partial \phi_1 \partial \phi_1}(\phi_1, \dots \phi_n) & \dots & \frac{\partial^2 f}{\partial \phi_1 \partial \phi_n}(\phi_1, \dots \phi_n) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \phi_n \partial \phi_1}(\phi_1, \dots \phi_n) & \dots & \frac{\partial^2 f}{\partial \phi_n \partial \phi_n}(\phi_1, \dots \phi_n) \end{bmatrix} \begin{bmatrix} \Delta\phi_1 \\ \vdots \\ \Delta\phi_n \end{bmatrix}$$

then $\frac{\partial f}{\partial \phi_1}(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n), \dots \frac{\partial f}{\partial \phi_n}(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n)$ will all be approximately zero. So starting from $(\phi_1, \dots \phi_n)$, we expect that $(\phi_1 + \Delta\phi_1, \dots \phi_n + \Delta\phi_n)$ will be closer to a minimum of f.

This gives an iterative procedure, the multivariate version of Newton's method, for optimising f.

3.5. Newton's method for the APCI model

In our implementation we will update each set of parameters separately, rather than updating them all in one step (i.e. we update the α_x, then the β_x, then the κ_t, then the γ_{t-x}). This simplifies the algebra and computer code, while still converging quickly.

In pseudocode we have (using subscripts L and H for the lowest and highest values of an index):

1. Initialise the procedure:
   - 1a. Initialise all parameters: α_x, β_x, κ_t and γ_{t-x}.
   - 1b. Calculate mortality rates based on the initial parameters
   - 1c. Calculate the objective function
2. Do repeatedly, until the objective function stabilises:
   - 2a. Calculate $(\Delta\alpha_{x_L}, \dots \Delta\alpha_{x_H})$ and adjust the parameters to $(\alpha_{x_L} + \Delta\alpha_{x_L}, \dots \alpha_{x_H} + \Delta\alpha_{x_H})$
   - 2b. Calculate updated mortality rates
   - 2c. Calculate $(\Delta\beta_{x_L}, \dots \Delta\beta_{x_H})$ and adjust the parameters to $(\beta_{x_L} + \Delta\beta_{x_L}, \dots \beta_{x_H} + \Delta\beta_{x_H})$
   - 2d. Calculate updated mortality rates
   - 2e. Calculate $(\Delta\kappa_{t_L}, \dots \Delta\kappa_{t_H})$ and adjust the parameters to $(\kappa_{t_L} + \Delta\kappa_{t_L}, \dots \kappa_{t_H} + \Delta\kappa_{t_H})$
   - 2f. Calculate updated mortality rates
   - 2g. Calculate $(\Delta\gamma_{c_L}, \dots \Delta\gamma_{c_H})$ and adjust the parameters to $(\gamma_{c_L} + \Delta\gamma_{c_L}, \dots \gamma_{c_H} + \Delta\gamma_{c_H})$
   - 2h. Calculate updated mortality rates
   - 2i. Update parameters to allow for identifiability
   - 2j. Calculate the objective function
3. Calculate mortality improvements

We will consider the case of updating the α_x (i.e. step 2a in the pseudocode above) in detail, and state the analogous results for other parameters.

In order to implement Newton's method for the APCI model we need to be able to calculate

$$\frac{\partial \text{Objective}}{\partial \alpha_i}$$

and

$$\frac{\partial^2 \text{Objective}}{\partial \alpha_i \partial \alpha_j}$$

for all α_i and α_j.

Since:

$$\text{Objective} = \text{Deviance} + \text{Penalty}(\alpha_x) + \text{Penalty}(\beta_x) + \text{Penalty}(\kappa_t) + \text{Penalty}(\gamma_{t-x})$$

we have:

$$\frac{\partial \text{Objective}}{\partial \alpha_i} = \frac{\partial \text{Deviance}}{\partial \alpha_i} + \frac{\partial \text{Penalty}(\alpha_x)}{\partial \alpha_i} + \frac{\partial \text{Penalty}(\beta_x)}{\partial \alpha_i} + \frac{\partial \text{Penalty}(\kappa_t)}{\partial \alpha_i} + \frac{\partial \text{Penalty}(\gamma_{t-x})}{\partial \alpha_i}$$

but, because $\text{Penalty}(\beta_x)$, $\text{Penalty}(\kappa_t)$ and $\text{Penalty}(\gamma_{t-x})$ are not affected by the α_x, this simplifies to:

$$\frac{\partial \text{Objective}}{\partial \alpha_i} = \frac{\partial \text{Deviance}}{\partial \alpha_i} + \frac{\partial \text{Penalty}(\alpha_x)}{\partial \alpha_i}$$

Similarly:

$$\frac{\partial^2 \text{Objective}}{\partial \alpha_i \partial \alpha_j} = \frac{\partial^2 \text{Deviance}}{\partial \alpha_i \partial \alpha_j} + \frac{\partial^2 \text{Penalty}(\alpha_x)}{\partial \alpha_i \partial \alpha_j}$$

Deviance terms

Using the chain rule we have:

$$\frac{\partial \text{Deviance}_{x,t}}{\partial \alpha_i} = \frac{\partial \text{Deviance}_{x,t}}{\partial m_{x,t}} \frac{\partial m_{x,t}}{\partial \log m_{x,t}} \frac{\partial \log m_{x,t}}{\partial \alpha_i}$$

Since:

$$\text{Deviance}_{x,t} = 2(D_{x,t} \log D_{x,t} - D_{x,t} - D_{x,t} \log E_{x,t} m_{x,t} + E_{x,t} m_{x,t})$$

we have:

$$\frac{\partial \text{Deviance}_{x,t}}{\partial m_{x,t}} = 2\left(E_{x,t} - \frac{D_{x,t}}{m_{x,t}}\right)$$

Also:

$$\frac{\partial m_{x,t}}{\partial \log m_{x,t}} = m_{x,t}$$

so:

$$\frac{\partial \text{Deviance}_{x,t}}{\partial \alpha_i} = 2(E_{x,t} m_{x,t} - D_{x,t}) \frac{\partial \log m_{x,t}}{\partial \alpha_i}$$

For the APCI model:

$$\frac{\partial \log m_{x,t}}{\partial \alpha_i} = 1 \quad \text{if } x = i \quad \text{and 0 otherwise}$$

So:

$$\frac{\partial Deviance}{\partial \alpha_i} = 2\sum_{x,t|x=i}(E_{x,t}m_{x,t} - D_{x,t}) \qquad \text{where the sum is over those cells where } x = i$$

as it only involves those cells where $x = i$.

For the other parameters we have similarly:

$$\frac{\partial Deviance}{\partial \beta_i} = 2\sum_{x,t|x=i}\left(E_{i,t}m_{i,t} - D_{i,t}\right)(t - \bar{t}) \qquad \text{where the sum is over those cells where } x = i \text{ and the term } (t - \bar{t}) \text{ arises from } \frac{\partial \log m_{x,t}}{\partial \beta_i}$$

$$\frac{\partial Deviance}{\partial \kappa_i} = 2\sum_{x,t|t=i}\left(E_{x,i}m_{x,i} - D_{x,i}\right) \qquad \text{where the sum is over those cells where } t = i$$

$$\frac{\partial Deviance}{\partial \gamma_i} = 2\sum_{x,t|t-x=i}\left(E_{x,t}m_{x,t} - D_{x,t}\right) \qquad \text{where the sum is over those cells where } t - x = i$$

Turning to the second-order derivatives:

$$\frac{\partial^2 \text{Deviance}_{x,t}}{\partial \alpha_i \partial \alpha_j} = 2\frac{\partial}{\partial \alpha_j}\left((E_{x,t}m_{x,t} - D_{x,t})\frac{\partial \log m_{x,t}}{\partial \alpha_i}\right) = 2\frac{\partial}{\partial \alpha_j}\left(E_{x,t}m_{x,t} - D_{x,t}\right)\frac{\partial \log m_{x,t}}{\partial \alpha_i} + 2(E_{x,t}m_{x,t} - D_{x,t})\frac{\partial^2 \log m_{x,t}}{\partial \alpha_i \partial \alpha_j}$$

Applying the chain rule again:

$$\frac{\partial}{\partial \alpha_j}\left(E_{x,t}m_{x,t} - D_{x,t}\right) = \frac{\partial}{\partial m_{x,t}}\left(E_{x,t}m_{x,t} - D_{x,t}\right)\frac{\partial m_{x,t}}{\partial \log m_{x,t}}\frac{\partial \log m_{x,t}}{\partial \alpha_j} = E_{x,t}m_{x,t}\frac{\partial \log m_{x,t}}{\partial \alpha_j}$$

For the APCI model:

$$\frac{\partial^2 \log m_{x,t}}{\partial \alpha_i \partial \alpha_j} = 0$$

so we have:

$$\frac{\partial^2 \text{Deviance}_{x,t}}{\partial \alpha_i \partial \alpha_j} = 2E_{x,t}m_{x,t}\frac{\partial \log m_{x,t}}{\partial \alpha_i}\frac{\partial \log m_{x,t}}{\partial \alpha_j}$$

So:

$$\frac{\partial^2 Deviance}{\partial \alpha_i \partial \alpha_j} = 0 \qquad \text{if } i \neq j$$

and

$$\frac{\partial^2 Deviance}{\partial {\alpha_i}^2} = 2\sum_{x,t|x=i} E_{i,t}m_{i,t}$$

Similarly for the other parameters:

$$\frac{\partial^2 Deviance}{\partial {\beta_i}^2} = 2\sum_{x,t|x=i} E_{x,t}m_{x,t}\,(t - \bar{t})^2 \qquad \text{where the term } (t - \bar{t}) \text{ arises from } \frac{\partial \log m_{x,t}}{\partial \beta_i}$$

$$\frac{\partial^2 Deviance}{\partial {\kappa_i}^2} = 2\sum_{x,t|t=i} E_{x,t}m_{x,t}$$

$$\frac{\partial^2 Deviance}{\partial {\gamma_i}^2} = 2\sum_{x,t|t-x=i} E_{x,t}m_{x,t}$$

Penalty terms

Again we will focus on the case of α_x, and then state the analogous results for other parameters.

The penalty function is:

$$\text{Penalty}(\alpha_x) = \lambda_\alpha \sum_x (\nabla_x^3 \alpha_x)^2 = \lambda_\alpha \sum_x (\alpha_x - 3\alpha_{x-1} + 3\alpha_{x-2} - \alpha_{x-3})^2$$

It is helpful to write this in matrix form as:

$$\text{Penalty}(\alpha_x) = \lambda_\alpha \underline{\alpha}^T \underline{D_\alpha}^T \underline{D_\alpha} \underline{\alpha}$$

where $\underline{\alpha}$ is a vector of the parameters α_x, and $\underline{D_\alpha}$ is the difference matrix:

$$\underline{D_\alpha} = \begin{bmatrix} +1 & -3 & +3 & -1 & 0 & 0 & 0 & \\ 0 & +1 & -3 & +3 & -1 & 0 & 0 & \\ 0 & 0 & +1 & -3 & +3 & -1 & 0 & \\ & & & & & & & \ddots \end{bmatrix}$$

This has size $(N-3) \times N$ where N is the size of $\underline{\alpha}$; i.e. the number of ages in the calibration data.

The penalty matrix $\underline{D_\alpha}^T \underline{D_\alpha}$ has size $N \times N$ and is:

$$\underline{D_\alpha}^T \underline{D_\alpha} = \begin{bmatrix} +1 & -3 & +3 & -1 & 0 & 0 & 0 & 0 & & \\ -3 & +10 & -12 & +6 & -1 & 0 & 0 & 0 & & \\ +3 & -12 & +19 & -15 & +6 & -1 & 0 & 0 & & \\ -1 & +6 & -15 & +20 & -15 & +6 & -1 & 0 & & \\ 0 & -1 & +6 & -15 & +20 & -15 & +6 & -1 & & \\ 0 & 0 & -1 & +6 & -15 & +20 & -15 & +6 & & \\ 0 & 0 & 0 & -1 & +6 & -15 & +20 & -15 & & \\ 0 & 0 & 0 & 0 & -1 & +6 & -15 & +20 & \ddots & \\ & & & & & & & & & +1 \end{bmatrix}$$

The elements of:

$$\frac{\partial \text{Penalty}(\alpha_x)}{\partial \alpha_i}$$

are given by:

$$2\lambda_\alpha \underline{D_\alpha}^T \underline{D_\alpha} \underline{\alpha}$$

and the elements of:

$$\frac{\partial^2 \text{Penalty}(\alpha_x)}{\partial \alpha_i \partial \alpha_j}$$

are given by:

$$2\lambda_\alpha \underline{D_\alpha}^T \underline{D_\alpha}$$

A similar result holds for derivatives of the penalty functions for the other parameters. For the κ_t terms, which have a second-order penalty function, the difference matrix is:

$$\underline{D_\kappa} = \begin{bmatrix} +1 & -2 & +1 & 0 & 0 & 0 & \\ 0 & +1 & -2 & +1 & 0 & 0 & \\ 0 & 0 & +1 & -2 & +1 & 0 & \\ & & & & & & \ddots \end{bmatrix}$$

3.6. Identifiability

There are multiple sets of parameters that could give exactly the same value for $\log m_{x,t}$ and hence the deviance. Specifically, the following transformations leave the values of $\log m_{x,t}$ unchanged for any values of $\theta_1, \dots, \theta_5$:

$$\alpha_x \longmapsto \alpha_x + \theta_1 - \theta_2(x - \bar{x}) + \theta_3(x - \bar{x})^2 + \theta_4$$
$$\beta_x \longmapsto \beta_x - 2\theta_3(x - \bar{x}) + \theta_5$$
$$\kappa_t \longmapsto \kappa_t + \theta_2(t - \bar{t}) + \theta_3(t - \bar{t})^2 - \theta_4 - \theta_5(t - \bar{t})$$
$$\gamma_c \longmapsto \gamma_c - \theta_1 - \theta_2(c - \bar{c}) - \theta_3(c - \bar{c})^2$$

So that the parameter values are uniquely determined, we use the following five identifiability constraints:

$\sum_t \kappa_t = \sum_t t\kappa_t = 0$ i.e. a linear fit to κ_t would be zero for all years t

$\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = 0$ i.e. a quadratic fit to γ_c would be zero for all cohorts c.

It would be possible to implement these constraints as part of the objective function, using Lagrangian multipliers. However we have found that doing so makes convergence extremely slow. Instead we allow for the identifiability constraints by making explicit adjustments to the parameters (in step 2i of the pseudocode).

The steps are:

(a) quadratic regression of γ_c against $c - \bar{c}$ to determine values of θ_1, θ_2 and θ_3.

(b) make the adjustments relating to θ_1, θ_2 and θ_3.

(c) linear regression of κ_t, after the adjustments in step (b), against $t - \bar{t}$ to determine θ_4 and θ_5.

(d) make the adjustments relating to θ_4 and θ_5.

For step (a) define:

$$E = \sum_c(\theta_1 + \theta_2(c - \bar{c}) + \theta_3(c - \bar{c})^2 - \gamma_c)^2$$

If we choose parameters θ_1, θ_2 and θ_3 to minimise E then we will make a quadratic fit to γ_c identically equal to zero. To minimise E we require that its partial derivatives with respect to the parameters that we are fitting are all zero:

$$\frac{\partial \mathrm{E}}{\partial \theta_1} = \frac{\partial \mathrm{E}}{\partial \theta_2} = \frac{\partial \mathrm{E}}{\partial \theta_3} = 0$$

We have:

$$\frac{\partial \mathrm{E}}{\partial \theta_1} = 2\sum_c(\theta_1 + \theta_2(c - \bar{c}) + \theta_3(c - \bar{c})^2 - \gamma_c)^2$$
$$\frac{\partial \mathrm{E}}{\partial \theta_2} = 2\sum_c\big(\theta_1(c - \bar{c}) + \theta_2(c - \bar{c})^2 + \theta_3(c - \bar{c})^3 - \gamma_c(c - \bar{c})\big)^2$$
$$\frac{\partial \mathrm{E}}{\partial \theta_3} = 2\sum_c(\theta_1(c - \bar{c})^2 + \theta_2(c - \bar{c})^2 + \theta_3(c - \bar{c})^4 - \gamma_c(c - \bar{c})^2)^2$$

We can express the requirement that all of these are zero in matrix form as:

$$\begin{bmatrix} \sum_c(c - \bar{c})^0 & \sum_c(c - \bar{c})^1 & \sum_c(c - \bar{c})^2 \\ \sum_c(c - \bar{c})^1 & \sum_c(c - \bar{c})^2 & \sum_c(c - \bar{c})^3 \\ \sum_c(c - \bar{c})^2 & \sum_c(c - \bar{c})^3 & \sum_c(c - \bar{c})^4 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \sum_c \gamma_c(c - \bar{c})^0 \\ \sum_c \gamma_c(c - \bar{c})^1 \\ \sum_c \gamma_c(c - \bar{c})^2 \end{bmatrix}$$

and then solve for the values of θ_1, θ_2 and θ_3.

Step (c) is similar. We solve the matrix equations:

$$\begin{bmatrix} \sum_t (t-\bar{t})^0 & \sum_t (t-\bar{t})^1 \\ \sum_t (t-\bar{t})^1 & \sum_t (t-\bar{t})^2 \end{bmatrix} \begin{bmatrix} \theta_4 \\ \theta_5 \end{bmatrix} = \begin{bmatrix} \sum_t \kappa_t (t-\bar{t})^0 \\ \sum_t \kappa_t (t-\bar{t})^1 \end{bmatrix}$$

for the values of θ_4 and θ_5.

3.7. Identifiability and the objective function

The identifiability transforms have no impact on the values of $\log m_{x,t}$ and so have no impact on the deviance. They also have no impact on the penalty functions for α_x, β_x, and γ_{t-x} as they use third-order differences and the transforms only involve second-order terms.

However the identifiability transforms do affect the penalty function for κ_t slightly. This means that when applying the iterative fitting process described in Section 3.4, the value of the objective function can rise slightly before converging. This is illustrated in Table 3.1. It shows the objective function falling for the first 40 iterations, and later rising slightly. We consider the impact of this to be minor and do not take any further action to address it.

Table 3.1: Change in deviance, penalty, and objective, by iteration (every ten, and final)

Iteration	Deviance	Penalty	Objective
0	51296859.84490	0.00000	51296859.84490
10	7657.84617	644.65329	8302.49946
20	7414.68540	676.42404	8091.10945
30	7396.36453	684.95146	8081.31599
40	7394.02202	687.03116	8081.05318
50	7393.57911	687.52884	8081.10795
60	7393.48025	687.64735	8081.12761
70	7393.45711	687.67554	8081.13265
80	7393.45163	687.68225	8081.13388
90	7393.45033	687.68384	8081.13417
93 (final)	7393.45019	687.68401	8081.13420

3.8. Overdispersion

Under the Poisson assumption for deaths, we expect the deviance to be equal to the number of degrees of freedom. In practice we see overdispersion; i.e. the deviance is higher than expected.

In CMI_2014 and CMI_2015 we made an allowance for overdispersion when fitting the p-spline model, by using the Quasi-Bayesian Information Criterion (QBIC) to determine the optimal amount of smoothing. For the proposed model the degree of smoothing is controlled through the hyperparameters, S. These can be considered to incorporate an implicit allowance for overdispersion, and so there is no need to make any explicit additional allowance for overdispersion. (This does however suggest that different hyperparameters might be appropriate if the method were applied to datasets with materially different amount of overdispersion.)

4. APCI model – parameters and smoothing

An important feature of the way that we use the APCI model is that we have hyperparameters (described in Section 3.2) that we use to control the smoothness of its fitted parameters and hence the smoothness of the age, period and cohort components of mortality improvements.

In this section we consider the values of the parameters of the APCI model and the resulting mortality improvements, and their sensitivities to choices for the smoothing parameters. We consider the proposed Core assumptions for the smoothing parameters ($S_\alpha = 7$, $S_\beta = 9$, $S_\kappa = 7.5$ and $S_\gamma = 7$; where $S_i = \log_{10} \lambda_i$) and the impact of changing these. The impact on life expectancies is considered in Section 6.

The results in this section are all calibrated to data for England & Wales for ages 20-100 and calendar years 1975-2015, with exposures adjusted as described in Section 5.9 of Working Paper 90.

4.1 Age components

Chart 4A shows the values of α_x for males and females for different choices of the age smoothing parameter S_α when the other smoothing parameters S_β, S_κ and S_γ keep their Core values. We show the effect of no smoothing[4], the Core assumption of $S_\alpha = 7$, and values of S_α that are one higher and one lower than the Core assumption.

Chart 4A: Impact of varying S_α on α_x

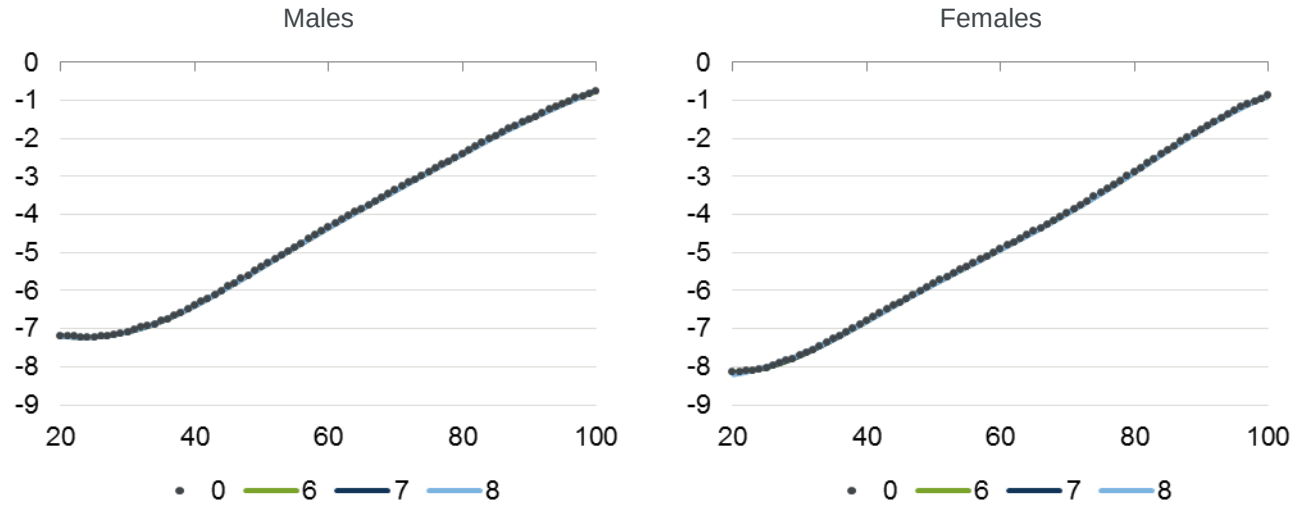
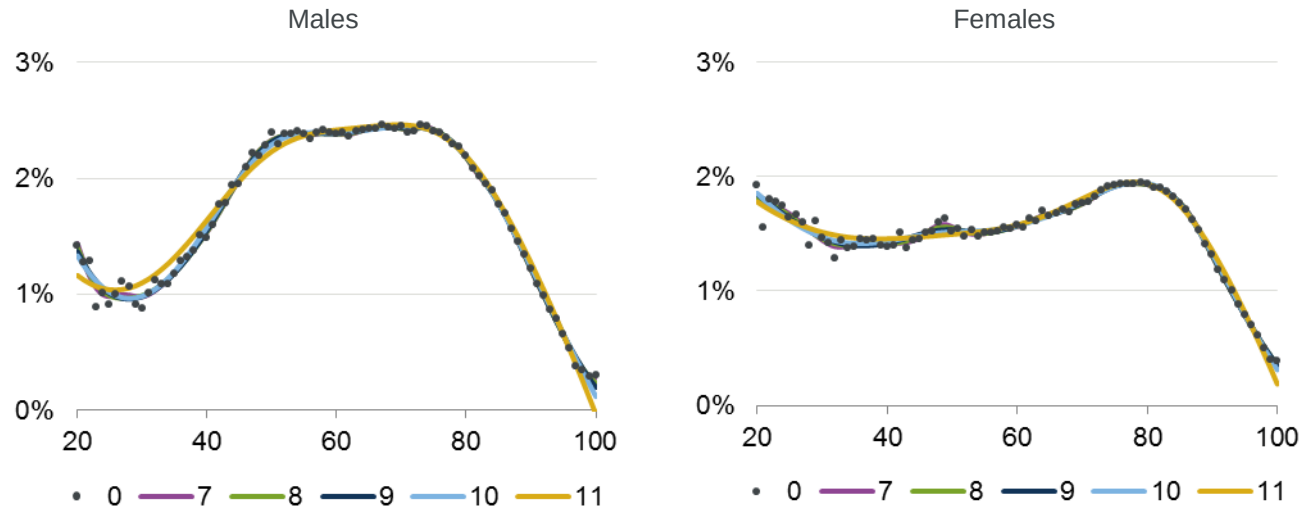


Chart 4A shows a plausible pattern for log-mortality by age, increasing roughly linearly for much of the age range, with some flattening at young and very-old ages. For α_x the choice of smoothing parameter seems largely unimportant, and it is hard to distinguish by eye between the smoothed and unsmoothed cases.

Chart 4B shows values of minus β_x, which corresponds directly to the age component of mortality improvements. For both males and females the improvements fall towards zero at the oldest ages. Compared to Chart 4A there is a bigger difference between the smoothed and unsmoothed parameters, but there is little visual difference between the values of S_β illustrated, except that $S_\beta = 11$ seems to over-smooth.

[4] By "no smoothing", we mean $\lambda_\alpha = 0$ rather than $S_\alpha = 0$. The latter would correspond to $\lambda_\alpha = 1$ and give rise to a very small amount of smoothing.

Chart 4B: Impact of varying S_β on *minus* β_x i.e. the age component of mortality improvements

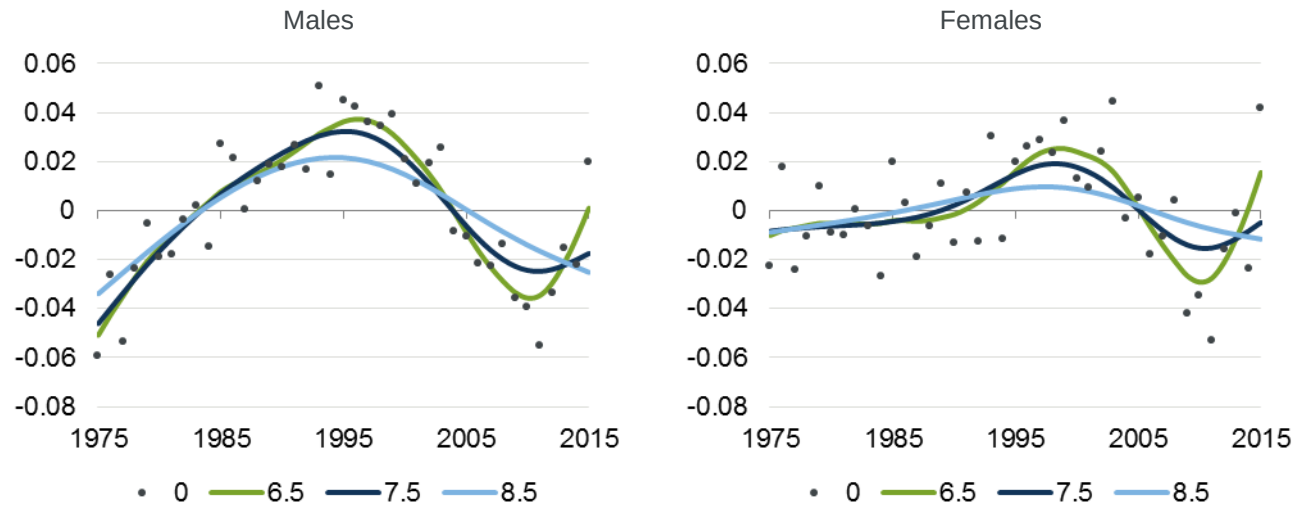


As Chart 4B suggests, and as shown in Section 6, the choice of S_β has little impact on life expectancy. Our choice of $S_\beta = 9$ is based on closer inspection of the shapes of mortality improvements which suggests that using a value of S_β lower than 9 may under-smooth, particularly for the 50-80 age range, whilst S_β above 10 over-smooths across the whole age range.

4.2 Period components

Chart 4C shows the period parameters, κ_t, for different choices of the period smoothing parameter, S_κ.

Chart 4C: Impact of varying S_κ on κ_t



While Chart 4C is of some interest in its own right, it is instructive to consider Chart 4D, which shows the impact of the smoothing parameter S_κ on the period components of mortality improvements, which are derived from κ_t.

While the pattern of improvements by age (in Chart 4B) is clear even before applying any smoothing, the improvements by period (Chart 4D) show considerable volatility from year to year. This is due to events such as cold or mild winters and the extent of infectious diseases such as influenza.

Chart 4D: Impact of varying S_κ on the period component of mortality improvements

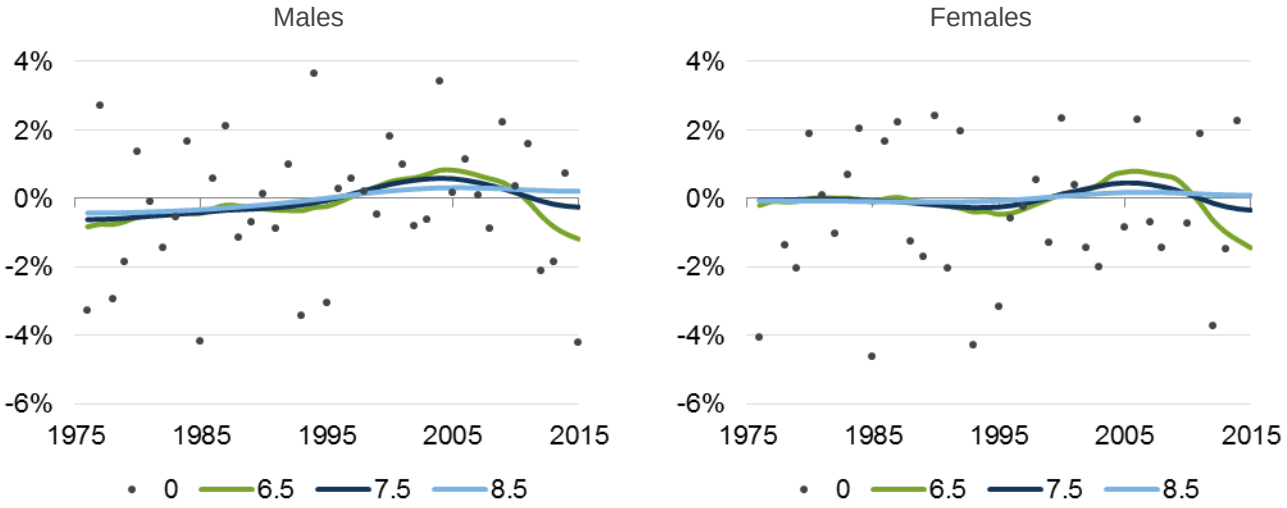
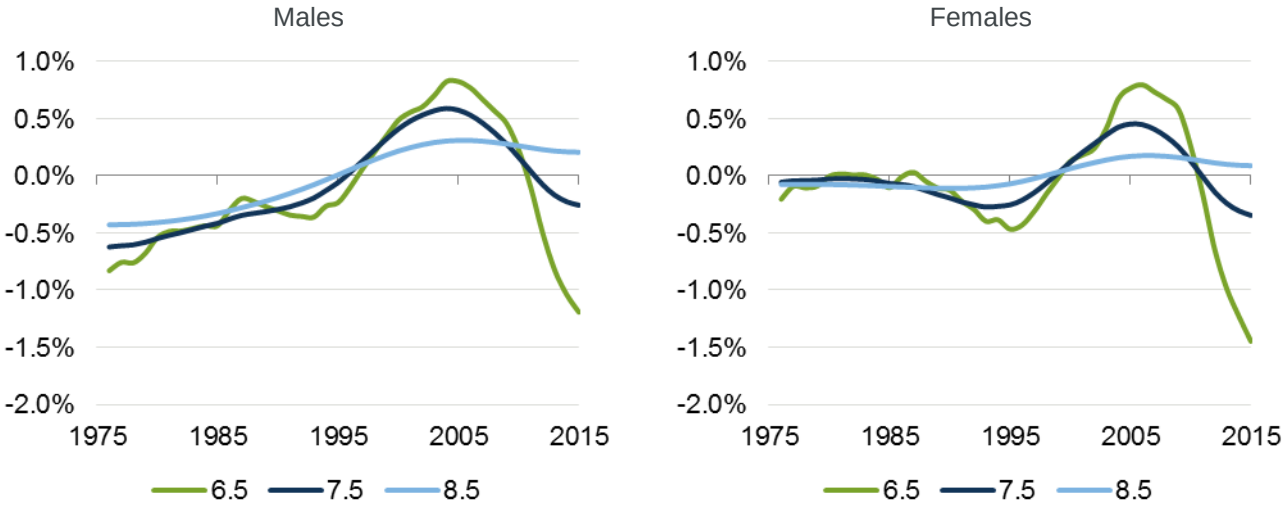


Chart 4E shows the same results as Chart 4D but excludes the unsmoothed case. This allows the *y*-axis to be expanded to show more detail of the smoothed parameters.

Chart 4E: Impact of varying S_κ on the period component of mortality improvements (alternative scale)



Our motivation for the choice of 7.5 as the Core parameter value is discussed in Section 9.3 of Working Paper 90. We consider that a value of 7.5 provides an appropriate degree of responsiveness to new data. e.g.:

- A value of 7 or less would give rise to a fall in life expectancy that is greater than under the current CMI method, and we perceive that a majority of users think this is too responsive.
- A value of 8 would produce improvements for females that are marginally higher in 2015 than in 2011, despite the unprecedented low improvements of 2011-2015.

4.3 Cohort components

Chart 4F shows the cohort parameters γ_{t-x}. Charts 4G and 4H show the impact of the smoothing parameter S_γ on the values of the cohort components of mortality improvements, with different *y*-axis scales.

(Note that when $S_\gamma = 0$ in Chart 4G, the 1875 cohort for males, for which we only have one observation, has a value of −21% that is off the scale of the chart).

Chart 4F: Impact of varying S_γ on γ_{t-x}

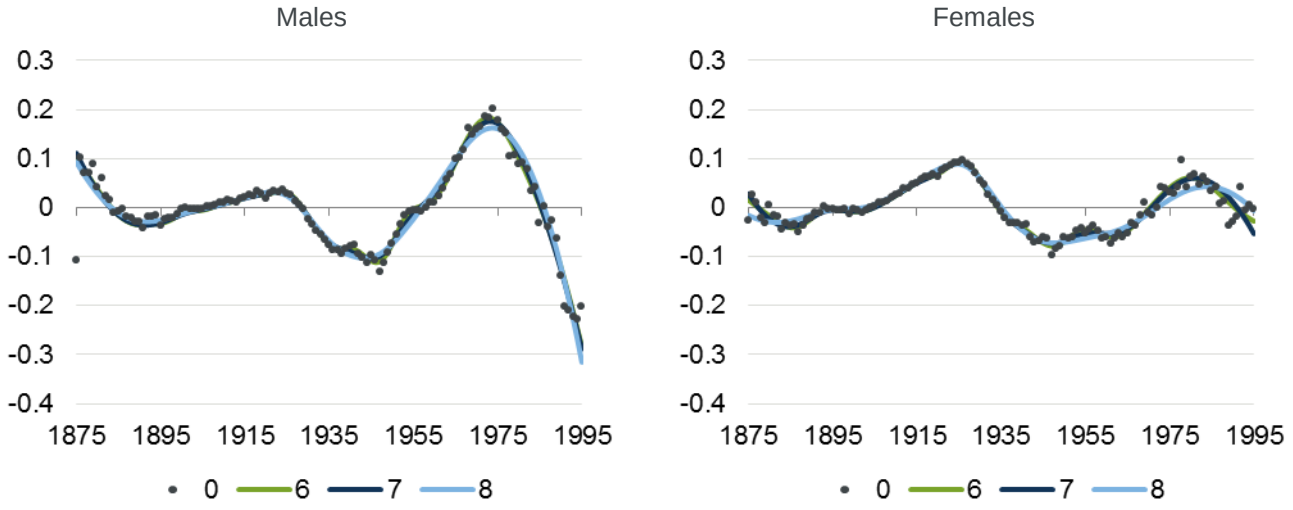


Chart 4G: Impact of varying S_γ on the cohort component of mortality improvements

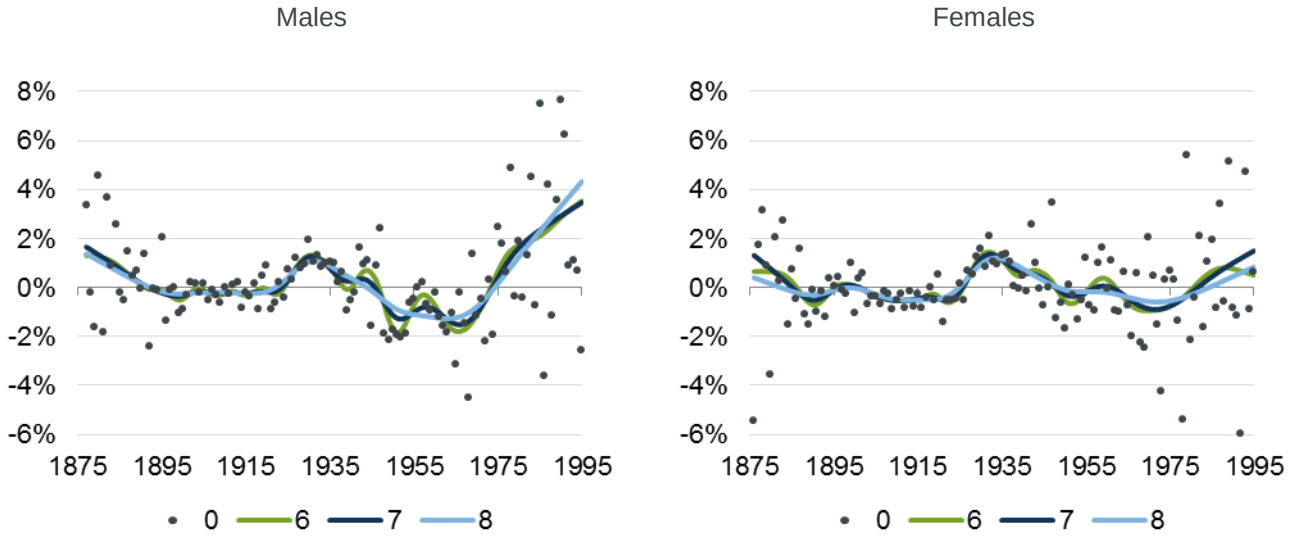
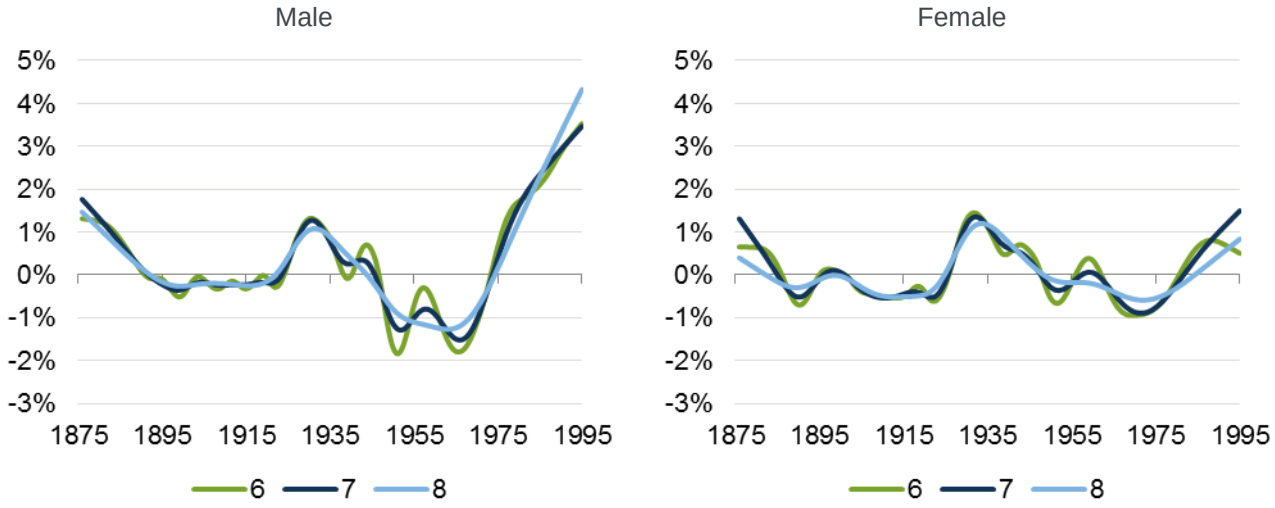


Chart 4H: Impact of varying S_γ on the cohort component of mortality improvements (alternative scale)

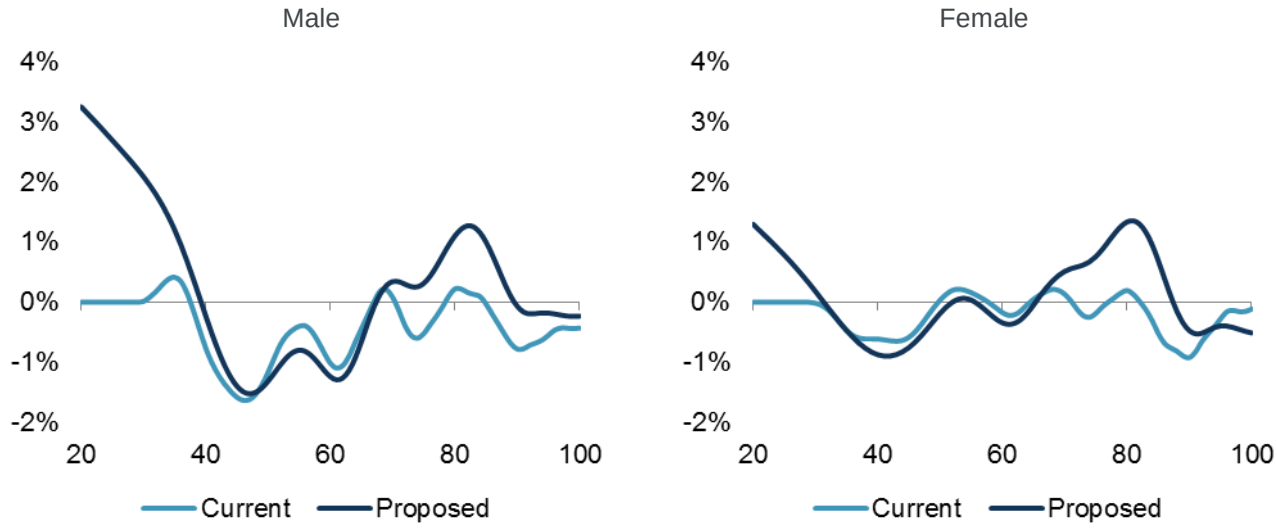


There is some subjectivity over the choice of S_γ, which controls the strength of cohort features in the Model.

Chart 4H shows that a choice of $S_\kappa = 8$ seems to over-smooth; for example Willets (2004) notes "two 'sub-cohorts' of the 1925 to 1945 cohort" and setting $S_\kappa = 8$ would remove the peak for the 1945 cohort (shown at age 70). Conversely a choice of 6 seems to under-smooth; producing cohort improvements that would be larger than under the current Model.

Chart 4I compares the cohort components under the current method and the proposed method (with parameter values of $S_\alpha = 7$, $S_\beta = 9$, $S_\kappa = 7.5$ and $S_\gamma = 7$). The cohort components are noticeably higher under the proposed approach at the youngest ages (due to the different identifiability constraints) and at around age 80 (where the total mortality improvements are higher). However the overall level of smoothing seems similar between the two approaches; i.e. the sizes of the peaks and troughs from ages 40 to 70 look broadly similar.

Chart 4I: Cohort components (by current age) in the current and proposed models



4.4 Impact of varying S_κ on all parameters

In the previous section we considered the impact on a set of parameters of changing its own smoothing hyperparameter. We may also see a "knock-on" effect. For example if we change S_κ then this will affect the values and smoothness of the parameters κ_t and may also affect the other parameters α_x, β_x and γ_{t-x} as it is the combination of the four sets of parameters that is used to fit mortality rates.

In this section we focus on S_κ, given its importance in controlling the responsiveness of the Model, and consider how all four sets of parameters vary when we change it. This is shown in Charts 4J to 4M.

Chart 4J: Impact of varying S_κ on α_x

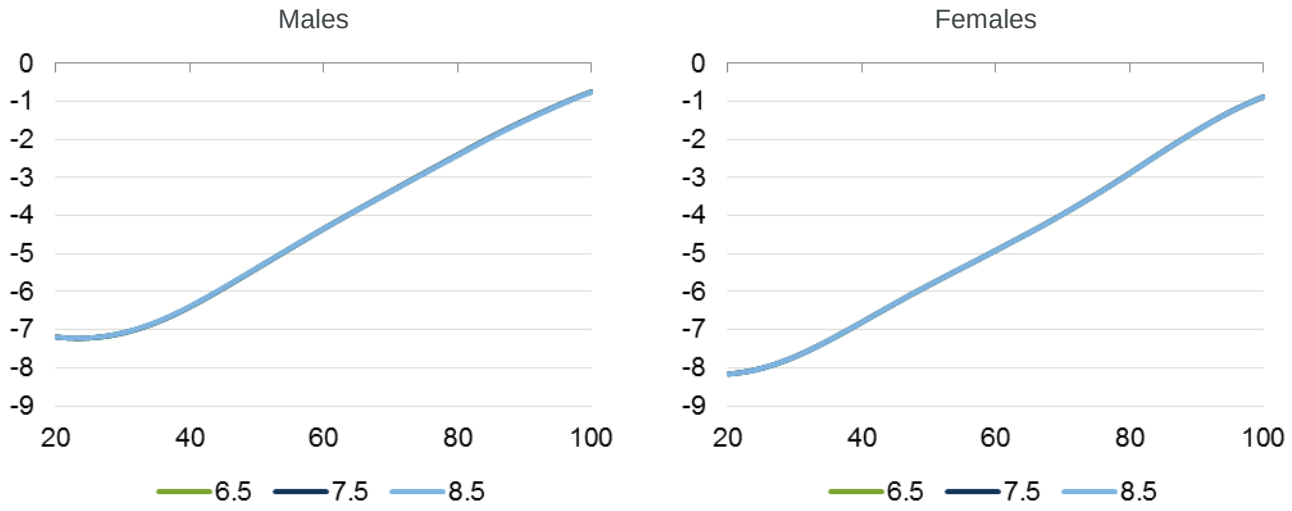


Chart 4K: Impact of varying S_κ on *minus* β_x i.e. the age component of mortality improvements

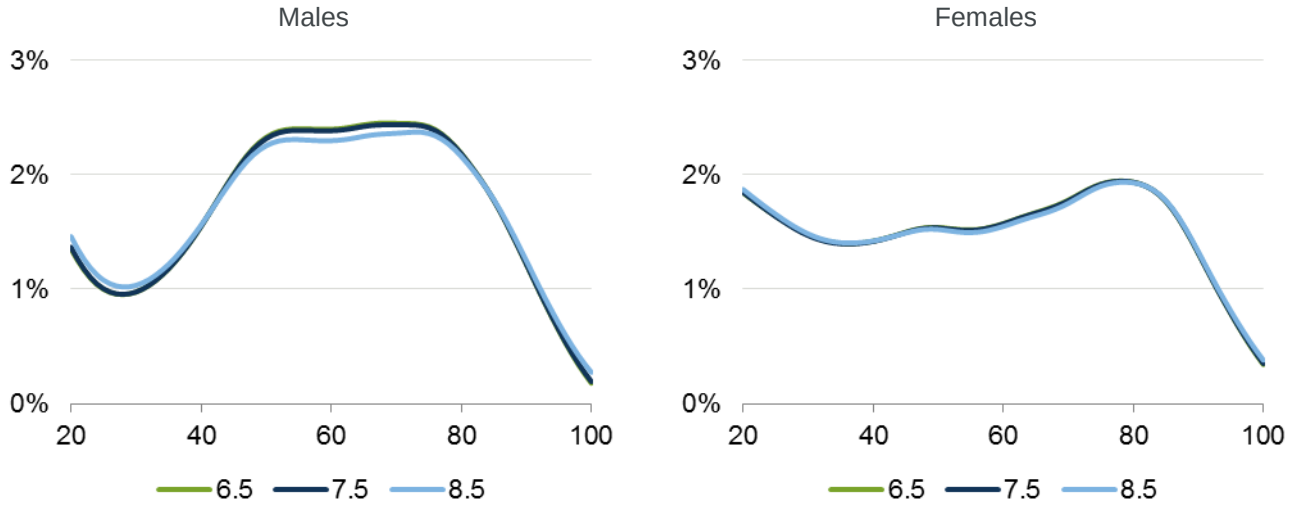


Chart 4L: Impact of varying S_κ on the period component of mortality improvements
This is the same as Chart 4E, but is shown for ease of reference.

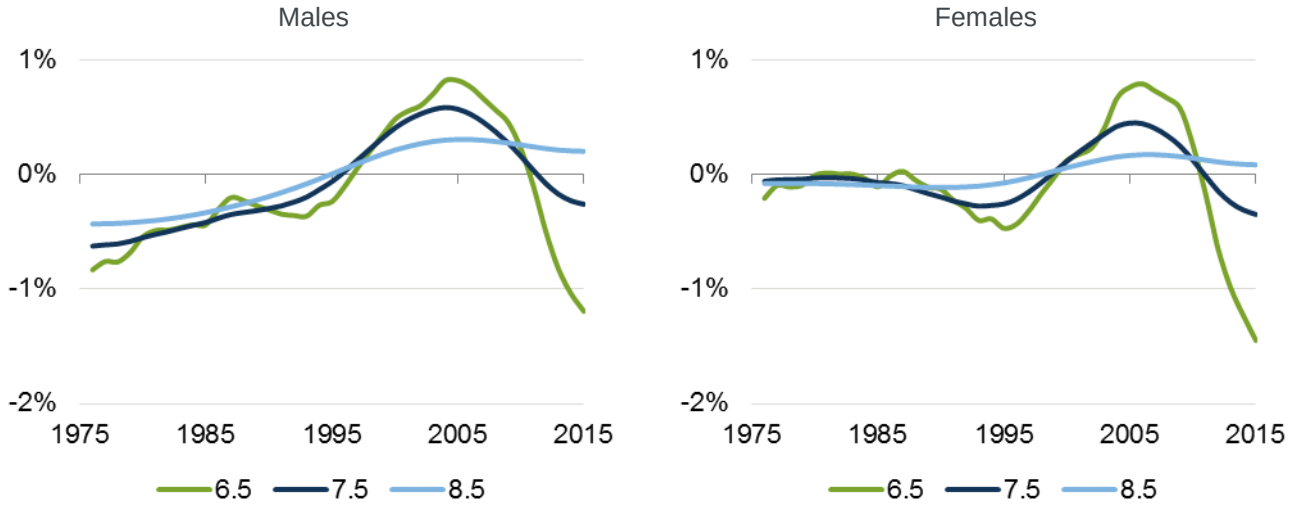
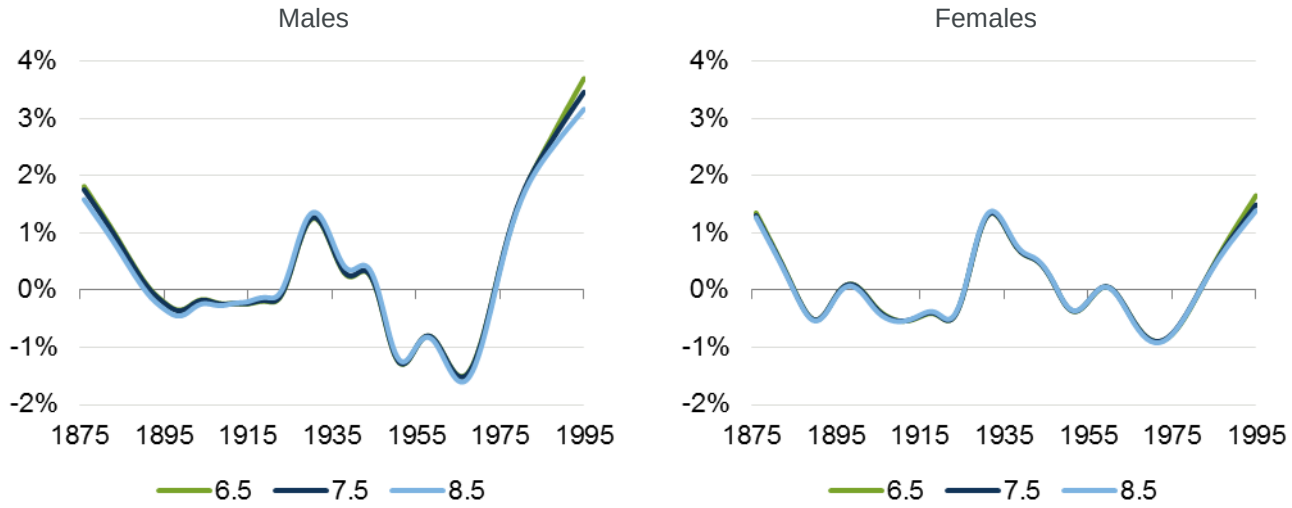


Chart 4M: Impact of varying S_κ on the cohort component of mortality improvements



We see that S_κ has a large impact on the period component of mortality improvements (Chart 4L), which it controls directly. It also affects the age and cohort components of improvements (Charts 4K and 4M) but the impact on these is fairly small, particularly for females. The α_x parameters (Chart 4J) are largely unaffected.

4.5 Impact on S when the volume of data changes

The values of the hyperparameters S_α, S_β, S_κ and S_γ have been set based on analysis of the results of fitting the APCI Model to datasets for England & Wales that cover 81 ages (20-100) and 41 years (e.g. 1975-2015). In this section we consider how these hyperparameters may need to change if the size of the dataset changes.

We first consider the case where the numbers of ages and years remain the same, but the numbers of deaths (actual and expected) change; e.g. fitting the proposed model to a larger or smaller country. In this case the expected deviance is unaffected by the number of deaths as it relates only to the number of degrees of freedom. This is based on the number of age/year cells, and the number of parameters used. Consequently we would expect the hyperparameters to have the same impact and apply the same amount of smoothing, as long as the overdispersion of the new dataset is broadly similar to that for England & Wales.

We next consider the case where the numbers of ages or years change. This has proved challenging to analyse and our attempts to find a neat algebraic approach have not been fruitful. Instead we consider empirical tests, and discuss the results.

Charts 4N to 4P show period components of mortality improvements, for S_κ of 7, 7.5 and 8, for three datasets: the "standard" dataset (ages 20-100, years 1975-2015), halving the age range (ages 60-100, years 1975-2015), and halving the time period (ages 20-100, years 1995-2015). Chart 4Q will be described later, but is placed here for ease of comparison.

Chart 4N: Period components of improvements; ages 20-100, years 1975-2015 ("standard")

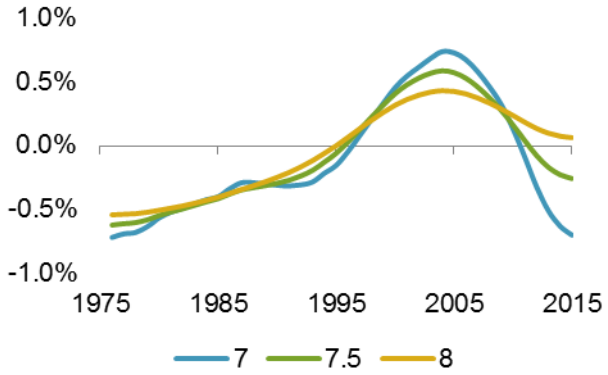


Chart 4O: Period components of improvements; ages 60-100, years 1975-2015 ("60-100")

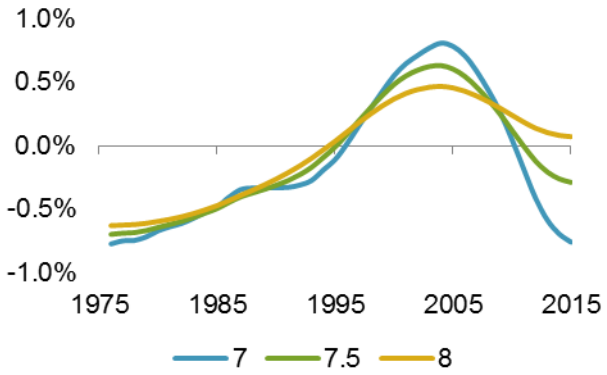


Chart 4P: Period components of improvements; ages 20-100, years 1995-2015 ("1995-2015")

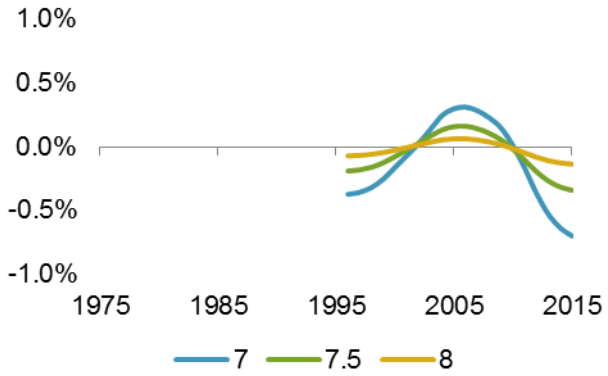


Chart 4Q: Period components of improvements; ages 20-100, years 1995-2015; S_κ reduced

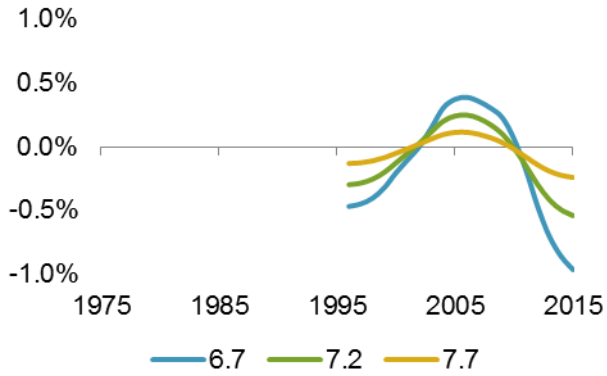


Table 4.1 shows, for each of these cases, the difference in the period component between 2005 and 2015; indicative of the fall from the peak to the current value.

Table 4.1: Fall in period component of mortality improvement from 2005 to 2015

	$S_\kappa = 7$	$S_\kappa = 7.5$	$S_\kappa = 8$
Standard	-1.43%	-0.83%	-0.36%
60-100	-1.54%	-0.89%	-0.38%
1995-2015	-1.00%	-0.50%	-0.20%

Results for ages 60-100 look similar to those for the standard case for ages 20-100. This suggests that changing the age range does not have a material effect on the smoothing of the period component of improvements. Although the results for ages 60-100 show a slightly higher fall in Table 4.1, this may reflect higher falls in mortality improvements at those ages, rather than any artefact of the smoothing process.

Results for the fit to 1995-2015 data look quite different to those for 1975-2015. The overall downward shifts between Charts 4N and 4P and between Charts 4O and 4Q reflect the identifiability constraints applied to the APCI model; as we constrain κ_t to have zero slope, the period components of improvements ($\kappa_{t-1} - \kappa_t$) have an average close to zero. In addition, Table 4.1 shows that the fall in mortality improvements between 2005 and 2015 is materially smaller when using the 1995-2015 dataset; this suggests that the period smoothing parameter has a different impact when the number of years in the data changes.

To address this we consider the impact of halving the value of λ_κ when we halve the number of years. This corresponds to reducing S_κ by 0.3. (Strictly; to halve λ_κ we would reduce S_κ by $\log_{10} 2$, which is 0.301; but this would seem to be spurious accuracy, given the subjective nature of S_κ). Chart 4Q (above) and Table 4.2 show the results when we do this.

Table 4.2: Fall in period component of mortality improvement from 2005 to 2015

	$S_\kappa = 7$	$S_\kappa = 7.5$	$S_\kappa = 8$
Standard (S_κ as stated)	-1.43%	-0.83%	-0.36%
1995-2015 (S_κ as stated)	-1.00%	-0.50%	-0.20%
1995-2015 (S_κ 0.3 lower)	-1.34%	-0.78%	-0.35%

The results for 1995-2015 after reducing S_κ look broadly similar to those for the standard case.

Having considered the specific case of adjusting S_κ to compensate for halving the time period, we now seek to generalise this. The example suggests a rule of thumb that:

$$S_\kappa(T_1) \approx S_\kappa(T_0) + \log_{10}\left(\frac{T_1}{T_0}\right)$$

where T_0 and T_1 are the number of years of data in different datasets, and $S_\kappa(T_0)$ and $S_\kappa(T_1)$ are the corresponding broadly-consistent values of S_κ.

Similarly for the other parameters, where X is the number of ages, and $X + T - 1$ is the number of cohorts:

$$S_\alpha(X_1) \approx S_\alpha(X_0) + \log_{10}\left(\frac{X_1}{X_0}\right)$$

$$S_\beta(X_1) \approx S_\beta(X_0) + \log_{10}\left(\frac{X_1}{X_0}\right)$$

$$S_\gamma(X_1 + T_1 - 1) \approx S_\gamma(X_0 + T_0 - 1) + \log_{10}\left(\frac{X_1+T_1-1}{X_0+T_0-1}\right)$$

In theory we should perhaps allow for the difference between the number of years of data, T, and the number of differences in the penalty function, $T - 2$, but this may be spurious as at this stage we only have a rule of thumb rather than a proven result.

We stress that this is a tentative result, based on limited empirical testing and an educated guess, rather than a rigorous derivation and proof. Further research may be able to verify or improve on this, or may show it to be mistaken, so it should not be relied on.

5. Projections

This section considers a number of issues related to the projection of mortality improvements:

- Section 5.1 sets out in detail the proposed method for projecting mortality improvements, including the conversion between the proposed "m-style" improvements and the traditional "q-style" improvements.
- Section 5.2 provides analysis to support the proposed change in the taper of the long-term rate.
- Section 5.3 presents and discusses an illustrative model to show the difficulty of estimating direction of travel. This complements the empirical tests in Section 8.2 of Working Paper 90.

5.1 Projection and conversion of mortality improvements

Section 6 of Working Paper 90 describes two definitions of mortality improvements.

"q-style" improvements are defined by:

$$MI_{x,t} = 1 - \frac{q_{x,t}}{q_{x,t-1}}$$

"m-style" improvements are defined by:

$$MI^{*}_{x,t} = \log m_{x,t-1} - \log m_{x,t}$$

The proposed model will use m-style improvements within the model (i.e. the initial improvements, long-term rate and projected improvements will be m-style) and these will then be converted to q-style improvements as the final outputs; for consistency with the current Model.

This section describes the calculation and projection of m-style mortality improvements (in steps 1 to 4) and conversion to q-style mortality improvements (steps 5 to 7). We do this by using the example of the proposed Core Model calibrated to data for 1975-2015.

1. Calibrate the APCI model to data for 1975-2015 to obtain values for its parameters.

2. Calculate initial age-period and cohort components of mortality improvements $MI^{AP}_{x,t}$ and $MI^{C}_{t-x,t}$ as:

$$MI^{AP}_{x,2015} = -\beta_x + \kappa_{2014} - \kappa_{2015} \qquad \text{for ages 20 to 100}$$

$$MI^{AP}_{x,2015} = \left(\frac{110-x}{110-100}\right) MI^{AP}_{100,2015} \qquad \text{for ages 101 to 109}$$

$$MI^{AP}_{x,2015} = 0 \qquad \text{for ages 110 to 150}$$

$$MI^{C}_{2015-x,2015} = \gamma_{2014-x} - \gamma_{2015-x} \qquad \text{for ages 20 to 100}$$

$$MI^{C}_{2015-x,2015} = \left(\frac{110-x}{110-100}\right) MI^{C}_{2015-100,2015} \qquad \text{for ages 101 to 109}$$

$$MI^{C}_{2015-x,2015} = 0 \qquad \text{for ages 110 to 150}$$

3a. Project age-period improvements as:

$$MI^{AP}_{x,2015+t} = L^{AP}_{x} + \left(MI^{AP}_{x,2015} - L^{AP}_{x}\right)\left(1 - 3\left(\frac{t}{T^{AP}_{x}}\right)^2 + 2\left(\frac{t}{T^{AP}_{x}}\right)^3\right) + D^{AP}_{x}\, t\left(1 - \frac{t}{T^{AP}_{x}}\right)^2 \qquad \text{for } 0 \le t \le T^{AP}_{x}$$

$$MI^{AP}_{x,2015+t} = L^{AP}_{x} \qquad \text{for } t > T^{AP}_{x}$$

where:

L^{AP}_{x} is the long-term rate for age x

T^{AP}_{x} is the convergence period for age x

D^{AP}_{x} is the direction of travel for age x (zero in the Core Model)

If the shape of convergence is specified in terms of the proportion remaining at midpoint (p_x^{AP}) then:

$$D_x^{AP} = \frac{1}{T_x^{AP}}(8p_x^{AP} - 4)(I - L)$$

3b. Project cohort improvements as:

$$MI_{c,2015+t}^{C} = L_c^C + \left(MI_{c,2015}^{C} - L_c^C\right)\left(1 - 3\left(\frac{t}{T_c^C}\right)^2 + 2\left(\frac{t}{T_c^C}\right)^3\right) + D_c^C t\left(1 - \frac{t}{T_c^C}\right)^2 \quad \text{for } 0 \le t \le T_c^C$$

$$MI_{c,2015+t}^{C} = L_c^C \quad \text{for } t > T_c^C$$

where:

$c = 2015 - x$

L_c^C is the long-term rate for cohort c (zero in the Core Model)

T_c^C is the convergence period for cohort c

D_c^C is the direction of travel for cohort c (zero in the Core Model)

3c. Project total improvements by adding age-period and cohort components; i.e.

$$MI_{x,t}^{*} = MI_{x,t}^{AP} + MI_{t-x,t}^{C} \quad \text{for } t \ge 2015$$

4. Determine $MI_{x,t}^{*}$ for all necessary ages and years:

- For ages 20-150, years 2016 onwards, projected (as in 3 above)
- For ages 20-100, years 1974-2015, calculated as $MI_{x,t}^{*} = \log m_{x,t-1} - \log m_{x,t}$ from the APCI model fit
- For ages 101-109, years 1976-2015, interpolated between $MI_{100,t}^{*}$ and nil at age 110
- For ages 110-150, years 1976-2015, assumed to be nil

5. Determine $\log m_{x,t}$:

- For ages 20-100 in 2015, taken directly from the fit of the APCI model
- For ages 101-150 in 2015, linear extrapolation based on $\log m_{99,2015}$ and $\log m_{100,2015}$ i.e. $\log m_{x,2015} = \log m_{100,2015} + (x - 100)(\log m_{100,2015} - \log m_{99,2015})$
- For ages 20-100, years 1976-2014, determined using $\log m_{x,t} = \log m_{x,t+1} + MI_{x,t+1}^{*}$
- For ages 20-100, years 2016 onwards, determined using $\log m_{x,t} = \log m_{x,t-1} - MI_{x,t}^{*}$

6. Convert to $q_{x,t}$ assuming that $q_{x,t} = 1 - \exp(-m_{x,t})$

7. Calculate $MI_{x,t} = 1 - \frac{q_{x,t}}{q_{x,t-1}}$

5.2 Tapering of the long-term rate

In Working Paper 90 we proposed that the long-term rate should taper to zero between ages 85 and 110, rather than between ages 90 and 120 as in the current Model. This section provides evidence to support the Committee's proposal.

Chart 5A plots the male and female age components of mortality improvements from the APCI model fitted to data for 1975 to 2015. This is compared against long-term rates of 1%, 1.5% and 2% p.a. below age 90, tapering to zero by age 120, as in the current Model.

Chart 5A shows that the current assumption implies a material increase in mortality improvements at centenarian ages. For example, a long-term rate assumption of 1.5% p.a. corresponds to a long-term rate of 1% p.a. at age 100. However the fitted age component of mortality improvements is just 0.19% p.a. for males and 0.35% p.a. for females.

Chart 5A: Age components of mortality improvements, and long-term rates tapering from age 90 to age 120

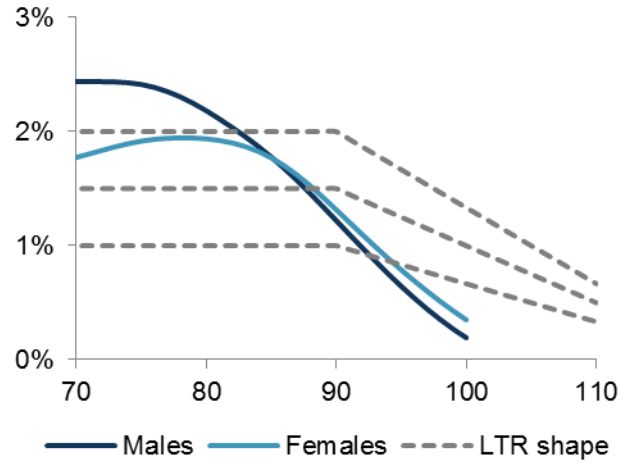


Chart 5B: Age components of mortality improvements, and long-term rates tapering from age 85 to age 110

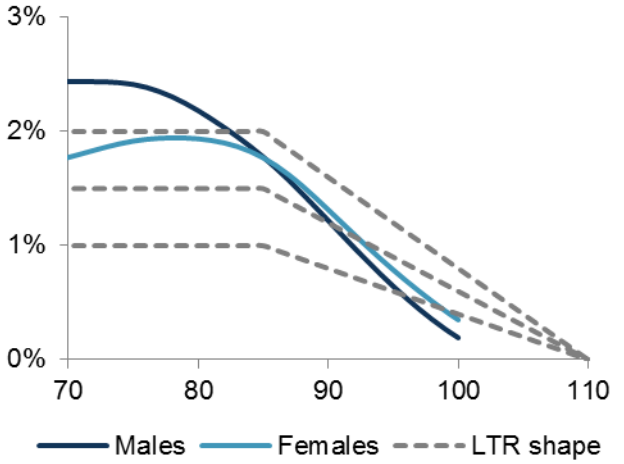


Chart 5B shows the alternative assumption with a taper between ages 85 and 110. The example of a long-term rate assumption of 1.5% p.a. produces a long-term rate of 0.6% p.a. at age 100 so still allows for mortality improvements at older ages to increase in the future, but this is now significantly lower than in the current Core Model.

Figure 3.13 of Working Paper 39 plotted smoothed mortality improvements by age for a range of years and observed that "the estimated improvement rates at high ages display a near-parallel shift over the past two decades, so that the age-point of reaching around 'zero', that is stable mortality rates, appears to have been increasing over time." This provided justification for a relatively high age for mortality improvements to taper to zero.

We have revisited the analysis using more recent data.

Chart 5C: Mortality improvements by age for 1985 to 2005 (as in Figure 3.13 of Working Paper 39)

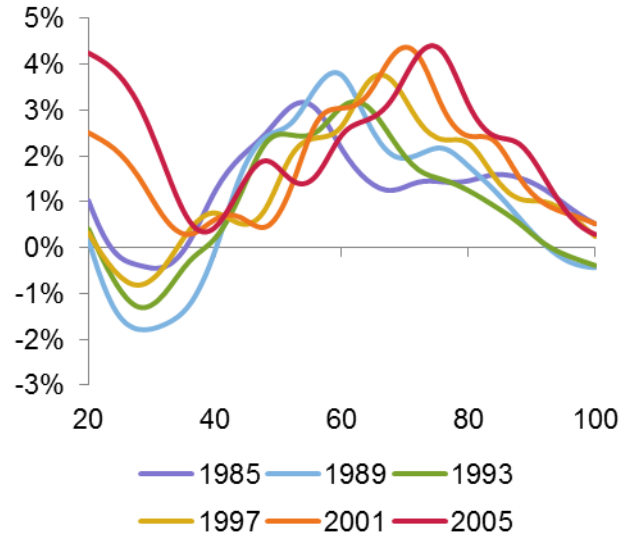


Chart 5D: Mortality improvements by age for 1993 to 2013

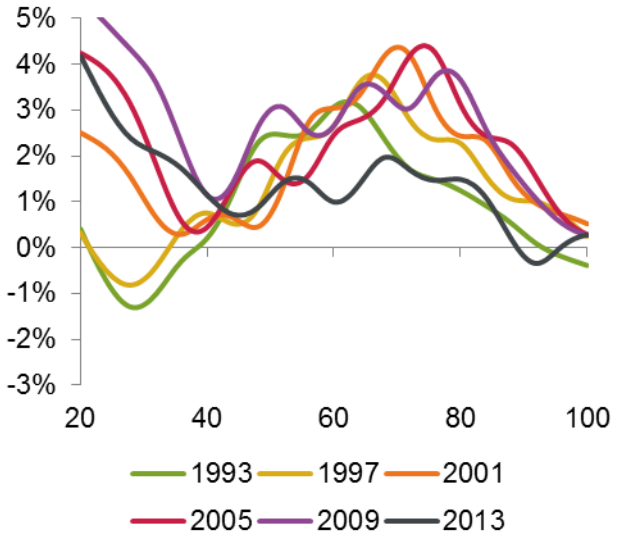


Chart 5C shows mortality improvements for each fourth year, using data that was available at the time of Working Paper 39. It shows that mortality improvements were moving to the right, suggesting that the age at

which mortality improvements would reach zero was increasing. Chart 5D includes more recent data and shows that this pattern has not continued; improvements for 2009 are similar to those for 2005, and those for 2013 are considerably lower. Chart 5D provides further support for the proposed tapering of the long-term rate.

When changing the taper of the long-term rate at older ages, users of the Model may also wish to consider whether the assumption at younger ages remains appropriate. Adopting the new taper in isolation, without adjusting the long-term rate assumption, would lead to lower liabilities.

5.3 Direction of travel

One of the concerns that has been raised about the current Model is the lack of allowance for "direction of travel" in mortality improvements. The Core parameterisation of the Model assumes a convergence function that is instantaneously flat at the start of the projection period, rather than allowing mortality improvements to rise or fall according to what happened in the recent past.

Section 8.2 of Working Paper 90 noted the difficulty of estimating direction of travel, by showing how it had varied over time. This section uses an illustrative model to provide some insight into why this difficulty arises.

Model – structure and properties

We model mortality as a combination of a smooth, slowly-varying, deterministic underlying trend $S(t)$ and annual noise:

$$\log m(t) = S(t) + \sigma Z(t)$$

where $Z(t)$ is a standard normal random variable, so the noise is independent and identically distributed (iid).

This could be considered as relating to an individual age or age-group or a standardised mortality ratio (SMR) across a wide age range.

Mortality improvement is the negative change in log mortality rates:

$$MI(t) = -\Delta \log m(t)$$

and "direction of travel" is the rate of change in this:

$$DoT(t) = \Delta MI(t) = -\Delta^2 \log m(t)$$

Considering the first and second order differences of $\log m(t)$ we have:

$$\Delta \log m(t) = [S(t+1) - S(t)] + \sigma[Z(t+1) - Z(t)]$$

$$\Delta^2 \log m(t) = [S(t+2) - 2S(t+1) + S(t)] + \sigma[Z(t+2) - 2Z(t+1) + Z(t)]$$

So that:

$$Var(\log m(t)) = \sigma^2$$

$$Var\big(MI(t)\big) = Var(-\Delta \log m(t)) = Var(\Delta \log m(t)) = 2\sigma^2$$

$$Var\big(DoT(t)\big) = Var(-\Delta^2 \log m(t)) = Var(\Delta^2 \log m(t)) = 6\sigma^2$$

Model – underlying trend

The choice of the smooth trend is somewhat arbitrary, but the following captures key features:

$$S(t) = a - bt + \frac{c}{k}\sin(kt) \text{ where } k = \frac{\pi}{T}$$

Here $S(t)$ is a combination of a linear trend and a periodic component with time T between the peak and trough of the periodic component.

Using a derivative rather than a difference, we have:

$$\frac{dS}{dt} = -b + c.\cos(kt)$$

so the smooth underlying component of the mortality improvement $\left(-\frac{dS}{dt}\right)$ has a mean of b and oscillates in the range $[b - c, b + c]$.

Also:

$$\frac{d^2S}{dt^2} = -ck.\sin(kt)$$

so the smooth underlying component of direction of travel $\left(-\frac{d^2S}{dt^2}\right)$ is in the range $[-ck, +ck]$.

The range of the smooth component of direction of travel is $2ck$ compared to $2c$ for mortality improvement; i.e. the range for direction of travel is $k = \frac{\pi}{T}$ times that of mortality improvement.

Calibration

Plausible parameters for this form of $S(t)$ are:

- $b = 2\%$ for the mean improvement;
- $c = 1\%$ for the "half-range" of the improvements, so improvements are in the range $[1\%, 3\%]$;
- $T = 25$ as the time between minimum and maximum improvements, as they vary slowly; and
- $a = 0$ for simplicity, as it does not affect mortality improvements.

Finally we set σ, the standard deviation of the noise component, to $\sigma = 1.5\%$ as a plausible value for a standardised mortality ratio; for example, based on comparing versions of the APCI model with smooth and noisy period terms.

Results

Charts 5A to 5C show one realisation of the random process for $\log m\,(t)$, shown for an illustrative 100-year period. We plot $\log m$, $-\Delta \log m$ and $-\Delta^2 \log m$ as dots, and the corresponding smooth components S, $-\Delta S$, and $-\Delta^2 S$ as lines.

Chart 5E: log mortality rate
$\log m(t)$

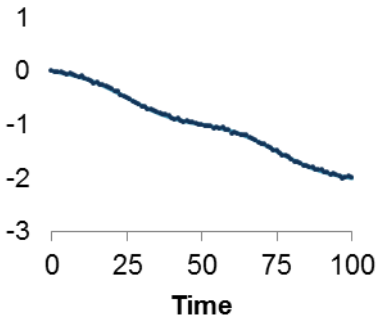


Chart 5F: Mortality improvement
$MI(t) = -\Delta \log m(t)$

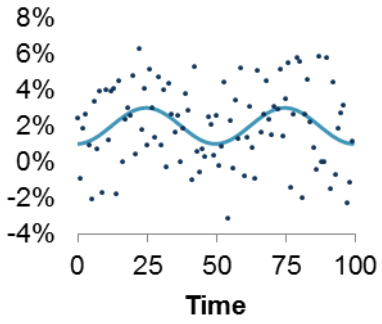
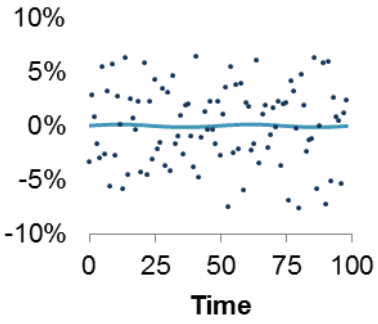


Chart 5G: Direction of travel
$DoT(t) = -\Delta^2 \log m(t)$



As we move from rates to improvements to direction of travel, the range of the smooth component narrows, but the standard deviation of the noise component increases. The proportionate error (the standard deviation divided by the range) increases by more than an order of magnitude as we move from improvement to direction of travel.

Table 5.1: Relative uncertainty of mortality rates, improvements, and direction of travel

	Standard deviation (StDev) of noise	Range of smooth component		Ratio (StDev ÷ Range)
Mortality rate	1.50%	200%	(-200% to 0%)	0.0075
Mortality improvement	2.12%	2%	(1% to 3%)	1.06
Direction of travel	3.67%	0.25%	(-0.13% to +0.13%)	14.6

Note that the "Ratio" figure for mortality rates depends on the length of the projection, as under this model rates have a long-term trend while improvements and direction of travel are bounded.

Our illustrative model and calibration suggests that direction of travel is significantly more difficult to estimate than mortality rates or improvements, and supports our decision not to allow for it in the Core parameterisation of the Model.

6. Sensitivity of life expectancy

This section considers the sensitivity of life expectancy to a wide range of parameters and model choices.

The cases considered are described in Table 6.1. All of the life expectancies in this section are cohort life expectancies calculated as at end 31 December 2015. They use base tables of S2PMA for males and S2PFA for females, and an illustrative long-term rate of 1.5% p.a. unless stated otherwise.

Table 6.1: Description of life expectancy calculations in this section

Short name	Description
Standard	As proposed in Working Paper 90
$S_\alpha = 6$	As standard, but reducing S_α from 7 to 6
$S_\alpha = 8$	As standard, but increasing S_α from 7 to 8
$S_\beta = 8$	As standard, but reducing S_β from 9 to 8
$S_\beta = 10$	As standard, but increasing S_β from 9 to 10
$S_\kappa = 6.5$	As standard, but reducing S_κ from 7.5 to 6.5
$S_\kappa = 8.5$	As standard, but increasing S_κ from 7.5 to 8.5
$S_\gamma = 6$	As standard, but reducing S_γ from 7 to 6
$S_\gamma = 8$	As standard, but increasing S_γ from 7 to 8
LTR of 1% p.a.	As standard, but reducing the long-term rate from 1.5% to 1.0% p.a.
LTR of 2% p.a.	As standard, but increasing the long-term rate from 1.5% to 2.0% p.a.
Old LTR tapering	As standard, but retaining the current tapering of the long-term rate (between ages 90 and 120, rather than between ages 85 and 110)
Old convergence	As standard, but retaining current cohort convergence periods
Direction of travel	As standard, but allowing for direction of travel, based on the fitted period component
Raw exposure	As standard, but applied to data without making the exposure adjustment
Ages 20-90	As standard, but fitted to the age range 20-90 rather than 20-100 (retaining the taper to nil improvements at age 110)
Ages 60-100	As standard, but fitted to the age range 60-100 rather than 20-100
Years 1995-2015	As standard, but fitted to the calendar years 1995-2015, rather than 1975-2015
Small cohorts	As standard, but fitted to data that excludes birth cohorts with five or fewer observations. i.e. fitted to cohorts from 1880-1990, rather than 1875-1995

Tables 6.2 to 6.5 show life expectancies for these cases, for males and females, in absolute terms, and relative to the standard case.

Table 6.2: Life expectancy for males

Age	25	35	45	55	65	75	85	95	105
Standard	63.826	52.812	41.993	31.872	22.396	13.737	6.775	2.829	1.586
$S_\alpha = 6$	63.825	52.810	41.991	31.870	22.396	13.736	6.775	2.829	1.585
$S_\alpha = 8$	63.834	52.824	42.006	31.882	22.402	13.739	6.777	2.830	1.586
$S_\beta = 8$	63.826	52.812	41.993	31.871	22.397	13.737	6.776	2.830	1.588
$S_\beta = 10$	63.823	52.808	41.988	31.875	22.396	13.736	6.774	2.824	1.582
$S_\kappa = 6.5$	63.043	51.998	41.165	31.108	21.770	13.312	6.556	2.736	1.566
$S_\kappa = 8.5$	64.283	53.283	42.473	32.349	22.825	14.054	6.951	2.908	1.603
$S_\gamma = 6$	63.826	52.817	41.979	31.904	22.379	13.709	6.775	2.832	1.585
$S_\gamma = 8$	63.824	52.812	42.011	31.843	22.406	13.756	6.770	2.828	1.587
LTR of 1% p.a.	62.125	51.481	41.044	31.271	22.073	13.601	6.736	2.822	1.586
LTR of 2% p.a.	65.491	54.142	42.956	32.489	22.730	13.877	6.815	2.836	1.587
Old LTR tapering	64.509	53.334	42.374	32.141	22.575	13.841	6.820	2.842	1.590
Old convergence	63.979	52.842	41.973	31.872	22.396	13.737	6.775	2.829	1.586
Direction of travel	63.776	52.759	41.936	31.821	22.360	13.719	6.770	2.828	1.586
Raw exposure	63.846	52.833	42.014	31.892	22.415	13.733	6.769	2.839	1.588
Ages 20-90	63.903	52.882	42.060	31.940	22.465	13.801	6.842	2.949	1.614
Ages 60-100	-	-	-	-	22.341	13.619	6.721	2.806	1.580
Years 1995-2015	63.537	52.521	41.669	31.344	21.698	13.203	6.585	2.813	1.585
Small cohorts	-	52.933	42.104	31.954	22.442	13.757	6.782	2.832	1.588

Table 6.3: Life expectancy for females

Age	25	35	45	55	65	75	85	95	105
Standard	65.682	54.744	44.137	34.028	24.329	15.244	7.636	3.268	1.696
$S_\alpha = 6$	65.677	54.737	44.132	34.024	24.326	15.243	7.635	3.267	1.695
$S_\alpha = 8$	65.703	54.779	44.166	34.046	24.341	15.247	7.638	3.271	1.698
$S_\beta = 8$	65.682	54.744	44.138	34.028	24.329	15.244	7.636	3.269	1.696
$S_\beta = 10$	65.679	54.741	44.134	34.029	24.328	15.242	7.635	3.265	1.693
$S_\kappa = 6.5$	64.863	53.886	43.269	33.228	23.661	14.772	7.381	3.156	1.673
$S_\kappa = 8.5$	66.028	55.105	44.503	34.377	24.632	15.471	7.766	3.327	1.708
$S_\gamma = 6$	65.684	54.740	44.134	34.044	24.327	15.229	7.637	3.270	1.695
$S_\gamma = 8$	65.695	54.774	44.146	34.028	24.348	15.248	7.639	3.261	1.697
LTR of 1% p.a.	64.055	53.458	43.202	33.409	23.975	15.085	7.588	3.259	1.695
LTR of 2% p.a.	67.254	56.013	45.076	34.658	24.691	15.407	7.684	3.277	1.696
Old LTR tapering	66.490	55.378	44.616	34.375	24.562	15.381	7.695	3.286	1.700
Old convergence	65.682	54.704	44.126	34.028	24.329	15.244	7.636	3.268	1.696
Direction of travel	65.626	54.684	44.074	33.971	24.285	15.221	7.629	3.266	1.696
Raw exposure	65.678	54.740	44.132	34.024	24.327	15.233	7.622	3.269	1.695
Ages 20-90	65.690	54.745	44.132	34.019	24.321	15.242	7.674	3.378	1.724
Ages 60-100	-	-	-	-	24.178	15.122	7.595	3.252	1.692
Years 1995-2015	65.517	54.567	43.845	33.489	23.698	14.817	7.650	3.415	1.731
Small cohorts	-	54.749	44.144	34.033	24.332	15.246	7.636	3.269	1.697

Table 6.4: Percentage change in life expectancy, compared to standard case, for males

Age	25	35	45	55	65	75	85	95	105
$S_\alpha = 6$	−0.00%	−0.00%	−0.00%	−0.01%	−0.00%	−0.00%	−0.01%	−0.00%	−0.08%
$S_\alpha = 8$	+0.01%	+0.02%	+0.03%	+0.03%	+0.02%	+0.02%	+0.02%	+0.03%	−0.00%
$S_\beta = 8$	+0.00%	+0.00%	+0.00%	−0.00%	+0.00%	+0.00%	+0.01%	+0.02%	+0.12%
$S_\beta = 10$	−0.00%	−0.01%	−0.01%	+0.01%	−0.00%	−0.00%	−0.01%	−0.16%	−0.26%
$S_\kappa = 6.5$	−1.23%	−1.54%	−1.97%	−2.40%	−2.80%	−3.10%	−3.23%	−3.30%	−1.28%
$S_\kappa = 8.5$	+0.72%	+0.89%	+1.14%	+1.50%	+1.91%	+2.31%	+2.60%	+2.80%	+1.08%
$S_\gamma = 6$	−0.00%	+0.01%	−0.03%	+0.10%	−0.08%	−0.21%	+0.00%	+0.10%	−0.10%
$S_\gamma = 8$	−0.00%	+0.00%	+0.04%	−0.09%	+0.04%	+0.14%	−0.07%	−0.03%	+0.05%
LTR of 1% p.a	−2.66%	−2.52%	−2.26%	−1.89%	−1.44%	−0.99%	−0.57%	−0.25%	−0.03%
LTR of 2% p.a.	+2.61%	+2.52%	+2.29%	+1.94%	+1.49%	+1.02%	+0.59%	+0.25%	+0.03%
Old LTR tapering	+1.07%	+0.99%	+0.91%	+0.84%	+0.80%	+0.76%	+0.66%	+0.48%	+0.25%
Old convergence	+0.24%	+0.06%	−0.05%	−0.00%	−0.00%	−0.00%	−0.00%	−0.00%	−0.00%
Direction of travel	−0.08%	−0.10%	−0.13%	−0.16%	−0.16%	−0.13%	−0.07%	−0.03%	−0.01%
Raw exposure	+0.03%	+0.04%	+0.05%	+0.06%	+0.08%	−0.03%	−0.09%	+0.36%	+0.10%
Ages 20-90	+0.12%	+0.13%	+0.16%	+0.21%	+0.31%	+0.47%	+0.98%	+4.24%	+1.78%
Ages 60-100	-	-	-	-	−0.25%	−0.86%	−0.79%	−0.83%	−0.38%
Years 1995-2015	−0.45%	−0.55%	−0.77%	−1.66%	−3.12%	−3.89%	−2.81%	−0.56%	−0.06%
Small cohorts	-	+0.23%	+0.27%	+0.26%	+0.20%	+0.14%	+0.09%	+0.09%	+0.09%

Table 6.5: Percentage change in life expectancy, compared to standard case, for females

Age	25	35	45	55	65	75	85	95	105
$S_\alpha = 6$	−0.01%	−0.01%	−0.01%	−0.01%	−0.01%	−0.01%	−0.01%	−0.02%	−0.05%
$S_\alpha = 8$	+0.03%	+0.06%	+0.07%	+0.05%	+0.05%	+0.02%	+0.03%	+0.10%	+0.13%
$S_\beta = 8$	+0.00%	+0.00%	+0.00%	−0.00%	−0.00%	+0.00%	+0.00%	+0.02%	+0.02%
$S_\beta = 10$	−0.00%	−0.01%	−0.01%	+0.00%	−0.00%	−0.01%	−0.01%	−0.10%	−0.13%
$S_\kappa = 6.5$	−1.25%	−1.57%	−1.97%	−2.35%	−2.75%	−3.10%	−3.33%	−3.42%	−1.36%
$S_\kappa = 8.5$	+0.53%	+0.66%	+0.83%	+1.02%	+1.25%	+1.49%	+1.71%	+1.82%	+0.71%
$S_\gamma = 6$	+0.00%	−0.01%	−0.01%	+0.05%	−0.01%	−0.10%	+0.02%	+0.07%	−0.06%
$S_\gamma = 8$	+0.02%	+0.05%	+0.02%	−0.00%	+0.08%	+0.02%	+0.04%	−0.23%	+0.05%
LTR of 1% p.a	−2.48%	−2.35%	−2.12%	−1.82%	−1.45%	−1.04%	−0.63%	−0.27%	−0.03%
LTR of 2% p.a.	+2.39%	+2.32%	+2.13%	+1.85%	+1.49%	+1.07%	+0.64%	+0.27%	+0.03%
Old LTR tapering	+1.23%	+1.16%	+1.09%	+1.02%	+0.96%	+0.90%	+0.78%	+0.55%	+0.28%
Old convergence	−0.00%	−0.07%	−0.03%	−0.00%	−0.00%	−0.00%	−0.00%	−0.00%	−0.01%
Direction of travel	−0.09%	−0.11%	−0.14%	−0.17%	−0.18%	−0.15%	−0.09%	−0.04%	−0.01%
Raw exposure	−0.01%	−0.01%	−0.01%	−0.01%	−0.01%	−0.07%	−0.18%	+0.03%	−0.02%
Ages 20-90	+0.01%	+0.00%	−0.01%	−0.03%	−0.03%	−0.02%	+0.50%	+3.37%	+1.69%
Ages 60-100	-	-	-	-	−0.62%	−0.80%	−0.53%	−0.49%	−0.20%
Years 1995-2015	−0.25%	−0.32%	−0.66%	−1.58%	−2.59%	−2.80%	+0.19%	+4.50%	+2.11%
Small cohorts	-	+0.01%	+0.02%	+0.02%	+0.01%	+0.01%	+0.01%	+0.03%	+0.05%

The tables show that:

- With regard to the smoothing parameters, life expectancies are:
  - Very insensitive to the choices made for $S_\alpha$and S_β.
  - Very sensitive to the choice made for S_κ, with the impact varying across the age range. At ages above 45, changing S_κ by 1 has a greater impact than changing the long-term rate by 0.5%.
  - Fairly insensitive to the choice of S_γ. The impact is not in the same direction at all ages, so the choice of S_γ may be relatively immaterial when averaged across a whole portfolio of lives
- The proposed tapering for the long-term rate reduces liabilities, compared to the old tapering. At ages above 75, the impact of the change in the tapering is larger than changing the long-term rate by 0.4%
- The proposed new cohort convergence pattern has a very small impact, apart from for the youngest males.
- As expected, introducing direction of travel leads to lower life expectancies, because the modelled mortality improvements are lower in 2015 than 2014. The directional impact of this element depends on the direction of travel at the start of the projection period.
- The impact of adjusting the exposure as set out in Section 2 is relatively small. The adjustment applied does have a knock-on effect on the amount of smoothing because without adjust the deviance is higher, which means that the smoothing parameters have less effect.
- Fitting to ages 20-90 has little impact on projected life expectancies until age 85. For ages 95 and 105, the impact is relatively high. This is because the tapering from age 90 to 110 means that we assume a higher mortality improvement than the proposed model produces.
- Fitting to ages 60-100 reduces life expectancies.
- Fitting to years 1995-2015 reduces life expectancies, apart from older females.
- Excluding cohorts with few observations is immaterial for females. For younger males, the impact is significant. This may be due to the impact on projections of the youngest male cohorts.

7. Progression of life expectancy

The previous section considers the sensitivity of life expectancy to a range of parameters and model choices. Those calculations all use the same dataset to show the impact of changes that we propose to make to the Model, on a consistent basis.

In contrast, this section primarily considers how estimates of life expectancy vary over time as new data has emerged:

- Section 7.1 compares life expectancy at 31 December 2015 for recent versions of the Model and the proposed model.
- Section 7.2 compares results from historical versions of the Model with hypothetical results if the proposed model had been applied at the time.

7.1. Life expectancy at 31 December 2015

This section compares life expectancy at 31 December 2015 for different models. Table 7.1 summarises the calculations that we consider.

Table 7.1: Summary of calculations considered in this section

Model	Release date	Actual data to	Estimated data to	Initial year
CMI_2014	November 2014	30 September 2014	31 December 2014	2011
CMI_2015	September 2015	31 July 2015	31 December 2015	2012
CMI_2015*	[March 2016]	31 December 2015	n/a	2012
Current method	[March 2016]	31 December 2015	n/a	2013
Proposed method	[March 2016]	31 December 2015	n/a	2015

CMI_2014 was released towards the end of 2014. It uses actual data to 30 September 2014 (based on ONS provisional weekly death data) and estimated data for the rest of 2014. The Committee retained the previous approach of stepping back two years from the last full year of data (i.e. stepping back two years from 2013) to give an initial year of 2011.

CMI_2015 was similar to CMI_2014 in that it used actual data for part of the year (to 31 July 2015 in this case) and estimated data for the rest of the year. A two-year step-back results in an initial year of 2012.

The remaining models all use the same dataset, actual data to the end of 2015 (again based on ONS provisional weekly death data). This is what would have been done if the CMI had released a version of the model in March 2016 (indicated by the release date in brackets), consistent with the move to a March release data for future versions of the Model announced in Working Paper 80. However these are not official releases of the Model.

CMI_2015* is similar to CMI_2015, apart from using actual data for the whole of 2015.

The “current method” has two differences to CMI_2015. As well as using different data, the two-year step-back gives an initial year of 2013, one year later than for CMI_2015.

For the “proposed method” there is no step-back, so the initial year is 2015.

Tables 7.2 to 7.7 show life expectancies for these cases, for males and females, in absolute terms, and relative to CMI_2015. As in Section 6, the life expectancies in this section are cohort life expectancies calculated at end 31 December 2015. They use base tables of S2PMA for males and S2PFA for females, and an illustrative long-term rate of 1.5% p.a.

Table 7.2: Life expectancy for males

Age	25	35	45	55	65	75	85	95	105
CMI_2014	65.115	53.903	42.922	32.564	22.717	13.813	6.930	2.998	1.642
CMI_2015	64.835	53.619	42.635	32.283	22.434	13.563	6.801	2.948	1.623
CMI_2015*	64.722	53.507	42.522	32.168	22.340	13.482	6.710	2.930	1.625
Current method	64.508	53.290	42.301	31.942	22.117	13.296	6.603	2.890	1.610
Proposed method	63.826	52.812	41.993	31.872	22.396	13.737	6.775	2.829	1.586

Table 7.3: Life expectancy for females

Age	25	35	45	55	65	75	85	95	105
CMI_2014	67.293	56.122	45.234	34.804	24.815	15.481	7.921	3.571	1.786
CMI_2015	66.969	55.787	44.906	34.477	24.474	15.186	7.765	3.500	1.755
CMI_2015*	66.836	55.646	44.769	34.344	24.351	15.092	7.704	3.494	1.755
Current method	66.648	55.457	44.576	34.136	24.133	14.895	7.579	3.446	1.737
Proposed method	65.682	54.744	44.137	34.028	24.329	15.244	7.636	3.268	1.696

Table 7.4: Change in life expectancy, compared to CMI_2015, for males

Age	25	35	45	55	65	75	85	95	105
CMI_2014	+0.280	+0.285	+0.286	+0.281	+0.283	+0.250	+0.129	+0.051	+0.019
CMI_2015	-	-	-	-	-	-	-	-	-
CMI_2015*	−0.113	−0.112	−0.113	−0.115	−0.093	−0.080	−0.091	−0.017	+0.002
Current method	−0.327	−0.329	−0.335	−0.341	−0.316	−0.267	−0.198	−0.057	−0.013
Proposed method	−1.010	−0.807	−0.643	−0.411	−0.037	+0.174	−0.026	−0.119	−0.037

Table 7.5: Change in life expectancy, compared to CMI_2015, for females

Age	25	35	45	55	65	75	85	95	105
CMI_2014	+0.324	+0.334	+0.327	+0.328	+0.341	+0.295	+0.156	+0.071	+0.031
CMI_2015	-	-	-	-	-	-	-	-	-
CMI_2015*	−0.134	−0.141	−0.137	−0.133	−0.123	−0.094	−0.061	−0.006	+0.000
Current method	−0.322	−0.330	−0.331	−0.340	−0.341	−0.291	−0.186	−0.054	−0.018
Proposed method	−1.287	−1.043	−0.769	−0.448	−0.145	+0.058	−0.129	−0.232	−0.059

Table 7.6: Percentage change in life expectancy, compared to CMI_2015, for males

Age	25	35	45	55	65	75	85	95	105
CMI_2014	+0.43%	+0.53%	+0.67%	+0.87%	+1.26%	+1.84%	+1.90%	+1.72%	+1.18%
CMI_2015	-	-	-	-	-	-	-	-	-
CMI_2015*	−0.17%	−0.21%	−0.26%	−0.36%	−0.42%	−0.59%	−1.34%	−0.59%	+0.11%
Current method	−0.50%	−0.61%	−0.78%	−1.06%	−1.41%	−1.97%	−2.91%	−1.94%	−0.82%
Proposed method	−1.56%	−1.50%	−1.51%	−1.27%	−0.17%	+1.29%	−0.38%	−4.03%	−2.28%

Table 7.7: Percentage change in life expectancy, compared to CMI_2015, for females

Age	25	35	45	55	65	75	85	95	105
CMI_2014	+0.48%	+0.60%	+0.73%	+0.95%	+1.39%	+1.94%	+2.01%	+2.02%	+1.76%
CMI_2015	-	-	-	-	-	-	-	-	-
CMI_2015*	−0.20%	−0.25%	−0.31%	−0.39%	−0.50%	−0.62%	−0.79%	−0.17%	+0.02%
Current method	−0.48%	−0.59%	−0.74%	−0.99%	−1.39%	−1.91%	−2.39%	−1.53%	−1.00%
Proposed method	−1.92%	−1.87%	−1.71%	−1.30%	−0.59%	+0.38%	−1.66%	−6.63%	−3.36%

The tables show that:

1. CMI_2015 gave lower life expectancies at all ages than CMI_2014, with the proportionate impact being higher at ages between 65 and 95.
2. CMI_2015* (using actual data for all of 2015, but still stepping back to 2012) gives lower life expectancies than CMI_2015 for all but age 105. This means that mortality in the latter part of 2015 was heavier than the neutral estimate used to produce CMI_2015.
3. The current method (using actual data for 2015, stepping back to 2012) would give lower life expectancies than CMI_2015. This is partly due to the use of actual data. Also, the current method uses an initial year of 2013 rather than 2012 which effectively means that more of the fall in improvements over 2015 is allowed for in the initial rates. This in turn means that the initial mortality improvements fall further than in CMI_2015 or CMI_2015*. The change in the initial year has a bigger impact than using actual data for the whole of 2015.

4. The proposed method gives higher life expectancies than the current method at the youngest and oldest ages. At ages 65 and 85, the proposed method gives a higher life expectancy than the current method (but lower than CMI_2015). At age 75, the proposed method actually gives a higher life expectancy than CMI_2015. It is perhaps worth noting that the raw improvements in England & Wales over 2015 for the age group 70-75 were higher than for surrounding cohorts.
5. The largest proportionate impact of the proposals is at ages 95 and 105 – as set out in Section 3, due to the proposed change in tapering ages, improvements are assumed to be significantly lower at these age groups than in previous models.

7.2. Progression of life expectancy

This section compares results from historical versions of the Model with the hypothetical results if the method proposed in Working Paper 90 had been in use at the time.

We define and compare:

- "Historical" life expectancies – those calculated using historical versions of the Model.
- "Proposed" life expectancies – those that would have been if the method and Core assumptions proposed in Working Paper 90 had been adopted at the time.

In calculating the "proposed" life expectancies the model is always calibrated to 41 years of data, but we use the same data that was used for the "historical" life expectancies at the time[5]. i.e. the data for CMI_2009, CMI_2010 and CMI_2011 has not been re-stated following the 2011 Census.

Table 7.8 summarises the calculations.

Table 7.8: Summary of life expectancy calculations

Model version	Publication date	Life expectancy as at	Base table	Calibration data (historical)	Calibration data (proposed)
CMI_2009	25 November 2009	31 December 2009	S1PxA	1961-2008	1968-2008
CMI_2010	23 November 2010	31 December 2010	S1PxA	1961-2009	1969-2009
CMI_2011	16 September 2011	31 December 2011	S1PxA	1961-2010	1970-2010
CMI_2012	8 February 2013	31 December 2012	S1PxA	1961-2011	1971-2011
CMI_2013	13 September 2013	31 December 2013	S1PxA	1961-2012	1972-2012
CMI_2014	24 November 2014	31 December 2014	S2PxA	1974-2014	1974-2014
CMI_2015	28 September 2015	31 December 2015	S2PxA	1975-2015	1975-2015

Charts 7A to 7F compare life expectancies on the historical and proposed bases for ages 45, 65 and 85 for males and females.

[5] A consolidated file of this data was published with Working Paper 84 as https://www.actuaries.org.uk/documents/cmi-working-paper-84-exposure-and-deaths-dataset-used-preparation-cmi2015

Chart 7A: Comparison of life expectancies for males aged 45

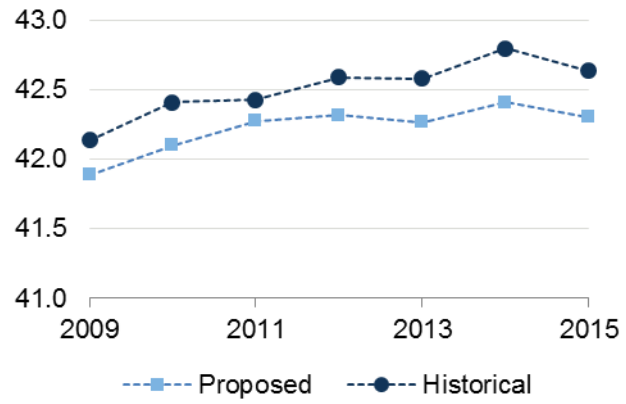


Chart 7B: Comparison of life expectancies for females aged 45

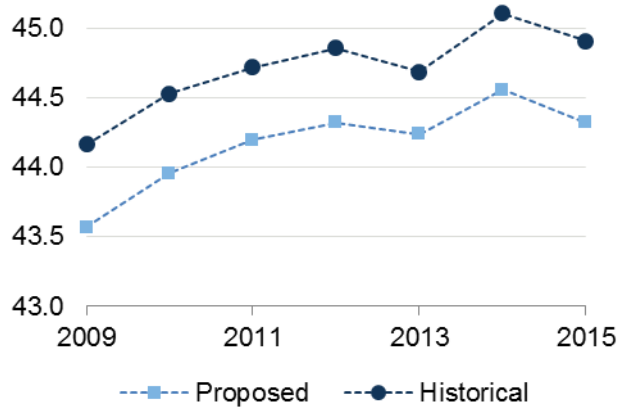


Chart 7C: Comparison of life expectancies for males aged 65

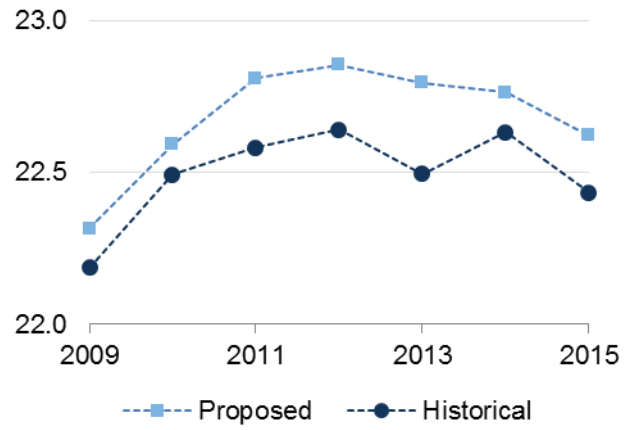


Chart 7D: Comparison of life expectancies for females aged 65

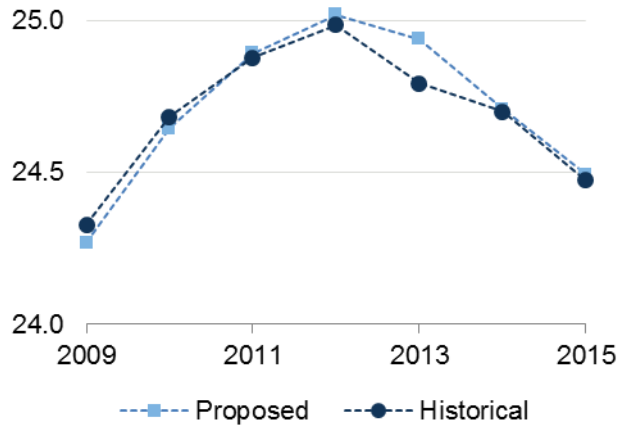


Chart 7E: Comparison of life expectancies for males aged 85

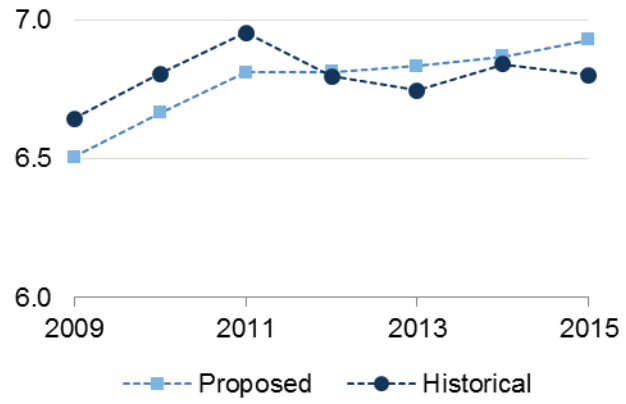
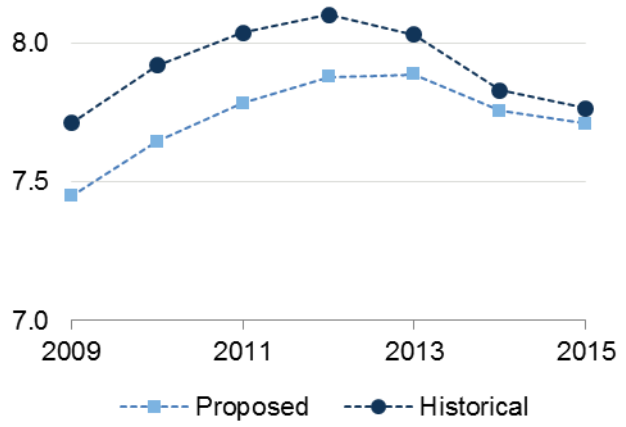


Chart 7F: Comparison of life expectancies for females aged 85



8. Convergence and critical damping

The basic principle of the current Model is to project rates of mortality improvement by interpolating between current (Initial) rates, which are estimated from historical data, and assumed long-term rates, which are set by users of the Model. This process is carried out separately for age-period and cohort components, and these are combined to give the overall mortality improvements.

In its review of the Model, the Committee considered various options for convergence from the initial rates to the long-term rate. In this section we describe the motivations for considering a change, and compares two possible options.

Two key criticisms have been made of the current Model:

(1) "Direction of travel". The current approach to convergence does not allow for the "direction of travel" of mortality improvements. Under the Core parameters the slope of mortality improvements is assumed to be flat at the start year of the model. The advanced parameters can be used to adjust the slope of improvements, but this is rather cumbersome as the advanced parameters specify the "proportion remaining at midpoint".

(2) "Experience item". We would ideally like projections from next year's model to be the same as those from this year's model, if mortality experience were as expected. Under the current approach this can only be achieved by shortening the convergence periods by one year each year; or equivalently specifying the convergence periods by using fixed calendar years. This is undesirable, not least because some convergence periods are as short as five years, so the ends of some convergence periods from the original CMI_2009 model have already been reached.

In this section we describe two methods:

A. One is based on the current method, using a cubic polynomial over a finite convergence period, and is the option proposed in Working Paper 90.

B. The other is based on "critical damping", with asymptotic convergence over an infinite period.

Both of these have the following features:

1. The initial rates can be specified.
2. The long-term rate can be specified.
3. Mortality improvements are flat in the very long term.
4. A parameter can be used to allow for direction of travel.

Option A offers a modest improvement that addresses criticism (1) by allowing for direction of travel. Option B would address criticisms (1) and (2), but would be a bigger change as convergence would no longer be over a finite period.

8.1 Cubic convergence

This is described in Working Paper 90, and is a slight amendment of the current method.

For a particular age or cohort, we write $f(t)$ for the mortality improvement at time t, where $f(t)$ is defined by:

$$f(t) = L + (I - L)\left(1 - 3\left(\frac{t}{T}\right)^2 + 2\left(\frac{t}{T}\right)^3\right) + Dt\left(1 - \frac{t}{T}\right)^2 \qquad \text{for } 0 \le t \le T$$

$$f(t) = L \qquad \text{for } t > T$$

with parameters:

L	the long-term rate of mortality improvements
I	the initial rate of mortality improvements
D	the "direction of travel"; i.e. the initial slope of mortality improvements
T	the convergence period

The function $f(t)$ has been chosen so that:

$f(0) = I$	matching the specified initial rate of improvements
$f(t) = L$ for $t \ge T$	matching the specified long-term rate of improvements
$f'(0) = D$	matching the specified direction of travel
$f'(t) = 0$ for $t \ge T$	improvements are flat beyond time T

Under the current approach, users of the Model specify the proportion remaining at mid-point, p. Under the proposed approach users could choose to specify D directly to specify the initial slope of mortality improvements. This would allow for "direction of travel".

If we set $D = \frac{1}{T}(8p - 4)(I - L)$ we obtain:

$$f(t) = L + (I - L)\left(1 + (8p - 4)\left(\frac{t}{T}\right) + (5 - 16p)\left(\frac{t}{T}\right)^2 + (8p - 2)\left(\frac{t}{T}\right)^3\right) \qquad \text{for } 0 \le t \le T$$

and so:

$$f\left(\frac{T}{2}\right) = L + p(I - L).$$

This shows that we can retain the current approach to convergence of specifying the proportion remaining at mid-point, p. Setting $p = \frac{1}{2}$ is equivalent to setting $D = 0$ and gives the current Core parameterisation.

8.2 Critical damping

Option B is based on the theory of "critical damping". We first introduce the function and then consider its motivation.

Under this approach the convergence pattern $g(t)$ is defined as:

$$g(t) = L + (I - L)\left(1 + \frac{t}{T}\right)\exp\left(-\frac{t}{T}\right) + Dt\exp\left(-\frac{t}{T}\right)$$

and we have:

$g(0) = I$	matching the specified initial rate of improvements
$\lim_{t\to\infty} g(t) = L$	matching the specified long-term rate of improvements
$g'(0) = D$	matching the specified direction of travel
$\lim_{t\to\infty} g'(t) = 0$	improvements are flat beyond time T

So Option B is similar to the "direction of travel" version of Option A. However while mortality improvements under Option A reach the long-term rate after a finite time T, mortality improvements under Option B tend asymptotically to the long-term rate. The parameter T is no longer a fixed convergence period. Instead we refer to it as the "relaxation time".

Option B is based on "critical damping"[6]. In that context "damping" refers to reducing the oscillations of a system, and "critical" means that the object returns to its equilibrium as quickly as possible without overshooting. In the context of mortality, we can think of a long-term rate as the equilibrium position, and current mortality improvements as deviations from that. Option B satisfies the differential equation:

$$\frac{d^2g}{dt^2} + \frac{2}{T}\frac{dg}{dt} + \frac{1}{T^2}(g - L) = 0$$

Critical damping addresses criticism (2) above, the "experience item". We prove this in Section 8.4.

Expressing the convergence in this form may also be helpful for some models as we could apply difference equations to parameters within a model, rather than requiring convergence to be bolted on afterwards. We return to this when considering "integrated" models in Section 10.

8.3 Comparison and examples

Our expressions for $f(t)$ and $g(t)$ could be simplified, but have been written in their current forms in order to highlight common features.

We can write $f(t) = L + (I - L)f_1(t) + Df_2(t)$ and $g(t) = L + (I - L)g_1(t) + Dg_2(t)$ so that the mortality improvement is the sum of the long-term rate and two components that model the excess over the long-term rate. Functions f_1 and g_1 determine the convergence between initial and long-term rates in the absence of direction of travel, and f_2 and g_2 determine the impact of the initial direction of travel.

Table 8.1 shows that f_1 and g_1 both have an initial value of 1, and initial slope of 0, and long-term values and slopes of zero; and f_2 and g_2 both have an initial value of 0, and initial slope of 1, and long-term values and slopes of zero.

Table 8.1: Components of convergence functions

Function	Initial value	Initial slope	Long-term value	Long-term slope
$f_1(t) = 1 - 3\left(\frac{t}{T}\right)^2 + 2\left(\frac{t}{T}\right)^3$	$f_1(0) = 1$	$f_1'(0) = 0$	$f_1(T) = 0$	$f_1'(T) = 0$
$g_1(t) = \left(1 + \frac{t}{T}\right)\exp\left(-\frac{t}{T}\right)$	$g_1(0) = 1$	$g_1'(0) = 0$	$\lim_{t\to\infty} g_1(t) = 0$	$\lim_{t\to\infty} g_1'(t) = 0$
$f_2(t) = t\left(1 - \frac{t}{T}\right)^2$	$f_2(0) = 0$	$f_2'(0) = 1$	$f_2(T) = 0$	$f_2'(T) = 0$
$g_2(t) = t\exp\left(-\frac{t}{T}\right)$	$g_2(0) = 0$	$g_2'(0) = 1$	$\lim_{t\to\infty} g_2(t) = 0$	$\lim_{t\to\infty} g_2'(t) = 0$

[As an aside, the functions f_1 and f_2 are equal to two of the four Hermite basis functions[7]. The other two are not necessary in our case.]

[6] https://en.wikipedia.org/wiki/Damping
[7] https://en.wikipedia.org/wiki/Cubic_Hermite_spline

The core calibration of the current Model has all the necessary parameter values for the cubic convergence function. For the critical damping approach we need to specify values for the relaxation time for each age and cohort.

As a starting point, for comparison with the current Model, these could be a constant multiple of the convergence periods. For the purpose of illustration in this paper we have taken this to be one-third of the convergence period. This is a pragmatic round-number choice which seems to give broadly similar life expectancies.

These four functions are plotted in Charts 8A and 8B for sample parameters. We use $T = 20$ for f_1 and f_2 and $T = 6\frac{2}{3}$ for g_1 and g_2 since T plays a different role in each case.

Chart 8A: Functions f_1 and g_1

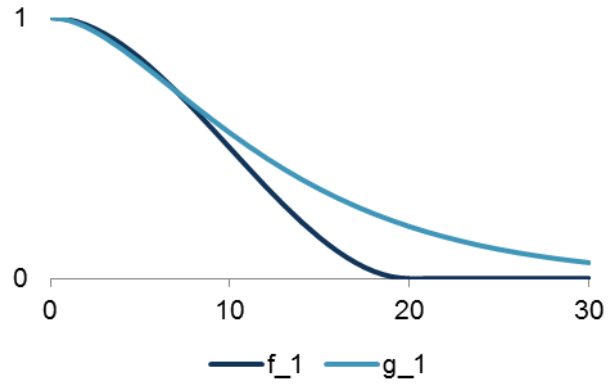
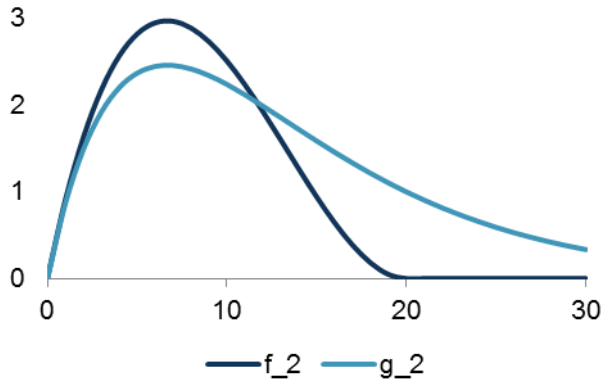


Chart 8B: Functions f_2 and g_2



Charts 8C and 8D show convergence patterns $f(t)$ and $g(t)$ using the same values of T, and with sample values of $I = 3\%$ and $L = 1.5\%$; with $D = +0.2\%$ in Chart 8C and $D = -0.2\%$ in Chart 8D.

Chart 8C: Functions f and g with $D = +0.2\%$

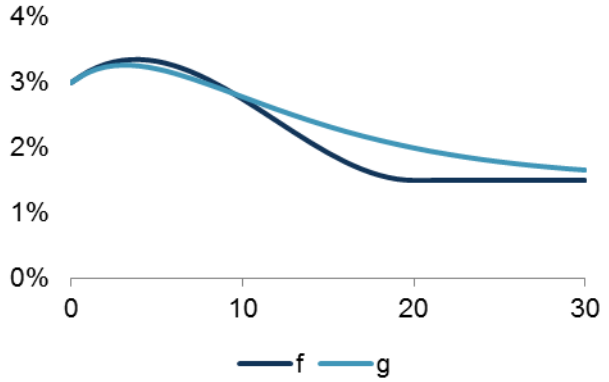
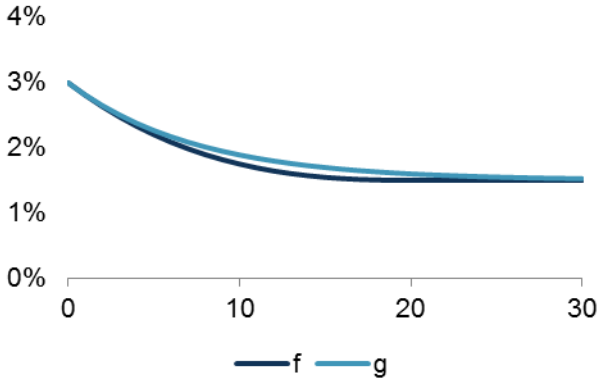


Chart 8D:Functions f and g with $D = -0.2\%$



Critical damping offers an advantage over the current method, as it avoids the “experience item” criticism, and adopting it need not have a material impact on life expectancies. However it would represent a conceptual change from a finite convergence period to asymptotic convergence.

8.4 Proof of "no experience item" with critical damping

As discussed, a useful property of a projection method is that if experience is as expected, recalibrating the projection at a future time should give exactly the same projection.

In this section we demonstrate that critical damping satisfies this property.

Calibrating at time t = 0

Write $g(t; I_0, D_0)$ for the mortality improvements, for a particular age or cohort, projected using the critical-damping convergence function, calibrated to the initial improvement I_0 and initial direction of travel D_0 at time $t = 0$. From the definition:

$$g(t; I_0, D_0) = L + (I - L)\left(1 + \frac{t}{T}\right)\exp\left(-\frac{t}{T}\right) + D_0 t\exp\left(-\frac{t}{T}\right)$$

Write $\omega = \frac{1}{T}$ to simplify notation, and rearrange the terms to get:

$$g(t; I_0, D_0) = L + \exp(-\omega t)[D_0 t + (I_0 - L)(1 + \omega t)] \quad (1)$$

By the product rule, this has slope:

$$g'(t; I_0, D_0) = \exp(-\omega t)\,[D_0 + (I_0 - L)\omega] - \omega.\exp(-\omega t)[D_0 t + (I_0 - L)(1 + \omega t)]$$

which simplifies to:

$$g'(t; I_0, D_0) = \exp(-\omega t)[(1 - \omega t)D_0 - \omega^2 t(I_0 - L)] \quad (2)$$

Setting $t = 0$ confirms that $g(0; I_0, D_0) = I_0$ and $g'(0; I_0, D_0) = D_0$.

Calibrating at time t = N

We can also define $h(t; I_N, D_N)$ for mortality improvements calibrated at a future time $t = N$ to the mortality improvements and slope at that point.

We have, similar to (1):

$$h(t; I_N, D_N) = L + \exp(-\omega(t - N))[D_N(t - N) + (I_N - L)(1 + \omega(t - N))] \quad (3)$$

We can calculate the improvement and direction of travel at time N using (1) and (2):

$$I_N = g(N; I_0, D_0) = L + \exp(-\omega N)[D_0 N + (I_0 - L)(1 + \omega N)] \quad (4)$$

$$D_N = g'(N; I_0, D_0) = \exp(-\omega N)[(1 - \omega N)D_0 - \omega^2 N(I_0 - L)] \quad (5)$$

Equivalence of the two approaches

Substituting (4) and (5) into (3) gives:

$$h(t; I_N, D_N) = L + \exp(-\omega(t - N))[\exp(-\omega N)[(1 - \omega N)D_0 - \omega^2 N(I_0 - L)](t - N) + \exp(-\omega N)[D_0 N + (I_0 - L)(1 + \omega N)](1 + \omega(t - N))] \quad (6)$$

Cancel the $\exp(-wN)$ terms:

$$h(t; I_N, D_N) = L + \exp(-\omega t))\big[[(1 - \omega N)D_0 - \omega^2 N(I_0 - L)](t - N) + [D_0 N + (I_0 - L)(1 + \omega N)](1 + \omega(t - N))\big]$$

Group the terms in D_0 and $(I_0 - L)$:

$$h(t; I_N, D_N) = L + \exp(-\omega t))\left[D_0[(1 - \omega N)(t - N) + N(1 + \omega(t - N))] + (I_0 - L)\big[(1 + \omega N)\big(1 + \omega(t - N)\big) - \omega^2 N(t - N)\big]\right]$$

Then simplify to get:

$$h(t; I_N, D_N) = L + \exp(-\omega t))[D_0 t + (I_0 - L)(1 + \omega t)]$$

This shows that $h(t; I_N, D_N) \equiv g(t; I_0, D_0)$; i.e. if experience is as expected, then recalibrating this projection method at any future time will give the same projection.

9. Other models – calculating initial improvements

The Committee considered a wide range of models before making the proposals described in Working Paper 90.

We considered two broad approaches:

- a "two-phase" approach, as in the current Model and our proposal, where we first determine initial mortality improvements and their age-period and cohort components, and then project them; and
- an "integrated" approach that combines the two phases into one.

This section summarises our investigations into possible models to determine initial improvements under the two-phase approach. Section 10 considers the integrated approach.

9.1 Smoothing

For our purpose, we want whatever model we choose to produce a smooth central projection rather than a stochastic projection. The current Model achieves this by using splines. Our approach applies regularisation penalties directly to parameters, as described in Section 7.4 of Working Paper 90.

Eilers et al (2015) note that "If the data are observed on an equidistant grid and only smoothed values on that grid are wanted, one can just as well use the identity matrix as a basis". This applies in our case, as we are only interested in annual mortality rates and improvements, i.e. a regular grid with annual spacing. The use of the "identity matrix as a basis" is equivalent to our regularisation approach.

9.2 Model criteria

Before introducing candidate models we consider four desirable features that we would prefer a model to have.

A good fit to historical data, without over-fitting to artefacts of the data

We can assess the goodness of fit by looking at deviance and deviance residuals. However we recognise that we want our fitted initial mortality improvements to be smooth. We could improve the deviance by making the fit less smooth, but that would not then be helpful for projection. We considered the use of information criteria, as used in the current Model, but did not find these to be helpful, given the aim of a smooth model.

Parameters that have a tangible real-world interpretation and can be seen to be plausible

The current Model split improvements into age, period and cohort components. This seems helpful to us, and we look for the parameters and improvements to be plausible e.g. cohort parameters should represent cohort effects, rather than being used to model effects by age or period.

When considering Basis models (in Section 9.7) we prefer a model whose parameters can be interpreted more simply as "mortality at low/medium/high ages" rather than "mortality, slope, and curvature".

Robustness to changing datasets

Within the Model we project age-period and cohort components of improvements differently. Because of this, the answers are sensitive to the split of improvements. We expect that in a suitable model we should get similar answer for parameters if we vary the age range of the calibration data.

Sensitivity to the range of years is arguably less important; we recognise that the levels of mortality improvements do change over time and changing the range of years puts more or less emphasis on certain periods.

All other things being equal, a simpler model with a reasonable run-time

A simpler model is likely to be easier to understand and to communicate, and to have fewer poorly-understood artefacts. A short run-time is convenient and allows more tests and sensitivities to be considered. Also, a longer run-time can be indicative of a fragile model that suffers from problems with identifiability of parameters.

9.3 Candidate formulae

We distinguish between a longevity "model" and its "formula":

- The "formula" is how mortality is expressed as a combination of components; e.g. the formula for Lee-Carter is $\alpha_x + \beta_x \kappa_t$
- The "model" encompasses broader aspects; e.g. it includes the formula, identifiability constraints, the fitting process and the data used.

Some models define their formulae in terms of $\text{logit}\, q_{x,t}$ and some in terms of $\log m_{x,t}$. Whilst we considered formulae from models that were defined in both ways, we have only applied these formulae to $\log m_{x,t}$.

Table 9.1 shows the formulae that we considered, together with the papers where they originally appeared.

For ease of reference we have named the formulae:

- The "C" column refers to the names used in Cairns et al (2009) and Cairns et al (2014). Beware that other authors have used the name "M9" to refer to different models; e.g. O'Hare and Li (2011) use it to refer to Plat (2009).
- The "HR" column refers to the names used in Haberman and Renshaw (2011).
- The "CMI" column has names for formulae introduced in this paper:
  - Names beginning with "A" are variants of the APCI formula.
  - Names beginning with "B" are variants of "basis" formulae, which we describe in Section 9.7.

Table 9.1 Formulae referred to in this section

Names C	Names HR	Names CMI	Formula	Original paper
M1	LC	-	$\alpha_x + \beta_x \kappa_t$	Lee and Carter (1992)
-	LC2	-	$\alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}$	Renshaw and Haberman (2003)
M2	M	-	$\alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(2)}\gamma_{t-x}$	Renshaw and Haberman (2006)
-	H1	-	$\alpha_x + \beta_x \kappa_t + \gamma_{t-x}$	Renshaw and Haberman (2006)
M3	H0	-	$\alpha_x + \kappa_t + \gamma_{t-x}$	Currie (2006)
M5	M5	-	$\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}$	Cairns et al (2006)
-	-	-	$\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}$	Currie (2010)
M6	M6	-	$\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x}$	Cairns et al (2009)
M7	M7	-	$\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + ((x-\bar{x})^2 - \sigma^2)\kappa_t^{(3)} + \gamma_{t-x}$	Cairns et al (2009)
M8	M8	-	$\kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (x_c - x)\gamma_{t-x}$	Cairns et al (2009)
M9	-	B3Q	$\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + ((x-\bar{x})^2 - \sigma^2)\kappa_t^{(3)} + \gamma_{t-x}$	Cairns et al (2014)
-	M5*	-	$\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (\bar{x}-x)^+\kappa_t^{(3)}$	Haberman and Renshaw (2011)
-	M6*	B3L	$\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (\bar{x}-x)^+\kappa_t^{(3)} + \gamma_{t-x}$	Plat (2009)
-	M7*	B4X	$\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (\bar{x}-x)^+\kappa_t^{(3)} + ((\bar{x}-x)^2 - \sigma^2)\kappa_t^{(4)} + \gamma_{t-x}$	Haberman and Renshaw (2011)
-	M8*	-	$\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (\bar{x}-x)^+\kappa_t^{(3)} + (x_c - x)\gamma_{t-x}$	Haberman and Renshaw (2011)
-	-	-	$\alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + ((\bar{x}-x)^+ + [(\bar{x}-x)^+]^2)\,\kappa_t^{(3)} + \gamma_{t-x}$	O'Hare and Li (2011)
-	-	A1	$\alpha_x + \beta_x\,(t-\bar{t}) + \kappa_t + \gamma_{t-x}$	
-	-	A2	$\alpha_x + \beta_x\,(t-\bar{t}) + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x}$	
-	-	A3	$\alpha_x + \beta_x\,(t-\bar{t}) + \kappa_t + \psi_x\gamma_{t-x}$	
-	-	A4	$\alpha_x + \beta_x\,(t-\bar{t}) + \kappa_t + \phi_t + \gamma_{t-x}$	
-	-	B3	$\alpha_x + B_x^{(1)}\kappa_t^{(1)} + B_x^{(2)}\kappa_t^{(2)} + B_x^{(3)}\kappa_t^{(3)} + \gamma_{t-x}$	
-	-	B4	$\alpha_x + B_x^{(1)}\kappa_t^{(1)} + B_x^{(2)}\kappa_t^{(2)} + B_x^{(3)}\kappa_t^{(3)} + B_x^{(4)}\kappa_t^{(4)} + \gamma_{t-x}$	

9.4 Initial assessment

This section summarises our initial assessment of formulae developed by others. Some can be rejected based on their structure, or on existing published work. Others were investigated further by us.

Cohort terms

The formulae M1/LC, LC2, M5, M5* and Currie (2010) have no cohort terms. Given the prominence of cohort effects in mortality improvements in England & Wales in recent years, and its inclusion in the current Model, we do not consider these formulae to be suitable for our purpose.

Robustness and convergence

We do not consider formula M2/M due to concerns raised by others. Cairns et al (2009) noted that M2 "seems to produce results that lack robustness, because the parameter estimates jump to a qualitatively quite different solution when we use less data" and Currie (2016) "found that for a particular parameterization convergence could be very fast, very slow or even fail completely".

For similar reasons we do not consider M8 either. Cairns et al (2009) found that "for some datasets, the M8 fitting program was very slow to converge. We found a similar problem with M2 and put this down to the possible existence of multiple maxima in the likelihood function and the consequential risk of parameter instability".

Haberman and Renshaw (2011) raise concerns about formula M8* (as well as M2/M) and we exclude it for this reason.

M3/H0/APC

The M3/H0 formula (also known as the APC formula) is a special case of the APCI formula (by setting $\beta_x \equiv 0$). Chart 4B shows that the β_x parameters in the APCI model are significantly non-zero, so we consider the APC formula to be inferior to the APCI formula.

O'Hare and Li

The formula of O'Hare and Li (2011) is a variant of Plat (2009), that is intended to cope well with ages as young as 5. We do not consider this further as our focus is on the adult population.

Considered further

The formulae developed by others that we consider further are:

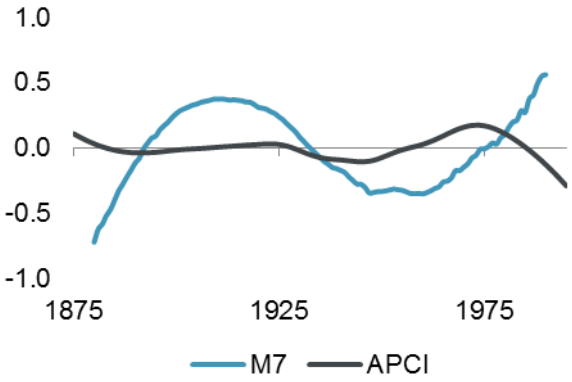
- M6 and M7 are considered together in Section 9.5.
- H1 is considered in Section 9.6.
- The M6*, M9 and M7* formulae are all examples of "Basis" formulae, introduced in Section 9.7:
  - The Plat formula, M6*, is considered in Section 9.8. We refer to this as B3L in our notation.
  - M9 is considered under the name B3Q in Section 9.9.
  - M7* is considered under the name B4X in Section 9.10.

9.5 Formulae M6 and M7

We understand that M7 is widely used for pensioner age ranges. However when fitting to the age range from 20 to 100 we find that the deviance residuals show a poor fit. The formula has a quadratic structure to the age-period components of mortality rates and improvements, and this struggles with the more complex shape of mortality rates over a wide age range.

Chart 9A compares the cohort parameters from the M7 and APCI formulae, both fitted to data for England & Wales males for ages 20-100 and years 1975-2015. (Note that the APCI case is smoothed, and the M7 case is not and excludes cohorts with fewer than five observations.)

Chart 9A: Cohort parameters for M7 and APCI



The cohort term from the M7 formula looks unrealistic, having larger values than for the APCI model, and a strongly cubic shape. The cohort term is being used primarily to reflect patterns of mortality rates by age, rather than purely cohort effects. Whilst the aggregate mortality improvements of the M7 formula are reasonable, the split between age-period and cohort components is not.

As a result, this formula seems unsuitable for modelling a wide age range; and the simpler nested M6 formula is also unsuitable.

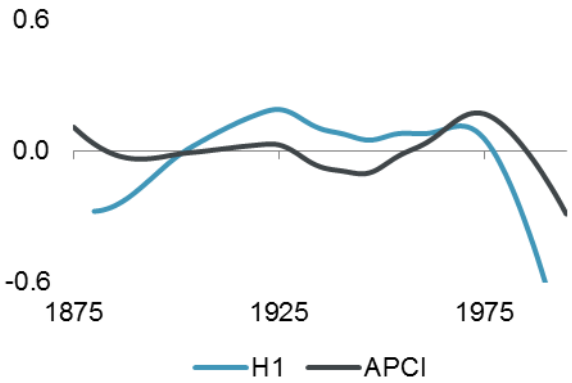
9.6 Formula H1

Chart 9B compares the cohort parameters for H1 and APCI (using the same approach as for Chart 9A). This shows that the shape of the cohort parameters is quite different to that of the APCI formula.

We also found that convergence of the H1 formula was slow, taking over 30 minutes to fit, compared to under a minute for the APCI formula. Slow convergence can be indicative of problems with identifiability.

We do not consider formula H1 further.

Chart 9B: Cohort parameters for H1 and APCI



9.7 "Basis" formulae

Several of the formulae in Table 9.1 (including M6*, M7*, M9 and O'Hare and Li) can be expressed by the generic formula

$$\log m_{x,t} = \alpha_x + \sum_{i=1}^{N} B_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x}$$

where the number of period terms, N, and the basis functions, $B_x^{(i)}$, are specified in advance (i.e. the $B_x^{(i)}$ are exogenous) and the other parameters are fitted. This is similar to equation (1) of Hunt and Blake (2015), but without multiplying the cohort term γ_{t-x} by an age function.

The differences between the formulae in this class are the size of the basis (the value of N) and the choice of basis functions $B_x^{(i)}$.

The resulting mortality improvements are:

$$MI_{x,t} = -\sum_{i=1}^{N} B_x^{(i)} \nabla_t \kappa_t^{(i)} - \nabla_c \gamma_c$$

We can think of these as having age-period ($-\sum_{i=1}^{N} B_x^{(i)} \nabla_t \kappa_t^{(i)}$) and cohort ($-\nabla_c \gamma_c$) components. The basis formulae have more flexibility than the APCI formula for the shape of age-period mortality improvements to change over time.

In the next section we discuss the Plat formula, and consider different ways to express it as a Basis formula.

9.8 Plat formula

The Plat (M6*) formula is defined by:

$$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (\bar{x} - x)^+ \kappa_t^{(3)} + \gamma_{t-x}$$

This is an example of a Basis formula with a basis of :

$$\{1, (x - \bar{x}), (\bar{x} - x)^+\}$$

We can also express the same formula in another way. For an age range of L to H, consider a basis of linear splines:

$$\left\{S^{(1)}\left(\frac{x-c_1}{k}\right), S^{(1)}\left(\frac{x-c_2}{k}\right), S^{(1)}\left(\frac{x-c_3}{k}\right)\right\}$$

where:

$S^{(1)}(u) = 1 - |u|$ if $u \in [-1, +1]$

$S^{(1)}(u) = 0$ otherwise

and k is the knot spacing with:

$$k = \frac{1}{2}(H - L)$$

and c_i are knot positions with:

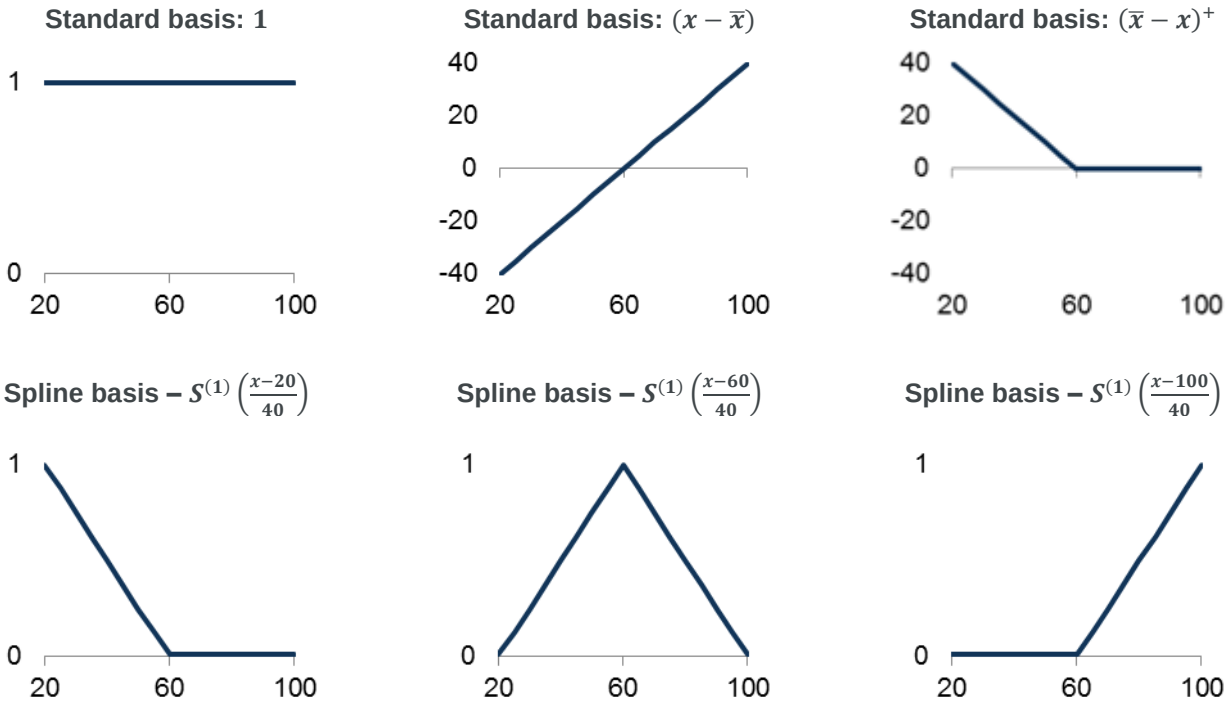
$$c_1 = L$$

$$c_2 = \frac{1}{2}(L + H)$$

$$c_3 = H$$

The two bases are illustrated in Chart 9C, for the age range 20-100, consistent with that used in the calibration phase of the proposed model.

Chart 9C: Standard and spline basis functions for the Plat (M6*) model



The two bases can be related by:

$$S^{(1)}\left(\frac{x-c_1}{k}\right)+S^{(1)}\left(\frac{x-c_2}{k}\right)+S^{(1)}\left(\frac{x-c_3}{k}\right)\equiv 1$$

$$\frac{1}{2}(H-L)\left(S^{(1)}\left(\frac{x-c_3}{k}\right)-S^{(1)}\left(\frac{x-c_1}{k}\right)\right)\equiv(x-\bar{x})$$

$$\frac{1}{2}(H-L)S^{(1)}\left(\frac{x-c_1}{k}\right)\equiv(\bar{x}-x)^+$$

In the language of linear algebra, these two bases have the same "span"; i.e. the set of shapes that can be made from:

$$S^{(1)}\left(\frac{x-c_1}{k}\right)\kappa_t^{(1)}+S^{(1)}\left(\frac{x-c_2}{k}\right)\kappa_t^{(2)}+S^{(1)}\left(\frac{x-c_3}{k}\right)\kappa_t^{(3)}$$

over the age range L to H is exactly the same as the set of shapes that can be made from:

$$\kappa_t^{(1)}+(x-\bar{x})\kappa_t^{(2)}+(\bar{x}-x)^+\kappa_t^{(3)}$$

Because the span is the same, both formulae have the same flexibility to fit mortality rates and so they will have the same goodness of fit. Given this, why consider using a different basis? There are two key reasons:

1. Using the basis of linear splines means that the fitted $\kappa_t^{(i)}$ in this formula have the same order of magnitude as each other. This should make it reasonable to specify a single smoothing parameter for all of the $\kappa_t^{(i)}$. If we instead used the original form of the Plat formula then the $\kappa_t^{(i)}$ would have very different orders of magnitude, and the smoothing parameters would have to vary accordingly.
2. We can generalise the approach of expressing the Plat model using a basis of three linear splines, to expressing other models using a basis of N splines of some power. We consider this in Section 9.10.

Plat (2009) says that the factor $\kappa_t^{(3)}$ is added to capture "the dynamics of mortality rates at lower ages (up to age 40/50) [which] can be (significantly) different at some times", due to the term $(\bar{x} - x)^+$ being non-zero only for the younger half of the age range. Re-expressing the formula as:

$$S^{(1)}\left(\frac{x-c_1}{k}\right)\kappa_t^{(1)} + S^{(1)}\left(\frac{x-c_2}{k}\right)\kappa_t^{(2)} + S^{(1)}\left(\frac{x-c_3}{k}\right)\kappa_t^{(3)}$$

shows that the effect of the Plat model is to allow for the dynamics of mortality rates to differ for younger and older ages, rather than giving any special prominence to younger ages.

For the Plat formula we have the identifiability constraints:

$\sum_t \kappa_t^{(1)} = \sum_t \kappa_t^{(2)} = \sum_t \kappa_t^{(3)} = 0$	i.e. $\kappa_t^{(i)}$ has a mean of zero for all i
$\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = 0$	i.e. a quadratic fit to γ_{t-x} would be zero

9.9 M9 formula

We saw above that the Plat formula can be re-expressed as a Basis formula using three linear splines. In a similar way, a Basis formula with a basis of three quadratic splines is equivalent to the Cairns-Blake-Dowd M9 formula.

The M9 formula is defined by:

$$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + ((x - \bar{x})^2 - \sigma^2)\kappa_t^{(3)} + \gamma_{t-x}$$

and has identifiability constraints:

$\sum_t \kappa_t^{(1)} = \sum_t \kappa_t^{(2)} = \sum_t \kappa_t^{(3)} = 0$	i.e. $\kappa_t^{(i)}$ has a mean of zero for all i
$\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = \sum_c c^3\gamma_c = 0$	i.e. a cubic fit to γ_{t-x} would be zero

This is an example of a Basis formula with a basis of:

$$\{1, (x - \bar{x}), ((x - \bar{x})^2 - \sigma^2)\}$$

We can also consider an alternative basis of quadratic basis splines:

$$B_x^{(i)} = S^{(2)}\left(\frac{x-c_i}{k}\right)$$

where $S^{(2)}(u)$ is:

$\frac{3}{4} - u^2$	if $u \in \left[-\frac{1}{2}, +\frac{1}{2}\right]$
$\frac{1}{2}u^2 - \frac{3}{2}\lvert u\rvert + \frac{9}{8}$	if $\lvert u\rvert \in \left[\frac{1}{2}, \frac{3}{2}\right]$
0	otherwise

and k is the knot spacing with:

$$k = H - L$$

and c_i are knot positions with:

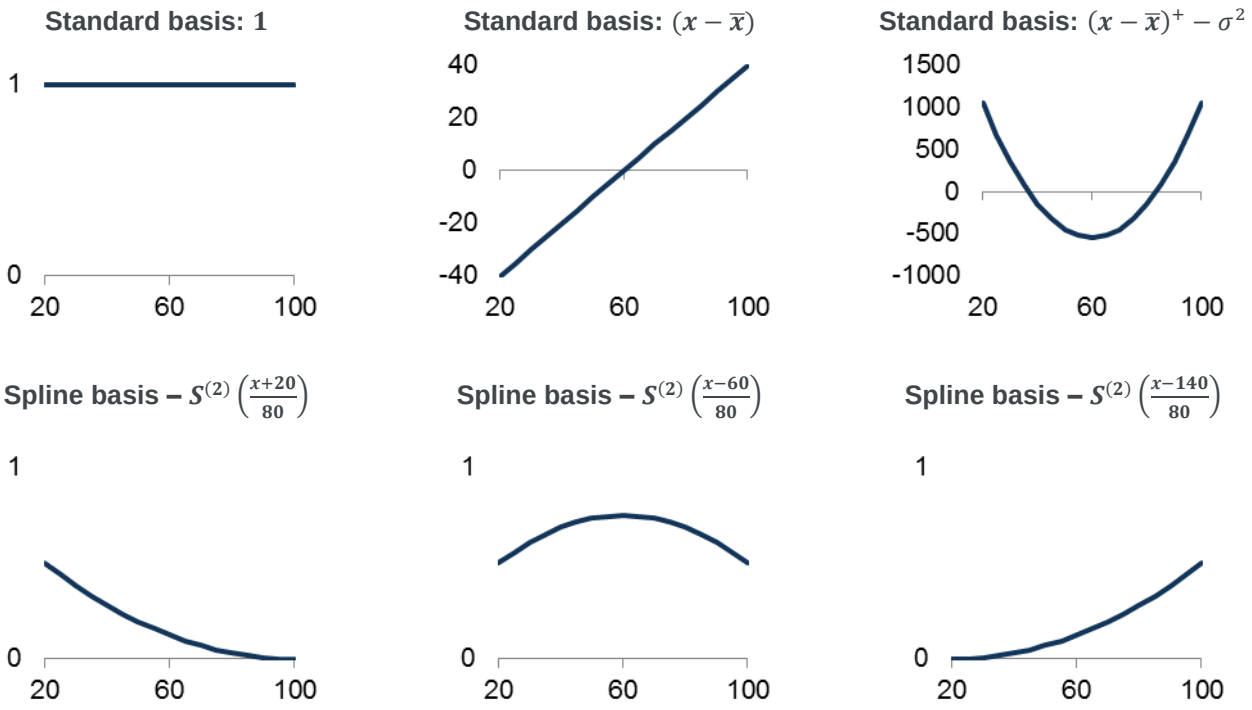
$$c_1 = \frac{1}{2}(3L - H)$$

$$c_2 = \frac{1}{2}(L + H)$$

$$c_3 = \frac{1}{2}(3H - L)$$

The two bases are illustrated in Chart 9D, for the age range 20-100, consistent with that used in the calibration phase of the proposed model.

Chart 9D: Standard and spline basis functions for the M9 model



Over the age range from L to H the two bases have the same span and can be related by:

$$B_x^{(1)} + B_x^{(2)} + B_x^{(3)} \equiv 1$$

$$\left(B_x^{(3)} - B_x^{(1)}\right)(H-L) \equiv (x-\bar{x})$$

$$\tfrac{1}{4}\left(3B_x^{(1)} - B_x^{(2)} + 3B_x^{(3)}\right)(H-L)^2 \equiv (x-\bar{x})^2$$

9.10 B3 and B4 formulae

The previous sections showed that the Plat and M9 formulae can be expressed as Basis formulae with three basis splines.

We refer to these, in consistent notation, as:

- B3L – three linear splines – Plat; and
- B3Q – three quadratic splines – M9.

We can extend this approach to a greater number of basis splines. In particular we consider:

- B4L – four linear splines;
- B4Q – four quadratic splines; and
- B4C – four cubic splines.

The number of identifiability constraints increases with the power of the basis splines. We require that:

$\sum_t \kappa_t^{(1)} = \sum_t \kappa_t^{(2)} = \sum_t \kappa_t^{(3)} = 0$ for B3; and additionally $\sum_t \kappa_t^{(4)} = 0$ for B4

For linear splines we constrain a quadratic fit to the cohort parameters to be zero; i.e.:

$$\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = 0$$

For quadratic splines we constrain a cubic fit to the cohort parameters to be zero; i.e.:

$$\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = \sum_c c^3\gamma_c = 0$$

For cubic splines we constrain a quartic fit to the cohort parameters to be zero; i.e.:

$$\sum_c \gamma_c = \sum_c c\gamma_c = \sum_c c^2\gamma_c = \sum_c c^3\gamma_c = \sum_c c^4\gamma_c = 0$$

We also define the non-standard formula B4X. This is equivalent to formula M7* and uses a mixture of linear and quadratic functions. We use three linear splines (as for B3L), and also a quadratic:

$$\frac{3}{2}\left(\frac{2x-H-L}{H-L}\right)^2 - \frac{1}{2}$$

This is a Legendre polynomial – it is a scaled version of the original $((\bar{x}-x)^2-\sigma^2)$, chosen so that it has a range of [-0.5,+1] to be broadly consistent with the linear splines.

For B4X we have used the same identifiability constraints as for B4L. We found that results for B4X offered no advantage over the conceptually simpler B4L, B4Q and B4C formulae, so we do not consider B4X further.

9.11 Preference among B3 and B4 formulae

In this section we consider the results of fitting the various B3 and B4 formulae compared with the APCI formula.

The purpose of these candidate formulae within the Model is to determine the age-period and cohort components of initial mortality improvements. The split of mortality improvements between age-period and cohort is particularly important because we assume different convergence periods when they are projected.

As a result, the shape of the cohort component was a key consideration when deciding between the various B3 and B4 models.

Charts 9E to 9H show cohort parameters and cohort improvements for fits of various Basis models. In each case results are from fits to data for England and Wales males for 1975-2015, and we apply regularisation to parameters so that they are smooth.

We note that we refined our approach to the APCI model after we had carried out the analysis in this section; i.e. the specification of the APCI model in charts 9E to 9J is slightly different to that proposed in Working Paper 90.

Chart 9E: Cohort parameters γ_{t-x} for B3

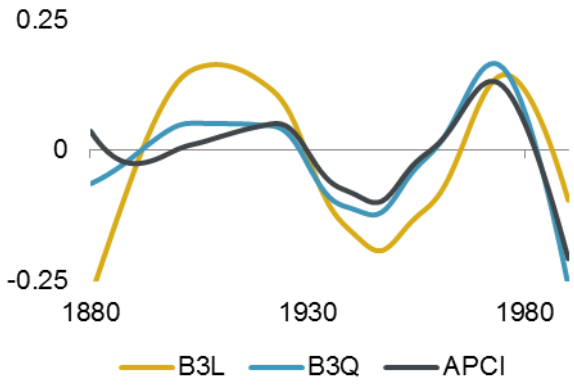


Chart 9F: Cohort parameters γ_{t-x} for B4

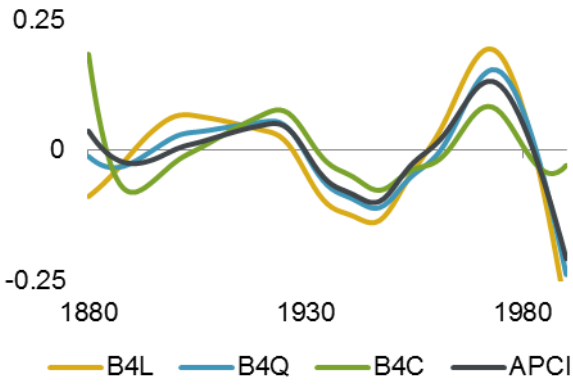


Chart 9G: Cohort improvements for B3

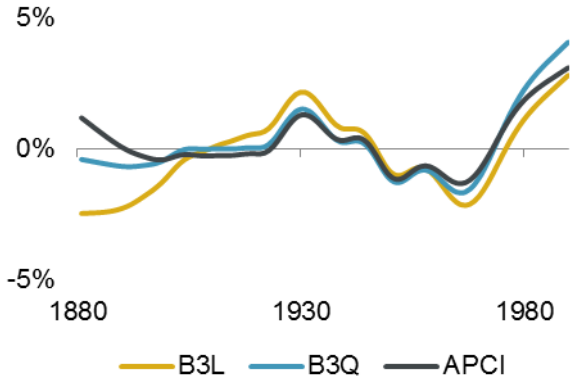


Chart 9H: Cohort improvements for B4

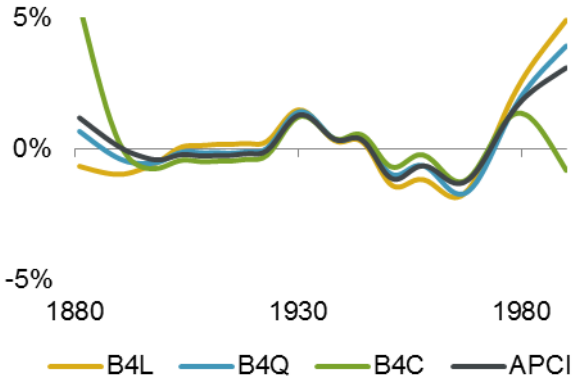


Chart 9E suggests that the cohort parameters for B3L are out of line with other runs. This appears to be due to the lack of a cubic element in the identifiability criteria – the difference between the cohort parameters for B3L and B3Q can be fitted very closely by a cubic polynomial.

Chart 9H suggests that the cohort improvements for B4C are out of line with other runs for the highest and lowest ages. In particular B4C suggests a negative cohort component at the youngest ages, which does not seem plausible in comparison with the pattern of crude mortality improvements. The B4C formula may suffer from having the most identifiability constraints – nine – so that identifiability transforms may mask genuine cohort effects.

Rejecting B3L and B4C leaves B3Q, B4L and B4Q. The cohort parameters for these are broadly similar to each other and to those for the APCI model. There is little to choose between these, either in terms of their parameters or their projected life expectancies. We prefer B3Q as it is a more familiar model (equivalent to the M9 formula that is an extension of the widely-used M7 formula) than B4L and B4Q which are our own inventions.

9.12 Comparison of M9/B3Q to APCI

In this section we compare B3Q, our preferred Basis formula, to the APCI formula that we propose to use in the Model. A key consideration, given the comments above regarding the sensitivity of the Model to the split into age-period and cohort components, is the stability of those components.

Charts 9I and 9J plot the age-period components of mortality improvements in 2005 (roughly the peak historical year for mortality improvements) for the B3Q and APCI models fitted to data for different age ranges: 20-100, 20-80, and 40-100. In each case this is for male England & Wales data for 1975-2015. We would hope to see that the improvements would be fairly stable as the age range changes; however the charts show clearly that the improvements from the APCI model are much more stable to changes in the age range used.

Chart 9I: Age-period component of mortality improvements in 2005 for the B3Q model

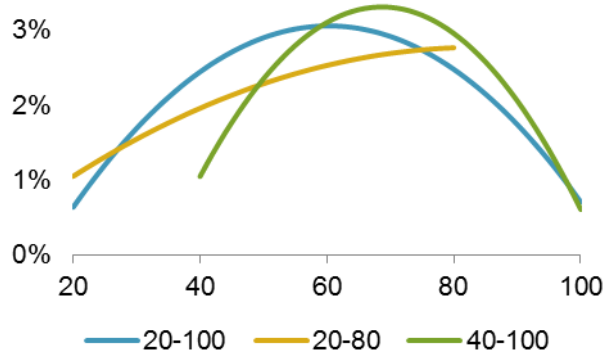
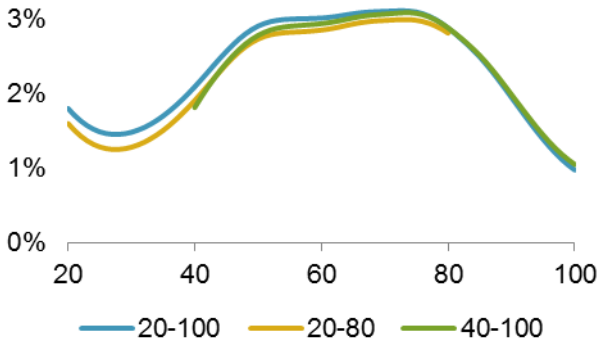


Chart 9J: Age-period component of mortality improvements in 2005 for the APCI model



9.13 Variants of the APCI formula

The standard APCI formula is:

$$\log m_{x,t} = \alpha_x + \beta_x(t - \bar{t}) + \kappa_t \ + \gamma_{t-x} \qquad \text{(A1)}$$

In this section we consider three variants.

Formula A2 allows the shape of mortality improvements by age to vary over time:

$$\log m_{x,t} = \alpha_x + \beta_x(t - \bar{t}) + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x} \qquad \text{(A2)}$$

In this formula a positive (or negative) value for $\kappa_t^{(2)}$ means that the age-period component of mortality is higher (or lower) at older ages. Fitting this formula materially improves the fit compared to A1, but does not seem helpful for projection. The values of $\kappa_t^{(2)}$ tell us what we already knew – that mortality in past four years has been particularly heavy at older ages – but it is not clear how that helps us to determine current levels of mortality improvements.

Formula A3 allows for "age-moderation of cohort":

$$\log m_{x,t} = \alpha_x + \beta_x(t - \bar{t}) + \kappa_t \ + \psi_x\gamma_{t-x} \qquad \text{(A3)}$$

Under the standard A1 formula the cohort component of mortality improvements for a particular cohort c is $\gamma_{c-1} - \gamma_c$ and so does change with time or age. Formula A3 multiplies this by an age-related parameter ψ_x. We have considered a number of shapes for ψ_x, all decreasing to zero at high ages, so that the cohort component of mortality improvements also falls to zero at high ages.

Allowing for age-moderation of cohort improves the fit for the APCI model, although convergence is quite slow, suggesting some difficulty with identifiability. While age-moderation of cohort improves the fit for the APCI model, a similar amendment to the M9/B3Q formula leads to a worse fit. This, together with the slow convergence, raises concerns that allowing for age-moderation of cohort may not be robust.

Formula A4 allows explicitly for annual noise:

$$\log m_{x,t} = \alpha_x + \beta_x(t - \bar{t}) + \kappa_t \ + \phi_t + \gamma_{t-x} \qquad \text{(A4)}$$

The intention is that κ_t represents smooth, persistent, underlying improvements and ϕ_t represents short-term, transient annual volatility. We would like to fit the model and then ignore the impact of ϕ_t so that we have smooth improvements. The formula itself does not distinguish between κ_t and ϕ_t so we require an additional penalty:

$$\lambda_\phi \sum_t \phi_t^2$$

so that the ϕ_t can be noisy, but cannot deviate far from zero; i.e. they cannot have any long-term trends.

Formula A4 fits the data better than A1, and varying the value of λ_ϕ allows for different splits between κ_t and ϕ_t, but the choice of λ_ϕ is not clear. We also tried adding a further identifiability constraint, but this did not prove fruitful.

Whilst A2, A3 and A4 are helpful in principle, and can fit the data better, they do not seem to offer any advantage in determining the initial components of underlying mortality improvements. We prefer to use A1, the proposed APCI formula, as this is simpler and seems to give acceptable and robust results.

Formula A1 seems to us to broadly satisfy the criteria of Section 9.2.

9.14 Assessment of formulae

In this section we summarise our assessment of formulae; firstly in the order in which the formulae were considered:

- M1/LC, LC2, M5, M5* and Currie (2010) lack cohort terms.
- M2/M, M8 and M8* all have concerns about their robustness and/or convergence.
- M3/H0/APC is a simpler form of the APCI formula and lacks the useful β_x parameters.
- O'Hare and Li (2011) is similar to M6*/B3L/Plat and intended to cope with very young ages, which we have no need for.
- M7 does not cope well with a wide age range, producing cohort parameters that seem implausible, as they are used to model the shape of mortality by age.
- M6 is a simpler version of M7.
- H1 has slow convergence, and its cohort parameters seem implausible.
- M7*/B4X shows no advantage over other, conceptually simpler, B4 formulae.
- B3L and B4C have cohort parameters and improvements which are out of line with other formulae.
- There is little to choose between B3Q, B4L and B4Q. We prefer B3Q as it is a more familiar formula. However the sensitivity of the age-period improvements in B3Q to changing age ranges is concerning, compared to the APCI formula.
- The variants A2, A3 and A4 give a better fit than the standard APCI formula, but they do not offer clear advantaged for projections.

Table 9.2 summarises this assessment in the same order and format as Table 9.1, for ease of comparison.

Table 9.2: Summary of our assessment of candidate formulae

Formula name			Summary of our assessment
M1	LC	-	No cohort terms
-	LC2	-	No cohort terms
M2	M	-	Concerns over convergence and the robustness of fitted parameters
-	H1	-	Slow convergence, and cohort parameters do not seem realistic
M3	H0	-	A simpler version of A1, lacking the β_x term
M5	M5	-	No cohort terms
Currie (2010)			No cohort terms
M6	M6	-	Subset of M7,
M7	M7	-	Cohort parameters do not seem realistic when fitted to the wide age range needed for the CMI Model
M8	M8	-	Concerns over convergence and the robustness of fitted parameters
M9	-	B3Q	Our second preference (behind the APCI formula A1) but parameters do not seem robust to changing the ages used for calibration
-	M5*	-	No cohort terms
-	M6*	B3L	Cohort terms seem out of line with other related models
-	M7*	B4X	More complex than B4L, B4Q and B4C and does not seem to offer any advantages
-	M8*	-	Concerns over convergence and the robustness of fitted parameters
O'Hare and Li (2011)			Similar to M6*/B3L/Plat, with a focus on very young ages that is not relevant for the CMI Model
-	-	A1	Our preferred formula, simple and robust, proposed for use in the next version of the CMI Model
-	-	A2	Allowing for varying shapes of age-period improvements improves the fit, but does not help with projections
-	-	A3	Age-moderation of cohort improves the fit for A3; but age-moderation does not seem robust when considering B3Q as well
-	-	A4	Allowing for annual noise improves the fit, but does not help with projections
-	-	B3	See specific comments on B3L and B3Q above
-	-	B4	B4L and B4Q give similar results to the simpler B3Q formula. Cohort terms for B4C seem out of line with other related models.

10. Other models – integrated approach

We have previously described the use of penalty functions to ensure smoothness of parameters and improvements (in Section 7.4 of Working Paper 90), and critical damping as a possible convergence function (in Section 8 of this paper).

In this section we put the two together and consider the possibility of an "integrated" approach that uses penalty functions to ensure convergence to the long-term rate.

In Section 8.2 we noted that our approach to critical damping for mortality improvements is based on the differential equation:

$$\frac{d^2g}{dt^2} + 2\omega\frac{dg}{dt} + \omega^2(g - L) = 0 \qquad (1)$$

where we write $\omega = 1/T$ to make notation clearer.

This has the solution:

$$g(t) = L + [(I - L)(1 + \omega t) + Dt]\exp(-\omega t)$$

where parameters I and D are set to define the initial mortality improvement and the initial slope of mortality improvements (i.e. the direction of travel).

Substituting the approximations:

$$\frac{d^2g}{dt^2} \approx g(t+1) - 2g(t) + g(t-1)$$

and

$$\frac{dg}{dt} \approx \frac{1}{2}(g(t+1) - g(t-1))$$

into (1) gives:

$$g(t+1) - 2g(t) + g(t-1) + 2\omega\frac{1}{2}\big(g(t+1) - g(t-1)\big) + \omega^2(g(t) - L) \approx 0$$

i.e.:

$$(1+\omega)g(t+1) + (\omega^2 - 2)g(t) + (1-\omega)g(t-1) - \omega^2 L \approx 0 \qquad (2)$$

This lets us express g by using the recurrence relation

$$g(t+1) = \frac{(2-\omega^2)g(t) + (\omega-1)g(t-1) + \omega^2 L}{(1+\omega)}$$

i.e.

$$g(t+1) = \frac{(2-\omega^2)}{(1+\omega)}g(t) + \frac{(\omega-1)}{(1+\omega)}g(t-1) + \frac{\omega^2 L}{(1+\omega)}$$

Note that:

- If $\omega = 0$ then $g(t+1) = 2g(t) - g(t-1)$ and we have linear extrapolation.
- If $g(t) = g(t-1) = L$ then $g(t+1) = L$, so once mortality improvements reach the long-term rate, they stay there.

Equation (2) is expressed in terms of $g(t)$, which is a mortality improvement. If we write $g(t)$ as the difference of two period terms; i.e. as:

$$g(t) = \kappa(t-1) - \kappa(t)$$

then (2) becomes:

$$(1+\omega)\kappa(t+1) + (\omega^2 - \omega - 3)\kappa(t) + (3 - \omega - \omega^2)\kappa(t-1) + (\omega - 1)\kappa(t-2) + \omega^2 L = 0 \qquad (3)$$

We can use the left-hand side of this formula to constrain the $\kappa(t)$ terms, by adding a penalty of:

$$\lambda \sum_t [(1+\omega)\kappa(t+1) + (\omega^2 - \omega - 3)\kappa(t) + (3 - \omega - \omega^2)\kappa(t-1) + (\omega - 1)\kappa(t-2) + \omega^2 L]^2$$

For future years, where there is no data, hence no deviance, this will force critical damping. For historical years, where there is data, this will enforce smoothness of the $\kappa(t)$ with the degree of smoothness depending on the value of λ.

Note that using a value of $\omega = 0$ means that we would extrapolate mortality improvements. This is a third-order penalty on the $\kappa(t)$. If we are not allowing for convergence within the penalty function then a better assumption would be that mortality improvements are constant; i.e. a second-order penalty on $\kappa(t)$, as proposed for the Core Model. Hence we should only use this approach where $\omega > 0$ with a "sensible" convergence period.

This integrated approach has the feature that the choice of long-term rate affects ones view of historical mortality improvements.

This can be seen as reasonable; for example:

- An actuary who assumes a long-term rate of 3% p.a. thinks that high levels of improvements will continue, and may be inclined to treat heavy recent mortality as just a blip, and assume comparatively high initial improvements.
- An actuary who assumes a long-term rate of 1% p.a. thinks that high levels of improvements will not persist, and may be inclined to treat heavy recent mortality as indicative of a new trend, and assume comparatively low initial improvements.

However the integrated approach would represent a significant change to the structure of the Model, which currently has a clear distinction between the two phases of determining initial improvements and then projecting them. The Committee felt that while the integrated approach is theoretically plausible, it may be seen as a step too far from the current method.

11. Predictive power

The Committee initially planned to select models based on pre-defined metrics of a given model's predictive power (amongst other notionally desirable qualities), which in turn were based on out-of-sample performance on past data. There is a large existing literature on this subject – see e.g. Gigerenzer and Brighton (2009) and Haldane (2012) for a discussion of a variety of approaches to the construction of predictive models.

After some investigation with inconclusive results, the Committee moved towards the view that:

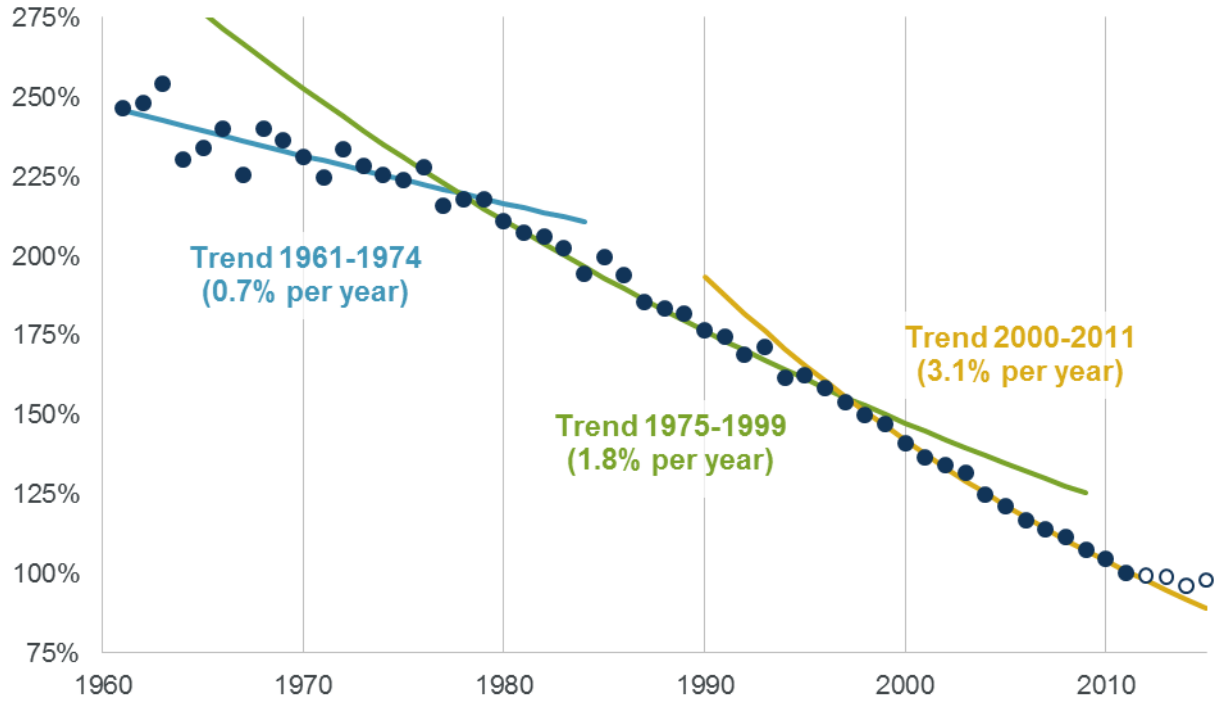
- selecting a longevity projections model purely on the basis of its past out-of-sample predictive performance is not robust given (a) the shortage of test data available and (b) the nature of longevity improvement; and
- instead, a simple, pragmatic approach to mortality projection was more appropriate.

There are strong reasons why selection criteria based purely on past predictions model are unlikely to be adequate:

- Mortality improvement varies significantly over time – by its very nature, if mortality is improving then the past drivers of improvements must fall away, and there is no natural law that requires the future (necessarily different) drivers to behave in the same way as the past drivers.
- Rates of mortality improvement (at least for England & Wales males) appear to persist over periods of a decade or even longer. Chart 11A shows this by plotting standardised mortality ratios (SMR) with three deliberately suggestive trend lines.
- There are at most five decades of relevant past data, i.e. at most five effective independent data points. There is very little one can do to test a model's predictive power if there only five independent data points (or even ten or twenty).
- Given the clear evidence of cohort effects in past England & Wales data, a projections model that does not project forward cohort-related features in the data is not considered acceptable by the Committee, However, the associated lack of identifiability of models with age, period and cohort parameters means that model calibration would need to make even stronger demands of the already small dataset.
- The inclusion in the model of a long-term improvement rate that is disconnected from mortality in the calibration period disables tests of predictive power because (a) the long-term rate is not derived objectively from past data within the model framework and (b) the tests of model prediction are strongly affected by the particular long-term rate chosen.

A simpler way of reaching the same conclusion is to note that a model that predicted England & Wales male mortality in Chart 11A would have to be one that – apart from the most recent four years – *always* predicted the future to have a higher rate of mortality improvement than the past, an approach we do not consider credible.

Chart 11A – Standardised mortality ratios (SMR), relative to 2011 for England & Wales males aged 50-89



The Committee was also cognisant of the school of thought that future mortality improvement should be at least partly determined using information other than solely the variation in recent mortality by age and time, such as:

- analysis of mortality patterns by cause of death; or
- the use of expert opinion to make explicit judgements about future longevity improvements.

Indeed the requirement of the Model for users to specify a long-term improvement rate, and the widespread use (for England & Wales mortality) of long-term improvement rates that are materially less than recent actual improvement rates, are tacit recognition of this.

Taking account of the fundamentally pragmatic rather than past-predictive nature of the model, the Committee has provided users with the ability to tweak the model's responsiveness to new data by adjusting the smoothing parameters S – and S_κ in particular – which in turn determine the model fit and projection.

12. Guide to software

The Committee is making software available as part of the consultation process. This section describes the software file “CMI Model consultation software v0.1.xlsm”. Further versions may be released in due course.

As noted in the disclaimer in the software:

> “This Software is being released as an illustrative version as part of the consultation process described in CMI Working Papers 90 and 91. It is being made available to allow users to investigate the CMI's proposals for the Model and to gain feedback on its functionality and accuracy. While the Software has been tested, the Software may contain bugs, errors and other problems. You should not rely on the Software without undertaking adequate checks of your own.”

The software is intended to allow interested parties to replicate the results in Working Paper 90 and this paper, and to consider the impact of particular parameter choices. It is not intended to offer the full functionality of the proposed model, and more complete software will be released to accompany CMI_2016 in March 2017. For example, the timing of calculated life expectancies is restricted to 1 January dates in this software, but will not be restricted in the final software.

The software contains Excel VBA macros. Depending on your computer setup you may need to explicitly enable macros in order to run the software.

12.1. Inputs

The sheet “Inputs” contains all of the inputs necessary to run the software. In its original state, all inputs are set to the proposed Core values.

The inputs are in three parts:

- Parameters needed for the APCI model
- Parameters for projection of mortality improvements
- Parameters for illustrative life expectancies

Once the inputs have been set, press the large “Run” button. Progress will be displayed in Excel’s statusbar. For the Core parameters a run takes of the order of 10-20 seconds on a typical PC. However we have found that convergence can be slower for some choices of parameters, including a calendar year range of 1995-2015.

12.2. Parameters needed for the APCI model

The parameters needed for the APCI model are described in Table 12.1.

Table 12.1: Parameters needed for the APCI model

Item	Core value	Notes
Dataset	ONS_EW_M or ONS_EW_F	The software will look for sheets with names starting with this, and ending in "_Exp" for exposures and "_Dth" for deaths. If you want to use different data, see Section 12.3 for details of the data format
Adjust exposures?	True	Set to True to apply the exposure adjustment described in Section 5.9 of Working Paper 90
If adjusting exposures, n	2	See Section 5.9 of Working Paper 90
If adjusting exposures, p	1%	See Section 5.9 of Working Paper 90
Minimum age to use for calibration	20	See Section 5.3 of Working Paper 90
Maximum age to use for calibration	100	See Section 5.3 of Working Paper 90
Age at which initial improvements taper to zero	110	See Section 8.6 of Working Paper 90
Minimum calendar year	1975	See Section 5.3 of Working Paper 90
Maximum calendar year	2015	See Section 5.3 of Working Paper 90
Number of cohorts to exclude	0	If a positive number is chosen, then age/year cells with fewer than that number of observations in the data will be excluded. This is not part of the proposed model but is included to allow sensitivity testing
Smoothing parameter S_α	7	See Section 7.4 of Working Paper 90
Smoothing parameter S_β	9	See Section 7.4 of Working Paper 90
Smoothing parameter S_κ	7.5	See Section 7.4 of Working Paper 90
Smoothing parameter S_γ	7	See Section 7.4 of Working Paper 90
Use direction of travel?	False	See Section 8.1 of Working Paper 90

12.3. Data format

The software requires data to be in a standard format that the Committee has used for its development and testing of the Model:

1. Corresponding exposures and deaths should be on separate sheets, with suffixes "_Exp" and "_Dth" and a common prefix.
2. Cells A1 to B16 contain "metadata", i.e. information about the data.
3. The name (cell B1) should match the sheet name.
4. Information on rows 2 to 10 provide optional information about the data.
5. The information in cells B11 to B16 should state the sex, type and age and year range of the data.
6. Row 18 contains header information for years, starting in column B. These should be consecutive ascending integers.
7. Column A contains header information for ages, starting in row 19. These should be consecutive ascending integers.
8. The data itself should start in cell B19.

12.4. Parameters for projection of mortality improvements

Columns J to M contain arrays by age of long-term rates and convergence periods. These all cover the age range 20 to 150.

The buttons above the columns will populate the columns with the Core assumptions. The button for the long-term rate for age-period improvements will require you enter a single assumption; this will be applied up to age 85, with the Core taper applied at older ages.

12.5. Parameters for illustrative life expectancies and annuities

The software will calculate illustrative life expectancies and annuities. Table 12.2 shows the parameters used to specify these.

Table 12.2: Parameters needed for illustrative life expectancies and annuities

Item	Notes
Base table qx	This specifies the q_x mortality rates to be used to calculate life expectancies and annuities
Base table timing	This specifies the date at which the "Base table qx" values apply
Use standard table	If you select one of the named standard tables from this list, then "Base table qx" and "Base table timing" will be populated for you. If the base table qx or timing is changed manually, then this will be reset to "{custom}"
Retirement age	This is used in the calculation of annuity values. For ages below retirement age, the annuity will be a deferred annuity
Interest rate	This is the (net) interest rate used to calculate annuities

A restriction of the current illustrative software is that calculated life expectancies and annuities will be as at 1 January of each year. The current Model does not have this restriction; neither will the published CMI_2016 Model.

Users should note that to change the software from males to females or vice versa they would need to change both the dataset used to calibrate the APCI model and the base table; rather than changing a single "gender" field. The software will produce a warning if it thinks there is an inconsistency between the genders of the dataset and the base table, but it does not prevent this (e.g. users may wish to see the impact of applying male mortality improvements to a female base table, for comparison).

12.6. Results workbook

The results of each run are written to a new Excel workbook. Its contents are shown in Table 12.3, and a contents sheet is also contained in the workbook itself.

12.7. "OutputChecks" sheet

The calculations in the software are all done using Excel Visual Basic for Applications (VBA) code. This has significant advantages for the speed and structure of the code, but may make the calculations less transparent to users who are unfamiliar with VBA.

To address this, the "OutputChecks" sheet shows detail of the projection of mortality improvements, including the conversion from m-style to q-style improvements, using Excel formulae. These can then be compared to the values calculated using VBA.

For various items of interest, the sheet:

1. calculates the item itself;
2. looks up the item from the relevant place in the output file; and
3. compares (1) and (2) to check that they agree.

For example, cells D38 to D42 convert from $\log m_{x,t}$ to $q_{x,t}$. Cell D38 looks up $\log m_{x,t}$ from the "logm_proj" sheet; cells D39:D40 convert this to $q_{x,t}$ using calculations on the "OutputChecks" sheet; cell D41 looks up $q_{x,t}$ from the "q_proj" sheet; and cell D42 compares the values of D40 and D41 and (hopefully) shows them to be the same.

Cells D8 and D9 can be varied by the user to investigate the derivation of an improvement at a particular age and year.

12.8. Passwords

In order to avoid accidental changes to results workbooks or to the software itself, the workbooks are protected with the password "CMI".

Table 12.3: Contents of the results workbook

Sheet name	Description
Contents	A contents sheet, containing information similar to this table.
Params_fit	The parameters used to specify the calibration process
Deaths	The deaths data used for calibration
Exposures	The exposures data used for calibration
ExposuresRaw	This sheet will only be shown if exposures have been adjusted by the software. If so, this will show the raw exposures data before adjustment; and adjusted cells on the Exposures and ExposuresRaw sheets will be highlighted
Params_APCI	The output parameters for the APCI model, including the derived components of mortality improvements, and direction of travel
Iterations	This shows the deviance, penalty function, and objective function after each iteration
logm_fit	Fitted values of the natural logarithm of central mortality rates
m_fit	Fitted values of the central mortality rates
DevRes	Deviance residuals and overall deviance
MI_fit	Fitted mortality improvements ("m-style")
Params_proj	The parameters used to specify the projection process
MI_m_proj	Projected mortality improvements ("m-style")
logm_proj	Projected values of the natural logarithm of central mortality rates
m_proj	Projected values of the central mortality rates
q_proj	Projected values of mortality rates (using calibration data) as at 1 January
MI_q_proj	Projected values of mortality improvements ("q-style")
Params_sample	The parameters used to specify the sample life expectancies
q_base_proj	Projected mortality rates (using the specified base table) as at 1 January
LE_P	Period life expectancies (using the specified base table) as at 1 January
LE_C	Cohort life expectancies (using the specified base table) as at 1 January
Annuity_P	Period annuities (using the specified assumptions) as at 1 January
Annuity_C	Cohort annuities (using the specified assumptions) as at 1 January
OutputChecks	This shows detail of the calculations. See Section 12.7 for details.

13. References

CMI documents

CMI Working Paper 3: "Projecting future mortality: a discussion paper" (2004)

CMI Working Paper 39: "A Prototype Mortality Projections Model: Part Two – Detailed Analysis" (2009)

CMI Working Paper 90: "CMI Mortality Projections Model consultation" (2016)

All CMI Working Papers can be accessed via the CMI website: https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/cmi-working-papers/numeric-listing-working-papers.

Please note that while this Working Paper is being made publically available, versions of the Model and some other outputs issued since March 2013 are restricted to those organisations and individuals who register as CMI users. Information on how to register is available on the CMI's website: http://www.actuaries.org.uk/research-and-resources/pages/how-access-cmi-outputs.

Non-CMI documents

Cairns AJG, Blake D and Dowd K (2006) "A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration". Journal of Risk and Insurance 73, 687-718.

Cairns AJG, Blake D, Dowd K, Coughlan GD, Epstein D, Ong A and Balevich I (2009) "A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States". North American Actuarial Journal, 13:1, 1-35.

Cairns AJG, Blake D, Dowd K, and Kessler A (2014) "Phantoms Never Die: Living with Unreliable Mortality Data". Working Paper.
http://www.macs.hw.ac.uk/~andrewc/papers/ajgc71.pdf

Cairns AJG, Blake D, Dowd K, and Kessler AR (2016) "Phantoms Never Die: Living with Unreliable Population Data". To appear in Journal of the Royal Statistical Society, Series A. DOI: 10.1111/rssa.12159

Currie ID (2006) "Smoothing and Forecasting Mortality Rates with P-splines".
http://www.macs.hw.ac.uk/~iain/research/talks/Mortality.pdf

Currie ID (2010) "A note on a model of Cairns, Blake and Dowd". Working Paper.

Currie ID (2016) "On fitting generalized linear and non-linear models of mortality", Scandinavian Actuarial Journal, 2016:4, 356-383.

Eilers PHC, Marx BD and Durban M (2015) "Twenty years of P-splines", SORT 39(2) 149-186

Gigerenzer G and Brighton H (2009) "Homo Heuristicus: Why Biased Minds Make Better Inferences", Topics in Cognitive Science 1, 107-143.

Haberman S and Renshaw A (2011) "A comparative study of parametric mortality projection models". Insurance: Mathematics and Economics, 48, 35-55.

Hunt A and Blake D (2015) "On the Structure and Classification of Mortality Models", Pensions Institute Discussion paper PI-1506.
http://www.pensions-institute.org/workingpapers/wp1506.pdf

Haldane AG (2012) "The dog and the Frisbee", Speech given at the Federal Reserve Bank of Kansas City's 36th economic policy symposium, "The Changing Policy Landscape", Jackson Hole, Wyoming.
http://www.bankofengland.co.uk/archive/Documents/historicpubs/speeches/2012/speech596.pdf

Lee RD and Carter LR (1992) “Modeling and Forecasting U.S. Mortality”. Journal of the American Statistical Association, 87:419, 659-671.

O'Hare C and Li Y (2011) “Explaining young mortality”. Insurance: Mathematics and Economics, 50, 12-25.

Plat R (2009) “On stochastic mortality modelling”, Insurance: Mathematics and Economics 45, 393-404.

Renshaw AE and Haberman S (2003) “Lee-Carter mortality forecasting with age-specific enhancement”. Insurance: Mathematics and Economics, 33 pp255-272.
http://www.sciencedirect.com/science/article/pii/S0167668703001380

Renshaw AE and Haberman S (2006) “A cohort-based extension to the Lee-Carter model for mortality reduction factors”. Insurance: Mathematics and Economics 38 pp556-570.
http://www.sciencedirect.com/science/article/pii/S0167668705001678

Willets RC (2004) “The Cohort Effect: Insights and Explanations”, British Actuarial Journal 10, 1027-1045.

Disclaimer: This Document has been prepared by and/or on behalf of Continuous Mortality Investigation Limited (CMI). The CMI does not accept any responsibility and/or liability whatsoever for the content or use of this Document. Whilst care has been taken during the development of the Document, CMI does not (i) warrant its accuracy; or (ii) guarantee any outcome or result from the application of this Document or of any of CMI's work (whether contained in or arising from the application of this Document or otherwise). You assume sole responsibility for your use of this Document, and for any and all conclusions drawn from its use. CMI hereby excludes all warranties, representations, conditions and all other terms of any kind whatsoever implied by statute or common law in relation to this Document, to the fullest extent permitted by applicable law. If you are in any doubt as to using anything produced by CMI, please seek independent advice.

Copyright: You may reproduce the contents of this Document free of charge in any format or medium provided it is:

1. reproduced accurately and is unaltered;
2. not used in a misleading context; and
3. correctly referenced and includes both CMI's Disclaimer notice set out above and CMI's copyright notice, as follows:

© Continuous Mortality Investigation Limited.

Continuous Mortality Investigation Limited ("CMI") is registered in England & Wales
Company number: 8373631
Registered Office: 7th floor, Holborn Gate, 326-330 High Holborn, London, WC1V 7PP

Correspondence address: Cheapside House, 138 Cheapside, London, EC2V 6BW
Email: info@cmilimited.co.uk
Tel: 020 7776 3820