Continuous Mortality Investigation

Institute and Faculty of Actuaries

Mortality Projections Committee

CMI_2018 v03 methods

Supplement to Working Paper 119

November 2019

Contents

Reliances and limitations

The purpose of the CMI Model is to allow users to produce projections of annual rates of mortality improvement. Specifically, the Model:

- is updated – typically annually – to reflect recent experience of mortality in the population of England & Wales, based on data published by the Office for National Statistics (ONS); and
- allows users the flexibility to modify projections tailored to their own views and purpose.

The CMI aims to produce high-quality outputs and takes considerable care to ensure that the Model and the accompanying documentation are accurate. However:

- We cannot guarantee their accuracy (see the Disclaimer on the last page of this document).
- There is a reliance on the underlying data, published by the ONS; although we have exercised judgement in the choice of age range and period and in the adjustments we apply to the data.
- We have also applied judgement and assumptions in:
  - the choice of the default parameters in the Core version of the Model; and
  - in deciding how to illustrate the results.
- Anyone using the Model should ensure that it is appropriate for their particular use and that suitable values are used for the parameters.

1. Introduction

The CMI Mortality Projections Committee publishes the CMI Mortality Projections Model ("the Model") and updates it to reflect the publication of new mortality data each year.

The method used for the latest version of the Model, CMI_2018 is similar to that used for the previous version, CMI_2017. The changes made are:

- An update to the dataset used to calibrate the Model, using the period 1978-2018 rather than 1977-2017.
- Making an adjustment to the exposure data at high ages, following the recommendations in Working Paper 106. This is described in Section 3.3.
- Changing the Core value of the period smoothing parameter, S_κ, from 7.5 to 7.

To accompany the release of CMI_2018, the Committee is publishing:

- A "results" paper, Working Paper 119, containing numerical results and discussion.
- A "methods" paper, containing technical details of the calculations, i.e. the formulae and algorithms.
- The CMI Model software, to allow users to perform their own calculations.
- A user guide to the software, describing how to use the software; i.e. how to specify parameter choices, and which buttons to press.
- Two spreadsheets that show how the CMI_2017 dataset has been derived from deaths and exposures data published by the Office for National Statistics (ONS).

These are all available from the CMI website to Authorised Users.

This is the "methods" paper. It is intended to be relatively concise and focus on what the Model does, rather than why the method has been adopted.

The Committee anticipates that it will publish a new results paper and software each year, to reflect changes in the Model due to new data. However, the methods paper should change little, if at all, from year to year; at least until the next review of the method.

1.1 Notation

Mathematical notation is explained when it is first introduced. All terms used are also collected in one place, in Section 7.

As the Model methods are not expected to change greatly over the next few years, it is helpful to illustrate them using algebraic equations (e.g. writing $Y_{\max}$ for the final year of calibration data) so that it is unambiguous how the method would vary for different calibration datasets. However we also illustrate the specific case of CMI_2018 (e.g. writing 2018 for the final year of calibration data).

2. High-level overview of methods used

In this section we provide a high-level overview of the methods used in the Model, before describing these in more detail in the following sections.

The Model is primarily a model of mortality improvements rather than mortality rates. Users of the Model can apply mortality improvements from the Model to their own choice of base mortality table to obtain projected mortality rates.

The basic approach of the Model is to project rates of mortality improvement by interpolating between current 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 summed to give the overall mortality improvements. The Model has a default ("Core") set of parameters, but users can also choose to use their own ("Extended" and/or "Advanced") parameters.

As a matter of policy, the Model is not fully specified – users must, as a minimum, select a value for the long-term rate, the intention being to encourage user engagement in determining future longevity improvement assumptions. Users do not need to consider the full calibration process if they are happy to use the Core assumptions for initial improvements.

The process is summarised below, with corresponding sections of this paper shown in square brackets:

- Construct a calibration dataset, consisting of deaths and exposure data for ages 20-100 and 41 calendar years (e.g. 1978-2018 for CMI_2018). This is based on Office for National Statistics (ONS) data for England & Wales, with some estimation and adjustment. [Section 3]
- Fit an APCI model of central mortality rates (m_x) to the calibration dataset. The fitting is based on maximum likelihood under a Poisson model for deaths. [Section 4]
- Derive mortality improvements from the APCI model – both total improvements and age-period and cohort components. [Sections 5.2 and 5.3]
- Project the age-period and cohort components of mortality improvements separately using the same form of convergence function to match desired long-term rates at and beyond the end of their convergence periods. [Section 5.4]
- Sum the age-period and cohort components to give the overall mortality improvement for each future year. [Section 5.4]
- So far, the mortality improvements use an "m-style" definition, introduced in CMI_2016. They are then converted to the "q-style" definition used in CMI_2009 to CMI_2015. [Sections 5.9]
- Calculate life expectancies and annuity values based on the projected mortality improvements and specified base table. [Section 6]

3. Construction of calibration dataset

We calibrate the APCI model, by using a rectangular dataset of deaths and exposures for ages X_{min} to X_{max} and years Y_{min} to Y_{max}. The data includes cohorts C_{min} to C_{max}, where $C_{min} = Y_{min} - X_{max}$ and $C_{max} = Y_{max} - X_{min}$.

For CMI_2018 the calibration data is for ages 20-100, years 1978-2018, and cohorts 1878-1998.

We anticipate that the calibration data for subsequent versions of the Model:

- will use the same age range; and
- will use a calendar year range from $Y_{max} - 40$ to Y_{max}, where Y_{max} is the year in the version name.

For example, we expect that CMI_2019 will be calibrated to ages 20-100, years 1979-2019, and cohorts 1879-1999.

The data used is for England & Wales and is obtained from the Office for National Statistics (ONS).

We expect that we will be able to obtain data for deaths and exposures for years Y_{min} to $Y_{max} - 1$ from the ONS in the required format, but we will need to make our own estimates for the final year of data, Y_{max}. This is the case for CMI_2018 where we need to estimate exposures for 2018, and estimate the distribution of deaths by age in 2018.

The CMI and others have noted that ONS population estimates have some unusual features for certain years of birth. To address this, we make two adjustments:

- a "high-age" adjustment, described in Section 3.3, that affects all exposures at ages 85 and above; and
- an "all-age" adjustment, described in Section 3.6, that affects only those age/year combinations where the data appears anomalous.

The remainder of this section describes the data sources, and the methods we use. We consider in turn:

- Raw exposures for years Y_{min} to $Y_{max} - 1$ (e.g. 1978-2017)
- Deaths for years Y_{min} to $Y_{max} - 1$ (e.g. 1978-2017)
- The "high-age" adjustments made to raw exposures
- Estimated exposures for year Y_{max} (e.g. 2018)
- Estimated deaths for year Y_{max} (e.g. 2018)
- The "all-age" adjustments made to exposures
- Data for earlier years (e.g. 1961-1977) that is not used in the calibration of CMI_2018, but which may be of interest.

The order of presentation is chosen so that no section depends on a later section; e.g. our estimate of exposures for Y_{max} depends on deaths and exposures in $Y_{max} - 1$; and our estimate of deaths in Y_{max} depends on exposures in Y_{max}.

In the sections below, deaths for age x in year t are denoted $D_{x,t}$, and exposures are denoted $E_{x,t}$.

3.1 Raw exposures for years Y_{min} to $Y_{max} - 1$ (e.g. 1978-2017)

The exposure data is approximated by mid-year population estimates for England & Wales.

The data for CMI_2018 is taken from a consolidated data file provided by the ONS following our request[1].

[1] Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/adhocs/009189populationestimatesanddeathsbysingleyearofageforenglandandwalesandtheuk1961to2017

3.2 Deaths for years Y_{min} to $Y_{max} - 1$ (e.g. 1978-2017)

The data used is for deaths registered in England & Wales, and is taken from the same consolidated ONS data file as for raw exposures.

3.3 High-age exposure adjustment

The exposure data for years Y_{min} to $Y_{max} - 1$ (e.g. 1978-2017) at high ages is adjusted using the method of Working Paper 106. This is done before estimating deaths and exposures for year Y_{max} (e.g. 2018).

We make the adjustments in Steps 1 to 6 of Section 3.1 of Working Paper 106. We exclude Step 7 at this stage, as it the same as the all-age adjustment in Section 3.6 of this paper, and is done after data for 2018 has been estimated.

3.4 Raw exposures for year Y_{max} (e.g. 2018)

A population estimate for England & Wales in the year Y_{max} is not available in time for the production of the Model, so we need to estimate it.

To do this, we first make an approximate estimate that projects the previous year's exposure, allowing only for the previous year's deaths. We then refine this adjustment, by considering how good the approximate estimate would have been in previous years.

The initial approximate estimate $E^0_{x,t}$ is defined by:

$$E^0_{x,t} = E_{x-1,t-1} - D_{x-1,t-1}$$

This is approximate for (at least) two reasons:

- Deaths in year $t-1$ are effectively used as a proxy for deaths in the second half of year $t-1$ and the first half of year t.
- No allowance is made for changes in population due to net migration.

The approximations are addressed by making adjustments. We define an adjustment $A_{G,t}$ for age group G (for ages 15-44, 45-64, 65-74, 75-84 and 85+ for consistency with the ONS) that compares actual ONS exposure data with the initial estimate:

$$A_{G,t} = \frac{\sum_{x \in G} E_{x,t}}{\sum_{x \in G} E^0_{x,t}}$$

This is calculated for the six[2] years prior to Y_{max} and we then take an average:

$$\bar{A}_G = \frac{1}{6}\left(A_{G,Y_{max}-1} + \cdots + A_{G,Y_{max}-6}\right)$$

and apply this to the exposure estimate for Y_{max}, so that:

$$E_{x,Y_{max}} = E^0_{x,Y_{max}} \times \bar{A}_{G_x}$$

where G_x is the age group that contains age x. We expect that the refined estimate $E_{x,Y_{max}}$ will be more reliable than the initial estimate $E^0_{x,Y_{max}}$ as $A_{G,t}$ is relatively stable over time.

[2] The choice of six years is intended to strike a balance between having sufficient years to reduce the impact of any single unusual year, and few enough years to reflect recent experience.

3.5 Deaths for year Y_{max} (e.g. 2018)

Deaths data by single year of age for the year Y_{max} is not available in time for the production of the Model, so we need to estimate it. For the year Y_{max} we base the data on weekly provisional figures by gender and age group[3].

We first calculate the total numbers of deaths for the year, for each age group, by summing the numbers of deaths for each week. We give partial weight to those weeks that span two calendar years, based on the number of working days that fall in each year (i.e. excluding weekends and public holidays), since this broadly reflects days on which register offices would be open for the registration of deaths.

For example, in CMI_2018 the number of deaths in 2018 for a particular gender and age band is the sum of:

- All of the deaths in the week ending on 5 January 2018 (since that week has four working days, January 2 to 5 inclusive, all of which are in 2018).
- All of the deaths in the weeks ending on 12 January 2018 to 28 December 2018 inclusive (since those weeks are contained entirely within 2018).
- One quarter of the deaths in the week ending on 4 January 2019 (since one day, December 31, of the four working days in that week falls in 2018).

At this stage the data is grouped by age band but we need to split this into single years of age for use in the APCI model.

For each age group G (15-44, 45-64, 65-74, 75-84 and 85+, as in Section 3.4) we calculate:

$$Scale_G = D_{G,Y_{max}}^{Grouped} \div \left(\sum_{x \in G} E_{x,Y_{max}} \times m_x^{NLT}\right)$$

Where $D_{G,Y_{max}}^{Grouped}$ is the number of deaths in the age group, $E_{x,Y_{max}}$ is the estimated exposure at each age in Y_{max}; and m_x^{NLT} is from the latest National Life Table for England & Wales[4]. For CMI_2018 this is the 2015-2017 table.

The ONS provides a single combined exposure for ages 105 and above, rather than exposures for each single year of age. In the calculation of $Scale$ for the highest age group, we use m_{105}^{NLT} for ages 105 and above, to avoid the spurious complexity of splitting that exposure into single years of age.

We expect that $Scale_G$ will be fairly close to 1 for all age-groups but it will differ slightly due to recent changes in mortality rates.

We then estimate the number of deaths at each age $D_{x,Y_{max}}$ by:

$$D_{x,Y_{max}} = E_{x,Y_{max}} \times m_x^{NLT} \times Scale_{G_x}$$

where G_x is the age group containing age x.

Our approach means that the pattern of deaths by age is plausible, based on estimated exposure and recent mortality rates, and the total number of deaths in each age group agrees with the ONS total; i.e.:

$$\sum_{x \in G} D_{x,Y_{max}} = D_{G,Y_{max}}^{Grouped}$$

[3] Available from http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales

[4] The National Life Table only runs to age 100 so we extrapolate exponentially beyond this age. Our method is described in Section 6.4.

3.6 All-age exposure adjustment

At this stage we have data for registered deaths $D_{x,t}$ and exposure estimates $E_{x,t}$ for the required age and calendar year range, based on ONS data. However, as noted above, there are some concerns about the data, and 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.

Our method, and the background and motivation behind it, is described in more detail in Section 2 of Working Paper 91.

We assume that the smoothed mortality rates $m_{x,T}$ in the age range $[X-n, X+n]$ in year T increase exponentially with age. Fitting an exponential model over that age range using least squares regression gives an estimate of the smoothed mortality rate $m_{x,T}$ for the specific point that we are considering:

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

If $D_{x,T} = 0$ for any age in the range $[X-n, X+n]$ then $\log m_{x,T}$ cannot be estimated in this way and no exposure adjustment will be made. This does not affect CMI_2018 and is unlikely to affect future Models, but may arise in other datasets, particularly for small populations.

We then calculate the deviance residual $r_{X,T}$ 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 rates increase exponentially 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 the data. Our previous research, in Appendix D of Working Paper 84, suggests that the problem is likely to be with the exposure data rather than the deaths data.

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

- if $|r_{X,T}| \le \Phi^{-1}\left(1-\frac{p}{2}\right)$ we use the unadjusted exposure $E_{X,T}$.
- However if $|r_{X,T}| > \Phi^{-1}\left(1-\frac{p}{2}\right)$ we replace the exposure by $\frac{D_{X,T}}{m_{X,T}}$.

For the Core model we set $p = 1\%$ and we use $n = 2$ for most ages. However for ages at and near the edges of the data we need to use a lower value of n, e.g. for ages $X_{min}+1$ and $X_{max}-1$ we use $n = 1$, and for ages X_{min} and X_{max} we make no adjustment.

3.7 Data for earlier years

The Model is calibrated to 41 years of data. The software also contains data for earlier years, which may be helpful to users in testing the Model's sensitivity to different time periods. However, the Committee cautions that it made a conscious decision not to calibrate the Model to a longer period of data, due to the concerns raised in Appendix D of Working Paper 74 about the quality of data in earlier years.

In the CMI_2018 software, data is provided from 1961 onwards and data for 1961-1977 inclusive is taken from the file in footnote 1 in this section. The exposure data for these earlier years has also been adjusted in line with Steps 1 to 6 of Section 3.1 of Working Paper 106, as described in 3.3 above.

4. Initial rates of mortality improvement

We use the Age-Period-Cohort Improvement (APCI) model to calculate initial mortality improvements for the Core Model. The APCI model is defined by:

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

where:

$m_{x,t}$	is the fitted central mortality rate
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 1978 to 2018, then $\bar{t}$ is 1998
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 terms by age relating to mortality improvements
κ_t	are parameter values for terms by period (i.e. calendar year)
γ_c	are parameter values for terms by cohort (i.e. birth year)

This section describes the process of fitting the APCI model. Working Papers 90 and 91 provide further detail about the APCI model, including the motivation for its design, and results and sensitivities.

4.1 Constraints on γ_c parameters

When fitting the APCI model, there are no constraints placed on the values of α_x, β_x or κ_t. However, as described in Section 3 of Working Paper 93, the values of γ_c are constrained, in order to restrict the cohort components of mortality improvements at the youngest and oldest ages. Without these constraints, the cohort components at those ages are large, and the Committee is concerned that these may be an artefact of the model or the data rather than reflecting genuine features of mortality improvement.

We constrain cohort components of initial mortality improvements to be nil at ages X^C_{Low} and below, and at ages X^C_{High} and above; i.e. cohort components of initial mortality improvements are nil for cohorts $Y_{max} - X^C_{Low}$ and later, and for cohorts $Y_{max} - X^C_{High}$ and earlier.

In order to achieve this, we fit values of γ_c for cohorts C^*_{min} to C^*_{max} where:

$$C^*_{min} = Y_{max} - X^C_{High}$$
$$C^*_{max} = Y_{max} - X^C_{Low} - 1$$

Cohort parameters for $c < C^*_{min}$ are set equal to $\gamma_{C^*_{min}}$; and those for $c > C^*_{max}$ are set equal to $\gamma_{C^*_{max}}$.

For CMI_2018, the Core values of X^C_{Low} and X^C_{High} are 30 and 110 respectively. This means that we fit cohort parameters γ_c directly for cohorts 1908 to 1987 inclusive. Cohort parameters for 1907 and earlier are set equal to γ_{1908}, and those for 1988 and later are set equal to γ_{1987}.

4.2 The objective function

For the purpose of the Model, we want to extract the underlying trends in mortality improvement 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:

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

Penalties for α_x and β_x

The penalty for the α_x terms is given by sums of squares of third-order differences:

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

where λ_α is a hyperparameter that can be used to control the smoothness of the α_x parameters.

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:

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

Note that $\underline{D_\alpha}$ has size $(N_X - 3) \times N_X$ where N_X is the number of ages; so $N_X = X_{max} + 1 - X_{min}$.

The penalty for the β_x terms is similar, with:

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

$$\text{Penalty}(\beta_x) = \lambda_\beta \underline{\beta}^T \underline{D_\beta}^T \underline{D_\beta}\,\underline{\beta}$$

and $\underline{D_\beta} = \underline{D_\alpha}$.

Penalty for κ_t

The penalty for κ_t is of second, rather than third order:

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

$$\text{Penalty}(\kappa_t) = \lambda_\kappa \underline{\kappa}^T \underline{D_\kappa}^T \underline{D_\kappa}\,\underline{\kappa}$$

where $\underline{D_\kappa}$ is of size $(N_Y - 2) \times N_Y$:

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

Penalty for γ_c

The penalty for γ_c differs, due to the constraints placed on γ_c described in Section 4.1. It is:

$$\text{Penalty}(\gamma_c) = \lambda_\gamma \underline{\gamma}^T \underline{D_\gamma}^T \underline{D_\gamma} \underline{\gamma}$$

Where $\underline{D_\gamma}$ is of size $(N_C^* + 1) \times N_C^*$, where N_C^* is the number of directly-fitted cohorts, so $N_C^* = C_{max}^* + 1 - C_{min}^*$.

$$\underline{D_\gamma} = \begin{bmatrix} +1 & -1 & 0 & & & & & \\ -2 & +3 & -1 & 0 & & & & \\ +1 & -3 & +3 & -1 & 0 & & & \\ 0 & +1 & -3 & +3 & -1 & 0 & & \\ & 0 & +1 & -3 & +3 & -1 & 0 & \\ & & & \ddots & \ddots & \ddots & \ddots & \\ & & & 0 & +1 & -3 & +3 & -1 \\ & & & & 0 & +1 & -3 & +2 \\ & & & & & 0 & +1 & -1 \end{bmatrix}$$

Use of S instead of λ

The penalty functions above use hyperparameters λ_α, λ_β, λ_κ and λ_γ to control the smoothness of the parameters. In practice these values are large and it is more convenient to work with their logarithms.

We define

$$S_\alpha = \log_{10} \lambda_\alpha; \quad S_\beta = \log_{10} \lambda_\beta; \quad S_\kappa = \log_{10} \lambda_\kappa; \quad S_\gamma = \log_{10} \lambda_\gamma$$

and

$$\lambda_\alpha = 10^{S_\alpha}; \quad \lambda_\beta = 10^{S_\beta}; \quad \lambda_\kappa = 10^{S_\kappa}; \quad \lambda_\gamma = 10^{S_\gamma}$$

4.3 Newton's method

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.

General case

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

Given a set of parameters $(\phi_1^{(i)}, \dots \phi_n^{(i)})$, if we calculate

$$\begin{bmatrix} \phi_1^{(i+1)} \\ \vdots \\ \phi_n^{(i+1)} \end{bmatrix} = \begin{bmatrix} \phi_1^{(i)} \\ \vdots \\ \phi_n^{(i)} \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 f}{\partial \phi_1 \partial \phi_1}(\phi_1^{(i)}, \dots \phi_n^{(i)}) & \dots & \frac{\partial^2 f}{\partial \phi_1 \partial \phi_n}(\phi_1^{(i)}, \dots \phi_n^{(i)}) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial \phi_n \partial \phi_1}(\phi_1^{(i)}, \dots \phi_n^{(i)}) & \dots & \frac{\partial^2 f}{\partial \phi_n \partial \phi_n}(\phi_1^{(i)}, \dots \phi_n^{(i)}) \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f}{\partial \phi_1}(\phi_1^{(i)}, \dots \phi_n^{(i)}) \\ \vdots \\ \frac{\partial f}{\partial \phi_n}(\phi_1^{(i)}, \dots \phi_n^{(i)}) \end{bmatrix}$$

then we expect that $(\phi_1^{(i+1)}, \dots \phi_n^{(i+1)})$ will give a lower value for f. Repeating this iterative step should converge towards the minimum of f as i increases.

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:

1. Initialise the procedure:
   - 1a. Initialise all parameters: $\alpha_x^{(0)}$, $\beta_x^{(0)}$, $\kappa_t^{(0)}$ and $\gamma_c^{(0)}$.
   - 1b. Calculate mortality rates based on the initial parameters
   - 1c. Calculate the objective function
2. Do repeatedly, until the objective function stabilises:
   - 2a. Update all of the α parameters from $\alpha_x^{(i)}$ to $\alpha_x^{(i+1)}$ using Newton's method
   - 2b. Calculate updated mortality rates
   - 2c. Update all of the β parameters from $\beta_x^{(i)}$ to $\beta_x^{(i+1)}$ using Newton's method
   - 2d. Calculate updated mortality rates
   - 2e. Update all of the κ parameters from $\kappa_t^{(i)}$ to $\kappa_t^{(i+1)}$ using Newton's method
   - 2f. Calculate updated mortality rates
   - 2g. Update the γ parameters where $C_{min}^* \le c \le C_{max}^*$ from $\gamma_c^{(i)}$ to $\gamma_c^{(i+1)}$ using Newton's method
   - 2h. Update the γ parameters where $c < C_{min}^*$ to be equal to $\gamma_{C_{min}^*}$
   - 2i. Update the γ parameters where $c > C_{max}^*$ to be equal to $\gamma_{C_{max}^*}$
   - 2j. Calculate updated mortality rates
   - 2k. Update parameters to allow for identifiability
   - 2l. Update the γ parameters where $c < C_{min}^*$ to be equal to $\gamma_{C_{min}^*}$
   - 2m. Update the γ parameters where $c > C_{max}^*$ to be equal to $\gamma_{C_{max}^*}$.
   - 2n. Calculate updated mortality rates
   - 2o. Calculate the objective function
3. Calculate mortality improvements

4.4 Calculating partial derivatives

We 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.

To implement Newton's method for the APCI model we need to be able to calculate the partial derivatives:

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

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 – first order

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:

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

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

For the β and κ parameters we have similarly:

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

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

For the γ parameters the position is slightly more complicated since we are constraining some of the parameters and are only directly fitting γ_c where $C^*_{min} \le c \le C^*_{max}$. We do:

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

The "$Include$" criterion is true if and only if one of the following holds:

- $t - x < C^*_{min}$ and $i = C^*_{min}$
- $C^*_{min} \le t - x \le C^*_{max}$ and $i = t - x$
- $C^*_{max} < t - x$ and $i = C^*_{max}$

Effectively we are considering the value of $\left(E_{x,t}m_{x,t} - D_{x,t}\right)$ over all age/year cells and assigning it either to $t - x$ if $C^*_{min} \le t - x \le C^*_{max}$ or to C^*_{min} or C^*_{max} otherwise.

Deviance terms – second order

Turning to the second-order derivatives:

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

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:

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

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:

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

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

$$\frac{\partial^2 Deviance}{\partial{\gamma_i}^2} = 2\sum_{x,t|Include(i,x,t)} E_{x,t}m_{x,t}$$ where the "$Include$" criterion is as before.

Penalty terms

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

Recall that the penalty function can be written 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 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.

4.5 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 $\theta_1, \ldots, \theta_5$:

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

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 2k of the pseudocode in Section 4.3).

The steps are:

(a) least-squares 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) least-squares 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 the error function for least-squares quadratic regression:

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

If we choose parameters θ_1, θ_2 and θ_3 to minimise LSE 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:

$$\frac{\partial LSE}{\partial \theta_1} = \frac{\partial LSE}{\partial \theta_2} = \frac{\partial LSE}{\partial \theta_3} = 0$$

We have:

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

$$\frac{\partial LSE}{\partial \theta_2} = 2\sum_c \left(\theta_1(c - \bar{c}) + \theta_2(c - \bar{c})^2 + \theta_3(c - \bar{c})^3 - \gamma_c(c - \bar{c})\right)^2$$

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

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.

5. Projection

This section describes the projection of mortality improvements and mortality rates.

5.1 Definition of mortality improvements

There are two definitions of mortality improvements used within the Model.

"q-style" improvements, consistent with CMI_2015 and earlier, are defined by:

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

"m-style" improvements, introduced in Working Paper 90 and used in CMI_2016 and later, are defined by:

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

The new Model uses m-style improvements for initial improvements, the long-term rate, and projected improvements, and these are then converted to q-style improvements as the final outputs; for consistency with previous versions of the Model.

5.2 Derivation of initial improvements from the APCI model

Initial age-period and cohort components of mortality improvements $MI^{AP(0)*}_{x,Y_{max}}$ and $MI^{C*}_{Y_{max}-x,Y_{max}}$ are calculated by:

$$MI^{AP(0)*}_{x,Y_{max}} = -\beta_x + \kappa_{Y_{max}-1} - \kappa_{Y_{max}} \quad \text{for ages } X_{min} \text{ to } X_{max}$$

$$MI^{AP(0)*}_{x,Y_{max}} = \left(\frac{X_{taper}-x}{X_{taper}-X_{max}}\right) MI^{AP(0)*}_{X_{max},Y_{max}} \quad \text{for ages } X_{max}+1 \text{ to } X_{taper}-1$$

$$MI^{AP(0)*}_{x,Y_{max}} = 0 \quad \text{for ages } X_{taper} \text{ and above}$$

$$MI^{C*}_{Y_{max}-x,Y_{max}} = \gamma_{Y_{max}-1-x} - \gamma_{Y_{max}-x} \quad \text{for ages } X_{min} \text{ to } X_{max}$$

$$MI^{C*}_{Y_{max}-x,Y_{max}} = \left(\frac{X_{taper}-x}{X_{taper}-X_{max}}\right) MI^{C*}_{Y_{max}-X_{max},Y_{max}} \quad \text{for ages } X_{max}+1 \text{ to } X_{taper}-1$$

$$MI^{C*}_{Y_{max}-x,Y_{max}} = 0 \quad \text{for ages } X_{taper} \text{ and above}$$

e.g. for CMI_2018 these are:

$$MI^{AP(0)*}_{x,2018} = -\beta_x + \kappa_{2017} - \kappa_{2018} \quad \text{for ages 20 to 100}$$

$$MI^{AP(0)*}_{x,2018} = \left(\frac{110-x}{110-100}\right) MI^{AP(0)*}_{100,2018} \quad \text{for ages 101 to 109}$$

$$MI^{AP(0)*}_{x,2018} = 0 \quad \text{for ages 110 to 150}$$

$$MI^{C*}_{2018-x,2018} = \gamma_{2017-x} - \gamma_{2018-x} \quad \text{for ages 20 to 100}$$

$$MI^{C*}_{2018-x,2018} = \left(\frac{110-x}{110-100}\right) MI^{C*}_{2018-100,2018} \quad \text{for ages 101 to 109}$$

$$MI^{C*}_{2018-x,2018} = 0 \quad \text{for ages 110 to 150}$$

5.3 Initial addition to mortality improvements

The "initial addition to mortality improvements", A, is used to adjust mortality improvements for years $t \leq Y_{max}$.

This can be specified either as a single value or as an array by age. If a single value A is specified, then this will be used to determine the initial additions at each age, using the same shape as the Core assumption for the shape of the long-term rate. i.e. in this case the initial additions $MI_x^{InitAdd*}$ are:

- For ages 85 and below A
- For ages 86 to 109 $A \times (110 - x) \div 25$
- For ages 110 and above 0

If an array by age for A is specified, no additional modifications (such as tapering at high ages) are made to the array. The initial addition is added to the age-period component of initial mortality improvements from the APCI model:

$$MI_{x,Y_{max}}^{AP*} = MI_{x,Y_{max}}^{AP(0)*} + MI_x^{InitAdd*}$$

Section 5.6 describes how the initial addition is applied to historical mortality improvements.

5.4 Projection of m-style mortality improvements

The Model projects age-period improvements as:

$$MI_{x,Y_{max}+t}^{AP*} = L_x^{AP} + \left(MI_{x,Y_{max}}^{AP*} - L_x^{AP}\right)\left(1 - 3\left(\frac{t}{T_x^{AP}}\right)^2 + 2\left(\frac{t}{T_x^{AP}}\right)^3\right) + D_x^{AP} t \left(1 - \frac{t}{T_x^{AP}}\right)^2 \quad \text{for } 0 \leq t \leq T_x^{AP}$$

$$MI_{x,Y_{max}+t}^{AP*} = L_x^{AP} \quad \text{for } t > T_x^{AP}$$

where:

$MI_{x,Y_{max}}^{AP*}$ is the initial rate of mortality improvement for age x

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

T_x^{AP} is the convergence period for age x, which must be greater than zero

D_x^{AP} is the direction of travel for age x

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

$$D_x^{AP} = \frac{1}{T_x^{AP}}\left(8Ppn_x^{AP} - 4\right)\left(MI_{x,Y_{max}}^{AP*} - L_x^{AP}\right)$$

The Model projects cohort improvements as:

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

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

where:

$c = Y_{max} - x$

$MI_{c,Y_{max}}^{C*}$ is the initial rate of mortality improvement for cohort c

L_c^C is the long-term rate for cohort c

T_c^C is the convergence period for cohort c, which must be greater than zero

D_c^C is the direction of travel for cohort c

If the shape of convergence is specified in terms of the proportion remaining at midpoint (Ppn_x^C) then:

$$D_x^C = \frac{1}{T_x^C}(8Ppn_x^C - 4)\left(MI_{x,Y_{max}}^{C*} - L_x^C\right)$$

The Model projects total improvements by adding age-period and cohort components and, for future years, the constant addition to mortality improvements; i.e.

$$MI^{*}_{x,t} = MI^{AP*}_{x,t} + MI^{C*}_{t-x,t} \quad \text{for } t = Y_{max}$$
$$MI^{*}_{x,t} = MI^{AP*}_{x,t} + MI^{C*}_{t-x,t} + MI^{ConstAdd*}_{x} \quad \text{for } t > Y_{max}$$

where:

$MI^{ConstAdd*}_{x}$ is the constant addition to mortality improvements, which does not vary by year.

5.5 CMI_2018 Core projection parameters

This section describes the parameters used for projecting mortality improvements in the Core CMI_2018 Model:

- The initial rates $MI^{AP*}_{x,Y_{max}}$ and $MI^{C*}_{c,Y_{max}}$ are derived from the APCI model, as described in Sections 5.2 and 5.3.
- The long-term rate L^{AP}_{x} is not specified. At a minimum the user must specify a single value, LTR. The rates L^{AP}_{x} are then:
  - For ages 85 and below LTR
  - For ages 86 to 109 $LTR \times (110 - x) \div 25$
  - For ages 110 and above 0
- The long-term rate L^{C}_{c} is nil for all ages.
- The convergence periods T^{AP}_{x} and T^{C}_{c} are shown in Table 5.1
- The directions of travel D^{AP}_{x} and D^{C}_{c} are nil for all ages and cohorts
- The initial addition, $MI^{InitAdd*}_{x}$, and constant addition, $MI^{ConstAdd*}_{x}$, to mortality improvements are both nil for all ages.

Table 5.1: Age-period and cohort convergence periods, based on age in year Y_{max}

Age (x) in Y_{max}	Age-period (T^{AP}_{x})	Cohort (T^{C}_{c})
20-49	10	x – 10
50-60	x – 40	40
61-79	20	100 − x
80-94	100 − x	100 − x
95-105	5	5
106-109	5	110 − x
110 and older	5	0

5.6 Calculation of historical m-style mortality improvements

We calibrate the APCI model to data for ages X_{min} to X_{max} (e.g. 20-100) and years Y_{min} to Y_{max} (e.g. 1978-2018). This means that mortality improvements are available for years $Y_{min}+1$ to Y_{max} (e.g. 1979-2018). We have described the calculation of initial improvements, in year Y_{max}, in Sections 5.2 and 5.3. For historical improvements, for years $Y_{min}+1$ to $Y_{max}-1$:

- For ages X_{min} to X_{max} (e.g. 20-100), historical mortality improvements are first calculated from the fitted APCI model as $MI^{Hist*}_{x,t} = \log m_{x,t-1} - \log m_{x,t}$.
- Improvements are nil at ages X_{taper} and above (e.g. ages 110 and above).
- For ages $X_{max}+1$ to $X_{taper}-1$ (e.g. 101 to 109), improvements are interpolated between $MI^{Hist*}_{X_{max},t}$ at age X_{max} and nil at age X_{taper} (e.g. between 100 and 110).

The initial addition is then applied to all historical years so that:

$$MI^{*}_{x,t} = MI^{Hist*}_{x,t} + MI^{InitAdd*}_{x} \qquad \text{for } Y_{min}+1 \le t \le Y_{max}-1$$

5.7 Calculation of mortality rates

To calculate $\log m_{x,t}$:

- For ages X_{min} to X_{max} in Y_{max}, taken directly from the fit of the APCI model
- For ages $X_{max}+1$ and older in Y_{max}, linear extrapolation based on $\log m_{X_{max}-1,Y_{max}}$ and $\log m_{X_{max},Y_{max}}$ i.e. $\log m_{x,Y_{max}} = \log m_{X_{max},Y_{max}} + (x - X_{max})(\log m_{X_{max},Y_{max}} - \log m_{X_{max}-1,Y_{max}})$
- For ages X_{min} to X_{max}, years $Y_{min}+1$ to $Y_{max}-1$, determined using $\log m_{x,t} = \log m_{x,t+1} + MI^{*}_{x,t+1}$
- For ages X_{min} to X_{max}, years $Y_{max}+1$ onwards, determined using $\log m_{x,t} = \log m_{x,t-1} - MI^{*}_{x,t}$

For CMI_2018:

- For ages 20-100 in 2018, taken directly from the fit of the APCI model
- For ages 101 and older in 2018, linear extrapolation based on $\log m_{99,2018}$ and $\log m_{100,2018}$ i.e. $\log m_{x,2018} = \log m_{100,2018} + (x - 100)(\log m_{100,2018} - \log m_{99,2018})$
- For ages 20-100, years 1979-2017, determined using $\log m_{x,t} = \log m_{x,t+1} + MI^{*}_{x,t+1}$
- For ages 20-100, years 2019 onwards, determined using $\log m_{x,t} = \log m_{x,t-1} - MI^{*}_{x,t}$

5.8 Derivation of direction of travel

The period component of mortality improvements is given by:

$$MI^{P*}_{t} = \kappa_{t-1} - \kappa_{t}$$

Direction of travel is defined as the change in this, so:

$$DoT_t = MI^{P*}_{t} - MI^{P*}_{t-1}$$
$$DoT_t = -\kappa_t + 2\kappa_{t-1} - \kappa_{t-2}$$

And the initial direction of travel is:

$$DoT_{Y_{max}} = -\kappa_{Y_{max}} + 2\kappa_{Y_{max}-1} - \kappa_{Y_{max}-2}$$

5.9 Calculation of q-style mortality rates and improvements

The Model converts from central to initial mortality rates by assuming that:

$$q_{x,t} = 1 - \exp(-m_{x,t})$$

then calculates:

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

6. Illustrative life expectancies and annuities

The Model produces life expectancies using an illustrative base table; and annuities using the same base table and a specified retirement age and interest rate.

Calculations in previous sections all use annual data, and the mortality rates $q_{x,t}$ in Section 5.9 are as at 1 January each year. However we want to calculate values at any calculation date and for a base table with any effective date.

6.1 Calculation of mortality rates

Define a reduction factor from $DateA$ to $DateB$ by:

$$RF(x, DateA, DateB) = q(x, DateB) \div q(x, DateA)$$

where the mortality rates are, as in Section 5.9, consistent with the calibration data.

We then use the reduction factors to project mortality rates from the illustrative base table at any date d as:

$$q_{Illustrative}(x, d) = \min(q_{Illustrative}(x, BaseTableDate) \times RF(x, BaseTableDate, d), 1)$$

Note that values of $q_{Illustrative}$ are limited to 1.

6.2 Calculation of reduction factors

The calculation of the reduction factors depends on the timing of $DateA$ and $DateB$

Case 1 – both at 1 January

If $DateA$ and $DateB$ are both on 1 January for some year, then the reduction factor can be calculated directly from the rates in Section 5.9.

Case 2 – both between consecutive 1 January dates

If $DateA$ and $DateB$ are both contained in the period "1 January Y" to "1 January $(Y + 1)$" (inclusive) for some year Y, then the reduction factor is defined as:

$$RF(x, DateA, DateB) = RF(x, 1 \text{ January } Y, 1 \text{ January } (Y + 1))^{p(DateA, DateB, Y)}$$

where p is the proportion of the year between those dates, defined by:

$$p(DateA, DateB, Y) = DayInterval(DateA, DateB) \div DayInterval(1 \text{ January } Y, 1 \text{ January } (Y + 1))$$

where $DayInterval(Date1, Date2)$ is the number of days in the interval from $Date1$ to $Date2$.

Some examples of $DayInterval$ are shown in Table 6.1. It is negative if $Date1$ is after $Date2$; and can be implemented in Excel by subtracting $Date1$ from $Date2$.

Table 6.1: Examples of $DayInterval$

$Date1$	$Date2$	$DayInterval(Date1, Date2)$
1 January 2016	11 January 2016	+10
1 April 2016	10 July 2016	+100
10 July 2016	1 April 2016	−100
1 January 2016	1 January 2017	+366
1 January 2017	1 January 2018	+365

Note that if $DateA$ and $DateB$ are consecutive 1 January dates, then Case 2 is consistent with Case 1.

Case 3 – Multiple years

If neither Case 1 nor Case 2 applies, then $DateA$ and $DateB$ span multiple years. In this case:

- Split the period from $DateA$ to $DateB$ into a number of sub-periods, each starting and/or ending on 1 January.
- Calculate a reduction factor for each of these sub-periods based on Case 1 or Case 2 as applicable.
- The reduction factor for the whole period is the product of the reduction factors for each sub-period.

For example, suppose we wish to project mortality rates from the PMA08 table, with an effective date of 1 July 2008, to a date of 6 April 2019. The reduction factor $RF(x, 1\text{ July }2008, 6\text{ April }2019)$ is the product of:

- $RF(x, 1\text{ July }2008, 1\text{ January }2009)$ using the method of Case 2
- $RF(x, 1\text{ January }2009, 1\text{ January }2010)$ using the method of Case 1
- $RF(x, 1\text{ January }2010, 1\text{ January }2011)$ using the method of Case 1
- …
- $RF(x, 1\text{ January }2018, 1\text{ January }2019)$ using the method of Case 1
- $RF(x, 1\text{ January }2019, 6\text{ April }2019)$ using the method of Case 2

6.3 Calculation of life expectancy and annuity value

Life expectancy is complete, rather than curtate, and annuity values are in advance. Both are calculated using q-style mortality rates.

Period life expectancy

Period life expectancy at age x in year y is calculated as:

$$LE_{Period}(x, y) = \tfrac{1}{2} + \sum_{t=1}^{150-x} S_{Period}(t)$$

where the survival index $S_{Period}(t)$ is defined recursively by:

$$S_{Period}(0) = 1$$
$$S_{Period}(t+1) = S(t) \times (1 - q_{Illustrative}(x+t, y))$$

Cohort life expectancy

Cohort life expectancy at age x in year y is calculated as:

$$LE_{Cohort}(x,y) = \frac{1}{2} + \sum_{t=1}^{150-x} S_{Cohort}(t)$$

where the survival index $S_{Cohort}(t)$ is defined recursively by:

$$S_{Cohort}(0) = 1$$
$$S_{Cohort}(t+1) = S(t) \times (1 - q_{Illustrative}(x+t, y+t))$$

Period annuity

The period annuity at age x in year y is calculated as:

$$Ann_{Period}(x,y,RetAge,i) = \sum_{t=\max(0,RetAge-x)}^{150-x} S_{Period}(t)\,(1+i)^{-t}$$

where the survival index $S_{Period}(t)$ is the same as for the life expectancy calculation, $RetAge$ is the retirement age, and i is the specified interest rate.

Cohort annuity

The cohort annuity at age x in year y is calculated as:

$$Ann_{Cohort}(x,y,RetAge,i) = \sum_{t=\max(0,RetAge-x)}^{150-x} S_{Cohort}(t)\,(1+i)^{-t}$$

where the survival index $S_{Cohort}(t)$ is the same as for the life expectancy calculation, $RetAge$ is the retirement age, and i is the specified interest rate.

Comparison

The life expectancies and annuity values are related by:

$$LE_{Period}(x,y) = Ann_{Period}(x,y,RetAge = x, i = 0\%) - \frac{1}{2}$$

$$LE_{Cohort}(x,y) = Ann_{Cohort}(x,y,RetAge = x, i = 0\%) - \frac{1}{2}$$

6.4 Extrapolation of base mortality tables

The software provides a number of standard base mortality tables that can be used to produce illustrative life expectancies and annuity values.

Most of these tables have been produced by the CMI. These have mortality rates up to age 120, and the mortality rates at age 120 and above is assumed to be 1.

The exception is the National Life Tables, produced by the Office for National Statistics, where the published tables only have mortality rates up to age 100. For the purpose of illustration in the software we have extrapolated rates to higher ages. This has been done by assuming that, for each gender, the difference in central mortality rates between the National Life Table and the A_C00 table (where the underscore represents sex) is constant for ages 100 to 119; i.e.:

- Calculate approximate central mortality rates for A_C00, and the National Life Table:

$$m_x^{A\_C00} = -\log(1 - q_x^{A\_C00}) \qquad \text{for } x \geq 100$$

- Extrapolate the National Life Table, assuming a constant margin between the central mortality rates:

$$m_x^{NLT*} - m_x^{A\_C00} = m_{100}^{NLT} - m_{100}^{A\_C00} \text{ for } 100 < x < 120$$

- Convert from central rates:

$$q_x^{NLT*} = 1 - \exp(-m_x^{NLT*}) \qquad \text{for } 100 < x < 120$$

7. Notation

This section collects the notation used in this paper in one place, for ease of reference:

- The notation is arranged thematically, rather than alphabetically.
- The "Meaning" either specifies what an expression means, where this can be done concisely, or refers to the relevant section of this paper for more information.
- For parameters with a single value, the final column shows the values used for the CMI_2018 Core Model.
- Input parameters are shown in bold

Expression	Meaning	CMI_2018 Core value
X_{min}	**Minimum age in the calibration data**	**20**
X_{max}	**Maximum age in the calibration data**	**100**
X_{taper}	**Age at which mortality improvements taper to nil**	**110**
X^{C}_{low}	**Cohort components of initial mortality improvements are nil at this age and below**	**30**
X^{C}_{high}	**Cohort components of initial mortality improvements are nil at this age and above**	**110**
Y_{min}	**Minimum calendar year in the calibration data**	**1978**
Y_{max}	**Maximum calendar year in the calibration data**	**2018**
C_{min}	Minimum cohort in the calibration data $C_{min} = Y_{min} - X_{max}$	1878
C_{max}	Maximum cohort in the calibration data $C_{max} = Y_{max} - X_{min}$	1998
C^{*}_{min}	Minimum cohort that is fitted directly $C^{*}_{min} = Y_{max} - X^{C}_{high}$	1908
C^{*}_{max}	Maximum cohort that is fitted directly $C^{*}_{max} = Y_{max} - X^{C}_{low} - 1$	1987
N_X	Number of ages in the calibration data $N_X = X_{max} + 1 - X_{min}$	81
N_Y	Number of calendar years in the calibration data $N_Y = Y_{max} + 1 - Y_{min}$	41
N_C	Number of cohorts in the calibration data $N_C = C_{max} + 1 - C_{min}$	121
N^{*}_{C}	Number of cohorts that are fitted directly $N^{*}_{C} = \mathrm{C}^{*}_{\mathrm{max}} + 1 - \mathrm{C}^{*}_{\mathrm{min}}$	80

Expression	Meaning	CMI_2018 Core value
$\bar{t}$	The mean of the calendar years in the calibration data $\bar{t} = \frac{1}{2}(Y_{min} + Y_{max})$	1998
$\bar{c}$	The mean of the cohorts in the calibration data $\bar{c} = \frac{1}{2}(C_{min} + C_{max})$	1938
$\boldsymbol{n}$	**Used in Section 3.5 to specify the age range used for exposure adjustments**	**2**
$\boldsymbol{p}$	**Probability threshold, used in Section 3.5 to specify exposure adjustments**	**1%**
λ	The weight placed on a penalty; e.g. λ_α applies to α_x	$\lambda_\alpha = 10^7$; $\lambda_\beta = 10^9$; $\lambda_\kappa = 10^7$; $\lambda_\gamma = 10^7$
$\boldsymbol{S}$	**Smoothing parameter; the base-10 logarithm of λ; e.g.$S_\alpha = \log_{10}\lambda_\alpha$ and $\lambda_\alpha = 10^{S_\alpha}$**	$\boldsymbol{S_\alpha = 7}$; $\boldsymbol{S_\beta = 9}$; $\boldsymbol{S_\kappa = 7}$; $\boldsymbol{S_\gamma = 7}$
$\boldsymbol{LTR}$	**Single-value specification of the age-period long-term rate L_x^{AP}**	**Specified by user**
L_x^{AP}	Long-term rate of age-period mortality improvements for age x	LTR if $x \leq 85$; Nil if $x \geq 110$; else interpolated
$\boldsymbol{L_x^C}$	**Long-term rate of cohort mortality improvements for age x**	**0**
$\boldsymbol{A}$	**Single-value initial addition to mortality improvements**	**0**
$MI_x^{InitAdd*}$	Initial addition to mortality improvements for age x	A if $x \leq 85$; Nil if $x \geq 110$; else interpolated
$\boldsymbol{T_x^{AP}}$	**Convergence period of age-period mortality improvements for age x**	**See Table 5.1**
$\boldsymbol{T_x^C}$	**Convergence period of cohort mortality improvements for age x**	**See Table 5.1**
$\boldsymbol{D_x^{AP}}$	**Initial direction of travel of age-period mortality improvements for age x**	**0**
$\boldsymbol{D_x^C}$	**Initial direction of travel of cohort mortality improvements for age x**	**0**
$\boldsymbol{Ppn_x^{AP}}$	**Proportion remaining at midpoint for age-period mortality improvements for age x**	**50%**
$\boldsymbol{Ppn_x^C}$	**Proportion remaining at midpoint for cohort mortality improvements for age x**	**50%**
x	Age at last birthday	
X	A specific age, used in Section 3.6	

Expression	Meaning
G	An age group used by the ONS (e.g. 15-44, 45-64, 65-74, 75-84, 85+)
G_x	The age group that contains age x
t	Calendar year
T	A specific calendar year, used in Section 3.6
c	Cohort, with $c = t - x$
$q_{x,t}$	Initial mortality rate, derived from the calibration data
$m_{x,t}$	Central mortality rate, derived from the calibration data
m_x^{NLT}	Central mortality rate from the latest ONS National Life Table for England & Wales
$D_{x,t}$	Number of deaths at age x in year t
$D_{G,t}^{Grouped}$	Number of deaths for age-group G in year t
$E_{x,t}$	Exposure at age x in year t
$E_{x,t}^0$	Initial approximate exposure estimate, used in Section 3.4
$A_{G,t}$	Adjustment made to an initial approximate exposure estimate, used in Section 3.4
$\bar{A}_G$	Adjustment made to an initial approximate exposure estimate, used in Section 3.4
$Scale_G$	Scaling factor used to estimate deaths in year Y_{max}, in Section 3.5
$r_{X,T}$	Deviance residual, used in Section 3.6
Φ	Cumulative distribution function of the standard Normal distribution
$\log$	Natural logarithm (base e)
$\log_{10}$	Base-10 logarithm

Expression	Meaning
α_x	APCI model parameters by age relating to mortality rates
β_x	APCI model parameters by age relating to mortality improvements
κ_t	APCI model parameters by period (i.e. calendar year)
γ_c	APCI model parameters by cohort (i.e. birth year)
$\underline{\alpha}$	Vector containing the values α_x
$\underline{\beta}$	Vector containing the values β_x
$\underline{\kappa}$	Vector containing the values κ_t
$\underline{\gamma}$	Vector containing the values γ_c
$\alpha_x^{(i)}$	Value of α_x after i iterations of Newton's method
$\beta_x^{(i)}$	Value of β_x after i iterations of Newton's method
$\kappa_t^{(i)}$	Value of κ_t after i iterations of Newton's method
$\gamma_c^{(i)}$	Value of γ_c after i iterations of Newton's method
$Deviance$	Deviance of the fitted APCI model
$Penalty$	Penalty term relating to a set of parameters; e.g. $\mathrm{Penalty}(\alpha_x)$ relates to α_x parameters
$\nabla^{(i)}$	An i^{th} order difference; defined recursively by $\nabla^{(i+1)}s_j = \nabla^{(i)}s_j - \nabla^{(i)}s_{j-1}$ *with* $\nabla^{(0)}s_j = s_j$.
$\underline{D}$	Difference matrix; e.g. $\underline{D}_\alpha$ corresponds to α_x
ϕ_i	Dummy parameters used when explaining Newton's method in Section 4.3
$\frac{\partial f}{\partial \phi_i}$	First order partial differential of function f with respect to parameter ϕ_i
$\frac{\partial^2 f}{\partial \phi_i \partial \phi_j}$	Second order partial differential of function f with respect to parameters ϕ_i and ϕ_j
$Include$	A criterion used in determining partial differentials of the $Deviance$ with respect to cohort parameters, in Section 4.4.
θ_i	Parameter used in specifying identifiability criteria, in Section 4.5
LSE	Error function for least-squares regression, in Section 4.5

Expression	Meaning
$MI_{x,t}$	"q-style" mortality improvement $MI_{x,t} = 1 - q_{x,t} \div q_{x,t-1}$
$MI^{*}_{x,t}$	"m-style" mortality improvement $MI^{*}_{x,t} = log\, m_{x,t-1} - log\, m_{x,t}$
$MI^{AP(0)*}_{x,t}$	Age-period component of "m-style" mortality improvement, derived directly from the APCI model
$MI^{AP*}_{x,t}$	Age-period component of "m-style" mortality improvement
$MI^{C*}_{c,t}$	Cohort component of "m-style" mortality improvement
MI^{P*}_{t}	Period component of "m-style" mortality improvement
DoT_t	Direction of travel derived from the APCI model
$MI^{ConstAdd*}_{x}$	Constant addition to mortality improvements
$RF(x, d_A, d_B)$	Reduction factor for age x from date d_A to date d_B
$q_{Illustrative}(x, d)$	Mortality rate for age x at date d from a chosen base table
$p(d_A, d_B, Y)$	Proportion of a year from dates d_A to d_B, both contained in the range 1 January Y to 1 January (Y+1)
$DayInterval(d_A, d_B)$	Number of days in the interval from d_A to d_B
LE_{Period}	Period life expectancy (complete)
S_{Period}	Survival index on a period basis
LE_{Cohort}	Cohort life expectancy (complete)
S_{Cohort}	Survival index on a cohort basis
Ann_{Period}	Period annuity in advance
Ann_{Cohort}	Cohort annuity in advance

References

Working Paper 74 “The CMI Mortality Projections Model, CMI_2014”. (2014)

Working Paper 84 “The CMI Mortality Projections Model, CMI_2015”. (2015)

Working Paper 90 “CMI Mortality Projections Model consultation”. (2016)

Working Paper 91 “CMI Mortality Projections Model consultation – technical paper”. (2016)

Working Paper 93 “CMI Mortality Projections Model: Consultation responses and plans for CMI_2016”. (2017)

Working Paper 106 “A proposed approach to closing off CMI mortality tables”. (2018)

Working Paper 119: “CMI Mortality Projections Model: CMI_2018” (2019)

“CMI_2018 software user guide” (2019)

These CMI working papers may be accessed and downloaded from the Mortality Projections Committee section of the Institute and Faculty of Actuaries’ website: https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/cmi-working-papers/mortality-projections.

Most of these outputs are only accessible to Authorised Users (i.e. to employees of subscribers and to researchers, for non-commercial use).

This page is deliberately blank.

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: Two London Wall Place, 123 London Wall, London, EC2Y 5AU
Email: info@cmilimited.co.uk
Tel: 020 7776 3820