European Actuarial Journal
https://doi.org/10.1007/s13385-022-00309-1

CASE STUDY

A general framework for analysing the mortality experience of a large portfolio of lives: with an application to the UK universities superannuation scheme

Andrew J. G. Cairns[1] · David Blake[2] · Kevin Dowd[3] · Guy D. Coughlan[4] · Owen Jones[5] · Jeffrey Rowney[4]

Received: 5 November 2021 / Revised: 21 January 2022 / Accepted: 9 March 2022
© The Author(s) 2022

Abstract
We propose a general framework that can be used to analyse the mortality experience of a large portfolio of lives. The objective of the framework is to provide a firm evidence base to support the setting of future mortality assumptions for the portfolio as a whole or subgroup-by-subgroup. The framework is developed in tandem with an analysis of the mortality of pensioners in the Universities Superannuation Scheme (USS), the largest funded pension scheme in the UK and one with a highly educated and very homogeneous membership. The USS experience was compared with English mortality subdivided into deprivation deciles using the Index of Multiple Deprivation (IMD). USS was found to have significantly lower mortality rates than even IMD-10 (the least deprived of the English deciles), but with similar mortality improvement rates to that decile over the period 2005–2016. Higher pensions were found to predict lower mortality, but only weakly so, and only for persons who retired on the first day of a month (mostly from active service). We found that other potential covariates derived from an individual's post/zip code (geographical region and the IMD associated with their local area) typically had no explanatory power. This lack of dependence is an important conclusion of the USS-specific analysis and contrasts with others that consider the mortality of more heterogeneous scheme memberships. Although the key findings are likely to be particular to USS, we argue that our analytical framework will be useful for other large pension schemes and life annuity providers.

Keywords Longevity risk · Pensioners' mortality · Index of multiple deprivation · Age standardised mortality rate · Occupation · Graphical diagnostics

✉ Andrew J. G. Cairns
A.J.G.Cairns@hw.ac.uk

Extended author information available on the last page of the article

Published online: 29 April 2022

1 Introduction

The actuarial and demographic literature has no shortage of analyses of both national and sub-population mortality. But it is quite rare to have the opportunity to analyse a sub-population that is quite as homogeneous as the Universities Superannuation Scheme (USS).

USS is a UK funded multi-employer pension scheme that provides mainly defined benefit pensions to university staff at almost all UK universities. The membership comprises mostly academics and senior professional services staff. The multi-employer nature of the scheme means that typical durations of active service that count towards a USS pension are high relative to many other schemes which, in combination with higher than national average salaries, lead to higher average pension amounts compared with other large schemes. The homogeneity of the scheme membership stems from the fact that most members are (effectively by definition) highly educated and have very similar jobs and working environments. This homogeneity leads to very specific conclusions in this paper about rates of mortality that could be quite different from other groups of lives.

In most developed countries, national mortality rates are easy to obtain from either national statistics agencies or the Human Mortality Database (HMD).[1] Sub-population mortality analyses also exist in a variety of settings. In the UK, for example, sub-population analyses or data are available for:

- the nations within the UK (Scotland, England, Wales and Northern Ireland: see HMD);
- geographical regions within, for example, England (see, for example, [1], and later in this paper);
- Index of Multiple Deprivation deciles (IMD) (see, for example, Department of Communities and Local Government, 2015 [15], and later in this paper);
- pension and insurance datasets (for example, [11]).

More detailed analyses of insurance or pension portfolios can take into account portfolio- or individual-specific information, such as pension amounts [11], industry [10] or geodemographic profile [28].

For other countries, covariates such as education [23] and income or affluence [2, 6, 29] have been found to be highly significant predictors of mortality rates at the national level.

In most cases, covariates such as geodemographic profile, region, affluence (or conversely deprivation) associated with local area of residence, and pension amount all have significant power to predict higher or lower levels of mortality. In turn, this leads to greater accuracy in the valuation of mortality-linked pension liabilities. Although our methodological approach was developed specifically to explain the idiosyncrasies of a particular pension scheme, we believe it can be adopted as a general framework for analysing the mortality experience of other large datasets.

[1] http://www.mortality.org.

The structure of the paper is as follows. Section 2 introduces the USS mortality data. Section 3 describes the CBDX mortality model [16] used to smooth some of the mortality data. Section 4 provides an empirical comparison of USS versus English mortality data by IMD decile, with more formal statistical tests in Sect. 5. Section 6 discusses a first-of-the-month effect, where the day of the month on which an individual retires turns out to be important. The significance of pension amount is considered in Sect. 7. Section 8 discusses future short- and long-term mortality rate improvement assumptions for the USS in the light of both national and international comparisons. Section 9 summarises the findings for the USS from the previous sections and makes recommendations based on those findings. Section 10 concludes with lessons about how the general framework can be used to analyse the mortality experience of other large pension schemes and life annuity providers. This paper builds on an earlier preliminary study of the longevity of USS members by Coughlan [13].

2 Description of the data

The USS pensioner dataset consists of 98,733 individuals who were directly in receipt of a USS pension at some time during the period 2005–2016 (inclusive).[2] Pensions in payment at the end of 2016 amounted to around £1.4 billion per annum. Pensioner 'types' comprise: primary pensioners (i.e., former active or deferred members who retired in 'normal health'); ill-health pensioners (former active members who took early retirement due to ill health, normally with enhanced benefits); spouse pensioners (i.e., spouses or long-term partners of deceased primary or ill-health pensioners); ex-spouse pensioners; child pensioners (i.e., children of deceased primary pensioners while they are still in full-time education); and other dependent pensioners. This analysis focuses on primary pensioners and spouses (approximately 93% of the lives and 95% of annual pension payments).

Data for each pensioner (in addition to 'type') comprise sex, dates of birth, retirement and death, pension amount, and their residential post/zip code. Pension amounts are either as at 1 January 2017 for survivors or at the date of death revalued to 1 January 2017 in line with general USS pension increases.

The data were subject to extensive data validation tests as well as a range of diagnostics designed to identify certain idiosyncratic features of the USS data. Data validation included identification of outliers in the data which were subsequently excluded from the analysis. Specifically, 93 out of 76,895 primary pensioners were considered to have inaccurate records including: age at retirement before 50 or after 80 (primary pensioners); date of birth before 1903; and date of exit preceded date of retirement.

[2] This number excludes a small proportion of records of two types with zero pensions. The first type consisted of individuals with duplicate records. The second type concerned individuals with trivial pensions that were commuted in full, and with whom there has been no subsequent contact by USS.

Table 1 Exposures (person years) and death counts for primary pensioners and spouse pensioners for the period 2005–2016

	Males	Females
Primary exposures	311,901	99,504
Spouse exposures	4908	72,740
Primary deaths	7229	1471
Spouse deaths	162	2219

With outliers removed, exposures[3] and deaths over the period 2005–2016 are given in Table 1. For comparison, exposures for the English national population over the 12 years from 2005–2016 and ages 65–89 are approximately 45 million person years (males) and 55 million person years (females), so USS pensioners' exposures are, on average, about 0.7% (males) and 0.3% (females) of the size of the corresponding English population.

Graphical diagnostics included scatterplots of pairs of quantities, such as date of birth, date of retirement, age at retirement, age at death and pension amount. The plots helped identify outliers and various idiosyncratic features of the data. Of most interest, we found that the day of the month on which primary retirees are recorded as having retired was strongly correlated (although not necessarily causally connected) with the age at retirement and the pension amount. The importance of this feature of the data is explored further in Sects. 6 and 7.

3 The CBDX model

The analysis below makes use of the CBDX model of Dowd et al. [16].[4] This results in some smoothing in both the age and time dimensions. Even though USS is one of the largest UK pension schemes, the pensioners population, at around 0.3% to 0.7% of the equivalent English population, is small enough that there will be considerable small-population noise in the crude death rates. Application of the CBDX model dampens this noise considerably by exploiting the generally stable relationship in sufficiently large populations between death rates at different ages and through time. The model for the underlying death rate $m(t, x)$ is:

$$\log m(t,x) = \alpha(x) + \kappa_1(t) + \kappa_2(t)(x - \bar{x}) \tag{1}$$

for calendar year t, age x last birthday (which we abbreviate to (t, x)) and birth cohort $t - x$, where

- $\alpha(x)$ is a non-parametric age effect,[5] similar to that in the Lee-Carter [21] model;

[3] Short for exposures to risk, which is not the same as population.

[4] A two-population generalisation of the Cairns–Blake–Dowd [4] model which builds on earlier proposed generalisations, including those of Plat [27] and Wen et al. [31, 32].

[5] Or base mortality table.

- $\bar{x}$ is the mid-point of the age range considered;
- $\kappa_1(t)$ and $\kappa_2(t)$ are period effects.

Adding a third (quadratic) age-period component turned out not to be helpful: it resulted in overfitting within each year together with an unstable fit from year to year at both high and low ages in response to randomly high and low death counts.

The (t, x) death count is denoted by $D(t, x)$ and, conditional on $m(t, x)$, is assumed to have a Poisson distribution with mean $m(t, x)E(t, x)$, where $E(t, x)$ is the central exposed-to-risk—the total number of years lived by pensioners in calendar year t at age x last birthday. $E(t, x)$ can be computed exactly for the USS dataset. In this analysis, we take a sequential approach that first estimates $\alpha(x)$ with smoothing and then the period effects, given $\alpha(x)$.

4 Comparison of USS mortality rates with English deprivation deciles

As previously mentioned, even for a pension scheme of the size of USS, there can be significant noise in both crude and fitted mortality rates due to small-population noise in the death counts, particularly at high and low ages where expected deaths are low. It is helpful, therefore, to consider the USS experience alongside larger national datasets, where we have much larger exposures and, potentially, more years. In this case, we compare the USS experience with the English national population subdivided into deprivation deciles. Deprivation decile number 1 (IMD-1) covers the most deprived areas in England, moving up to decile 10 (IMD-10) which covers the least deprived areas. Construction of the deciles is described in Department of Communities and Local Government [15] (main report plus technical report). IMD-based mortality rate data are available for the period 2001–2016 and can be obtained from the UK's Office for National Statistics (ONS).

4.1 Graphical diagnostics

The plots that follow present results for USS in comparison with data obtained from the ONS. In Figs. 1 and 2, we plot death rates for USS and for the English IMD-1 to IMD-10 deprivation deciles.

The IMD death rates are plotted in their (almost)[6] unsmoothed form. The death rates for the IMD deciles in the two figures *without any smoothing* show a reasonably clear ordering for both calendar year and age. We also see a pattern of inequality that is typical for a variety of countries and socio-economic measures: a high degree of inequality at ages 60 and 65 that gradually narrows at higher ages.

The IMD deciles exhibit two further features of note. First, annual improvements in mortality have been higher for the least deprived deciles (so the gap is getting wider). Second, all deciles have exhibited a slow down in recent years, consistent

[6] We have adjusted the IMD decile exposures using the methodology of Cairns et al. [5] to remove, as far as possible, perceived anomalies in the data.

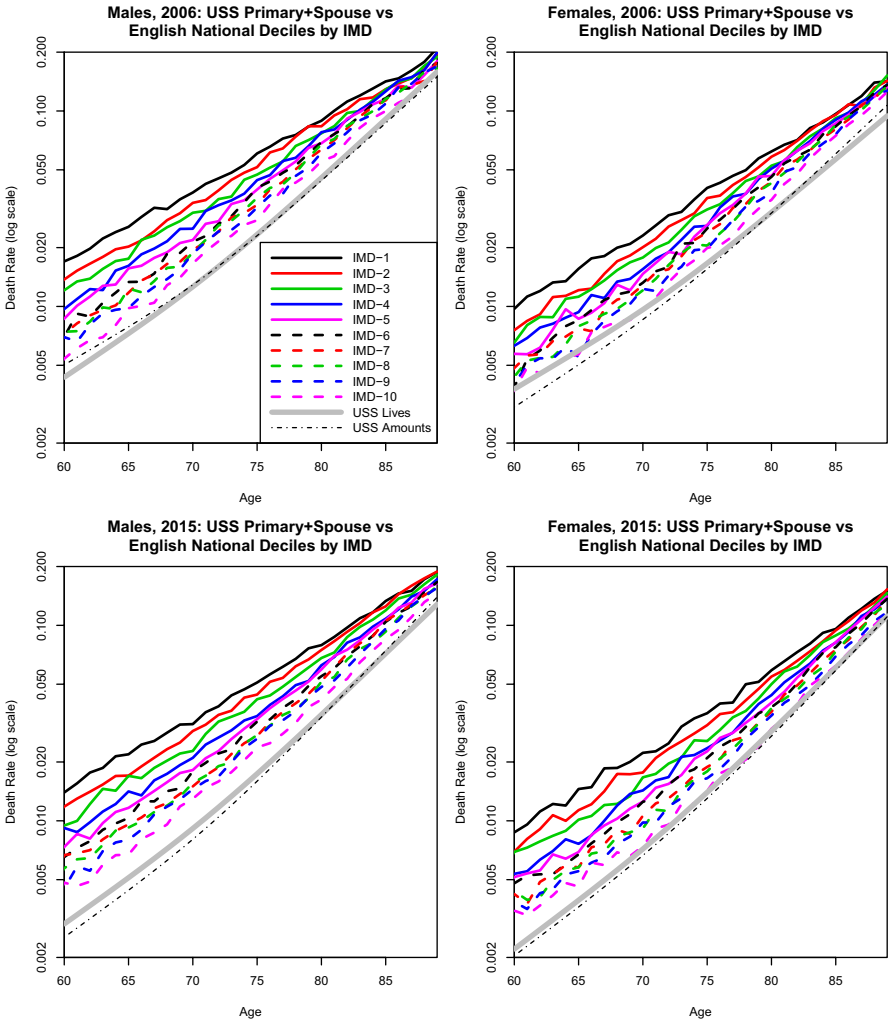

Fig. 1 Plots of mortality rates by age in specific years and for males and females. IMD rates are crude rates. USS rates follow model fitting to ages 60–89, and years 2005–2016

with the much-discussed national slow down in mortality rate improvements (but IMD-10 is still improving at a faster rate than IMD-1; see, for example, Office for National Statistics, [25]).

The curves for USS are based on aggregated deaths and exposures for primary pensioners and spouses based on lives (thick grey line) and amounts (dot-dashed line).[7] To mitigate small-population noise due to low death counts, the plots show

[7] Deriving mortality rates by amounts is a somewhat ad hoc technique that is not underpinned by a formal statistical model. However, it does provide an estimate of the rate at which total pension amounts by cohort will decline in the next year. Secondly, the difference between mortality rates by lives and amounts reveals, empirically, some dependency between mortality rates and pension amounts.

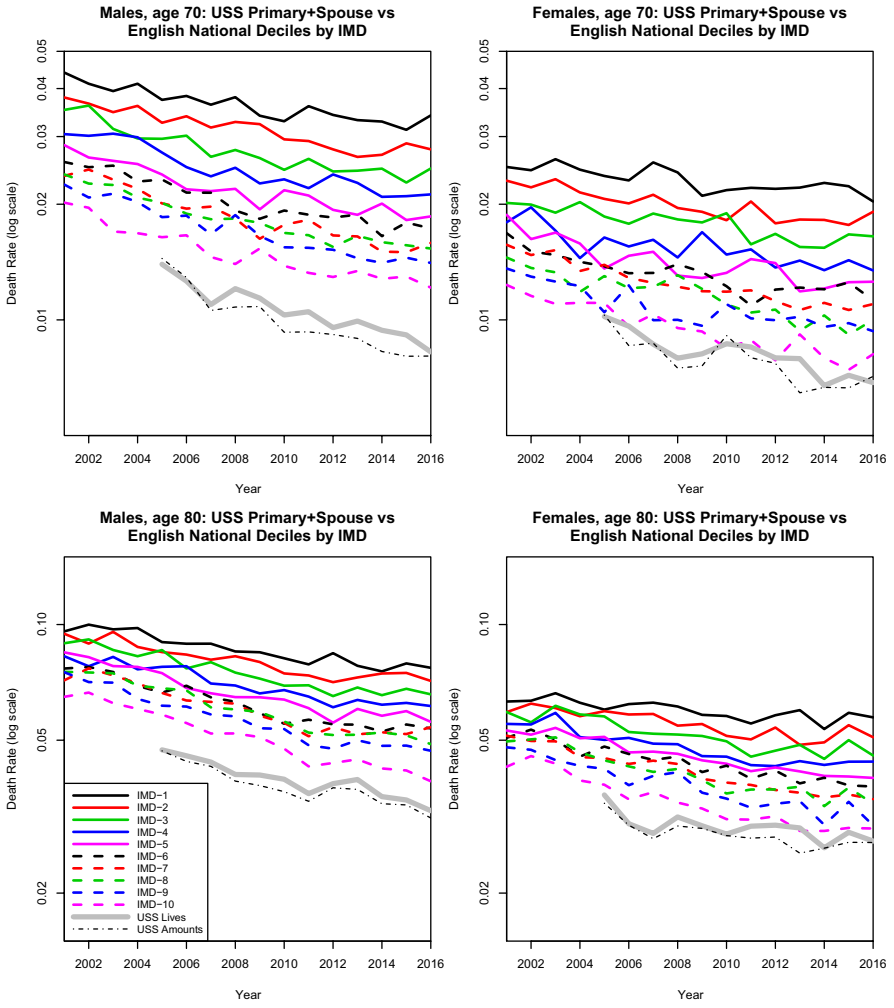

Fig. 2 Plots of mortality rates by year for specific ages and for males and females. IMD rates are crude rates. USS rates follow model fitting to ages 60–89, and years 2005–2016

fitted mortality rates based on the CBDX model. Mortality by amounts is lower than by lives, as is typical for other populations, implying that wealthier USS pensioners live longer than those who are less wealthy. However, the gap between the two is relatively small compared with other, less homogeneous populations (e.g., CMI standard tables S2PMA (amounts) and S2PML (lives); [11]).

We see that the USS death rates are generally quite low compared with the IMD death rates for both males and females. In particular, in the later years, the USS mortality rate curve is mostly below the least deprived English decile, IMD-10.

We next sacrifice the age-related detail in the data in order to compare USS with IMD *centiles* rather than deciles, ranging from centile 1 (most deprived) to centile

100 (least deprived). In Fig. 3, we plot Age Standardised Mortality Rates (i.e., the weighted average of the age-specific mortality rates, where the weights are the proportions of persons in the corresponding age groups of the European Standard Population (*ESP*) 2013). For population i, this is

$$ASMR_i(t) = \sum_{x=65}^{89} \hat{m}_i(t,x) ESP(x) / \sum_{x=65}^{89} ESP(x),$$

where $\hat{m}_i(t,x)$ is the crude death rate for population i, year t and age x. This clearly reveals the even stronger result that USS male mortality rates are lower than the average of the 1% least deprived areas in England, while for females they are around the level of the 1–2% least deprived.

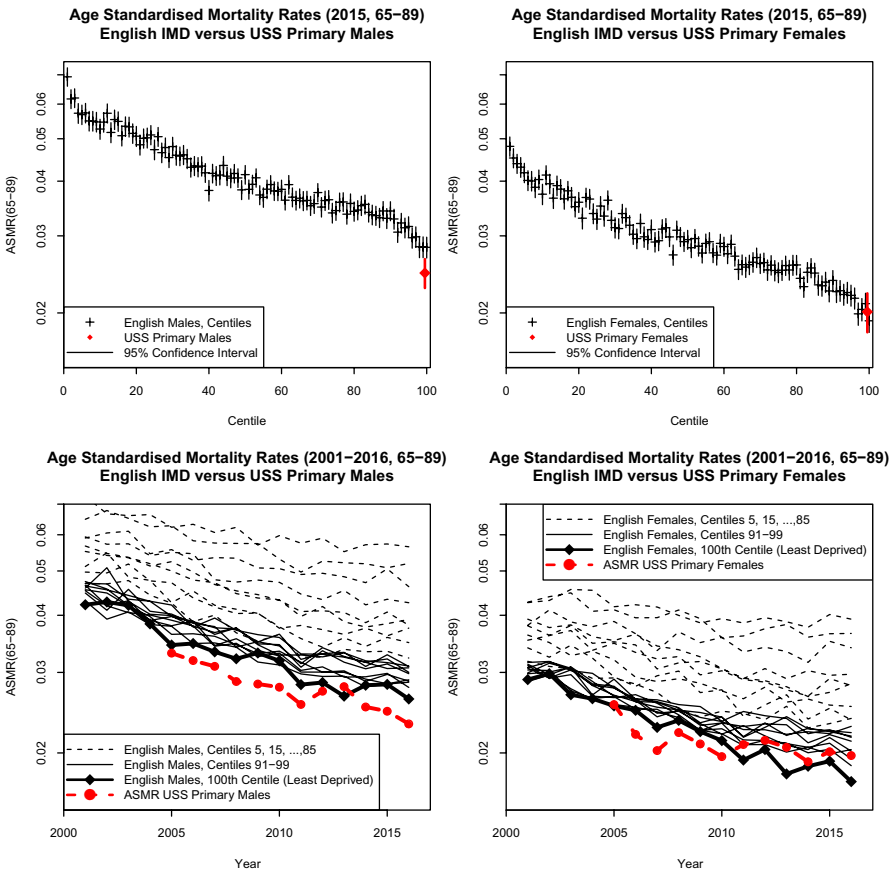

Fig. 3 Age standardised mortality rates for English deprivation centiles and for USS males and females. Standardisation is over ages 65–89 using the European Standard Population, 2013. Top row: ASMRs for 2015 in centiles with 95% confidence intervals. Bottom row: development of the ASMRs over time from 2001–2016 (selected IMD centiles) and 2005–2016 (USS)

The top row of Fig. 3 also shows how reliable the Index of Multiple Deprivation is as a predictor of high or low mortality: that is, there is a clear shape to the plot with the centile on the x-axis, with a steepening at each end.

The clear conclusions from these three figures are:

- Figure 1 shows that USS members have lower mortality rates than IMD-10 across all ages for men and women;
- Figure 2 shows that mortality rate improvements for USS members are similar to IMD-10;
- Figure 3 shows that on average USS males over 65 years of age are in the 1% of the UK population with the lowest mortality rates, while USS females over 65 years of age are in the 1–2% of the UK population with the lowest mortality rates.

5 Mortality rate analysis I: which explanatory variables are significant?

In this section, we conduct a series of formal statistical tests in order to determine whether there are unique features of the USS dataset that are significantly different from the larger, more heterogeneous IMD dataset. We also look at the significance of pension amount within the USS males data, and whether geographical region or the place of residence of a member within a specific IMD decile have significant power in explaining mortality rates.

5.1 Test 1: explanatory power of the IMD group

The IMD deciles discussed earlier consist of a collection of small geographical areas called *Lower Super Output Areas* (LSOAs), each having its own index of multiple deprivation (IMD).[8] Figures 1 and 2 indicate clearly that, *in the absence of other information*, the IMD decile to which an individual belongs by virtue of their place of residence is a significant predictor of mortality.

With the additional information that an individual is a member of USS, does this explanatory power change? The USS data allow us to identify the LSOA, and hence the IMD decile, to which each member belongs (see Fig. 4) if they are resident in England—which accounts for about 75% of members. Table 2 summarises the total exposures and deaths over ages 65–89 and years 2005–2016 in each IMD group. The eleventh Group, 'X', counts pensioners who have not been allocated to an English LSOA either because they live in other parts of the UK or overseas, or have an unidentified post/zip code. Unsurprisingly, exposures increase steadily from IMD-1 to IMD-10: that is, USS pensioners mainly live in less deprived areas of England.

[8] E.g., IMD-1 contains the 10% of LSOAs with the highest IMDs.

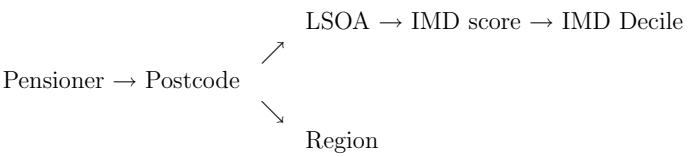

Fig. 4 The process of linking individual pensioners to an IMD decile (Test 1) or to a geographical region (Test 2)

Table 2 Total exposures and death counts for male and female primary pensioners aged 65–89, 2005–2016 allocated to IMD deciles by place of residence

	Males			Females		
IMD group	Exposures	Actual deaths	Expected deaths	Exposures	Actual deaths	Expected deaths
1	2601	72	63.0	984	22	17.1
2	7560	176	160.3	2602	38	34.1
3	9137	220	213.7	3511	49	51.0
4	13,303	285	288.9	5064	79	73.5
5	18,112	414	400.4	6726	81	91.9
6	22,748	511	540.3	7880	122	113.9
7	29,804	658	695.5	10,155	136	150.3
8	33,474	785	796.0	10,867	139	155.9
9	40,778	976	998.0	12,616	172	194.2
10	56,509	1375	1415.1	15,564	236	230.4
X	77,876	1757	1657.8	23,535	397	358.8

Expected deaths are calculated under the null hypothesis, H0. Individuals not resident in an identifiable Lower Super Output Area in England are allocated to Group X

For this part of the framework to be useful, we require a tool that maps post/zip code to a set of distinct socio-economic groups. Each group must have its own historical mortality experience and, ideally, this should be based on national population and deaths data aggregated into socio-economic groups using an appropriate statistical method. In this case study, the mapping tool combines a post-code-to-LSOA lookup table and the English IMD, which provides a deprivation measure for each LSOA.

In order to conduct hypothesis tests, we define $E(i, t, x)$ to be the exposed to risk for USS pensioners aged x in year t allocated to IMD group i, and $D(i, t, x)$ and $m(i, t, x)$ to be the corresponding death counts and death rates.

Assuming deaths, $D(i, t, x)$, follow a Poisson distribution with mean $m(i, t, x)E(i, t, x)$, we consider the following two hypotheses:

Table 3 Results of Test 1: explanatory power of the IMD group

	Test statistic, S	95% critical threshold	p-value	Conclusion
Males	233.9	295.5	0.835	Accept H0
Females	284.8	348.0	0.727	Accept H0

- H0: $m(i,t,x) = m(t,x)$ for all i, t, and x (i.e., there is no difference in death rates between the 11 groups, apart from differences between ages and over time), versus
- H1: $m(i,t,x) = A(i,x)m(t,x)$ where the $A(i, x)$ are constants to be estimated that depend on the IMD group, i, and age, x, but not year, t (that is, death rates differ across all groups).[9,10]

Under H0, there are $12 \times 25 = 300$ parameters to be estimated. Under H1, there are $12 \times 25 + (11 - 1) \times 25 = 550$ parameters to be estimated.[11] Let l_0 and l_1 be the maximum likelihoods under H0 and H1, respectively. We use the likelihood ratio test (in the spirit of Chen et al., [8]) which has test statistic $S = 2(l_1 - l_0)$ and we reject H0 at the α confidence level if S exceeds the α quantile of the *simulated* distribution of the test statistic under H0.[12]

For the data summarised in Table 2, the test statistic for males of $S = 233.9$ (Table 3) gives a p-value of 0.835. So we accept H0 and conclude that we have no evidence that the mortality rates of USS male primary pensioners are dependent on their IMD group. For females, we have a similar result: a test statistic of $S = 284.8$ and a p-value of 0.727. In other words, USS pensioners living in more deprived areas of England—for example, because of high house prices in other more desirable parts of cities such as London—do not suffer from higher rates of mortality.

This is a very significant finding. In the general population, the IMD group is a very strong predictor of high or low mortality, but this does not hold for USS male pensioners (see, for example, Figs. 1 and 2). A likely explanation is that the strong homogeneity of the USS membership—a very specific and highly educated occupational group—outweighs other information that has proved to be a useful predictor of mortality rates for other more heterogeneous segments of the general population.

5.2 Test 2: explanatory power of geographical region

5.2.1 Test 2A: regional differences

As an alternative to grouping by IMD decile, primary pensioners were grouped according to the region in which they are resident. There are nine English regions, and pensioners not in an identifiable LSOA are, again, allocated to the 'other' Group

[9] Hence, for two groups i and j, under H1, $m(i,t,x)/m(j,t,x) = A(i,x)/A(j,x)$ for all ages x: that is, the relationship between groups remains the same across all years.

[10] H1 does not require that the USS IMD groups are ordered—just that their death rates are different. If we reject H0, then the next step would be to investigate what sort of dependence there is on the IMD group.

[11] The inclusion of the $A(i, x)$ parameters introduces an identifiability problem: $m(i,t,x) = A(i,x)m(t,x) = (A(i,x)k(x))(m(t,x)/k(x))$ for arbitrary $k(x) > 0$. This problem is avoided through the use of a set of 25 identifiability constraints: i.e., $A(10, x) = 1$ for each x.

[12] Often in a Likelihood Ratio Test, we would use the Chi-squared approximation for the distribution of the test statistic. However, because many of the 3300 individual death counts are less than 5, this approximation is less good. To account for this, we simulated the test statistic under the assumption that H0 was true using the common death rate $m(t, x)$ and the given exposures $E(i, t, x)$.

X. Exposures and deaths are summarised in Table 4.[13] We carried out a similar likelihood ratio test to Test 1 (on the basis of the same H0 and H1, but with different groupings). For males, the Test 2A statistic $S = 228.7$ (Table 5) gave a p-value of 0.467 and for females, $S = 255.1$, with a p-value of 0.285. So, for both males and females, we accept H0 and conclude that we have no evidence that the mortality rates of primary pensioners are dependent on the region in which they live.[14]

5.2.2 Test 2B: north–south differences

Although the previous test does not find evidence for differences at the level of individual regions, it is possible, given the actual versus expected deaths in Table 4, to consider the possibility of a north-south divide. To test for this, 'the north' was taken to mean regions 1–5 and X, and 'the south' to be regions 6–9.[15]

Table 6 presents the results of Test 2B, and we see that for both males and females there is reasonable (although not strong) evidence to reject the null hypothesis. That is, we can conclude that there is a difference in mortality rates between the north and the south. Statistically, for females, the evidence is less strong, reflecting the smaller population size. However, empirically, for females, the actual versus expected gap is wider than for males.

5.3 Test 3: difference between USS and IMD-10

In this test, we compare USS male and female death rates separately for ages 65–89 and years 2005–2015 with IMD-10 death rates.

Death rates are denoted by $m(i, t, x)$, where $i = 1, 2$ correspond to IMD-10 and USS, respectively. The only modelling assumption is that

$$D(i,t,x) \sim \text{ Poisson } (m(i,t,x)E_i(t,x)).$$

We consider the following two hypotheses:

- H0: $m(2,t,x) = m(1,t,x) = m(t,x)$ for all t and x (i.e., death rates for USS are the same as those for IMD-10), versus
- H1: $m(2,t,x) \neq m(1,t,x)$ for all t and x (i.e., death rates for USS differ from those of IMD-10).

Let l_0 and l_1 be the maximum log-likelihoods under H0 and H1. The test statistic for the likelihood ratio test is $S = 2(l_1 - l_0)$ and we reject H0 at the α confidence level if S exceeds the α quantile of the χ^2_ν distribution where, here, $\nu = 12 \times 25 = 300$ is the

[13] Compared with the regional spread of the general population, USS primary pensioners are disproportionately concentrated in London and the South East.

[14] It was feasible to split Group X further and retest. As an example, we split Group X in two (Scotland and other). However, Scottish mortality was similar to Group X as a whole, the p-value increased, and H0 was again accepted.

[15] Most pensioners in group X reside in Scotland, Wales and Northern Ireland: countries that are contiguous to and have similar socio-economic characteristics to the northern English regions. Hence, X is grouped in with regions 1–5.

Table 4 Total exposures and death counts for male primary pensioners aged 65–89, 2005–2016 allocated to Region groups by place of residence

Region	Males			Females		
	Exposures	Actual deaths	Expected deaths	Exposures	Actual deaths	Expected deaths
1 North East	9095	204	216.5	2349	34	41.2
2 North West	24,886	587	559.8	5330	81	72.6
3 Yorkshire and The Humber	22,839	531	521.5	5954	113	84.2
4 East Midlands	16,658	411	360.1	4838	73	64.8
5 West Midlands	17,790	419	412.7	4616	80	71.3
6 East	26,679	580	665.7	8822	114	117.4
7 London	37,517	857	892.2	14,607	190	221.1
8 South East	51,865	1244	1273.8	21,371	304	324.5
9 South West	26,697	639	668.9	8083	85	115.1
X Other	77,876	1757	1657.8	23,535	397	358.8

Expected deaths are calculated under the null hypothesis, H0. Individuals not resident in an identifiable Lower Super Output Area in England are allocated to Group X

Table 5 Results of Test 2A: explanatory power of the region

	Test statistic, S	95% critical threshold	p-value	Conclusion
Males	228.7	264.4	0.467	Accept H0
Females	255.1	280.6	0.285	Accept H0

Table 6 Results of Test 2B: explanatory power of the north-south divide

	Test statistic, S	95% critical threshold	p-value	Conclusion
Males	44.5	37.7	0.009	Reject H0
Females	38.9	37.8	0.040	Reject H0

Table 7 Results of Test 3: difference between USS and IMD-10

	Test statistic, S	95% critical threshold	p-value	Conclusion
Males	569.70	341.40	0.000	Reject H0
Females	335.41	341.40	0.078	Accept H0

number of degrees of freedom. Table 7 shows that we reject H0 for males and accept it for females.

As a slight note of caution, we observe that the IMD-10 population also includes USS pensioners, implying that the underlying assumption that the death counts in the two populations are independent is not strictly correct. However, since the overlap between USS and IMD-10 is only a small proportion of IMD-10 (about 1.1% of IMD-10 based on Table 2), the impact of this absence of strict independence is likely to be negligible. A further implication follows from the fact that the USS data have lower mortality rates than IMD-10, namely the IMD groups themselves are not completely homogeneous.

Finally, as we noted earlier, the use of Chi-squared for the distribution of the test statistic under H0 is an approximation that works better the larger the population size. Nevertheless, the high value of the test statistic for males far outweighs any questions over the validity of the Chi-squared distribution. On the other hand, if the p-value for the test is fairly close to the critical value (typically 0.05), we should be more cognisant of the independence issue.

5.4 Test 4: age-related difference between USS and IMD-10 only

We consider next the following two hypotheses:

Table 8 Results of Test 4: age-related difference between USS and IMD-10 only

	Test statistic, S	95% critical threshold	p-value	Conclusion
Males	234.31	314.68	0.964	Accept H0
Females	270.21	314.68	0.570	Accept H0

- H0: $m(2,t,x) = A(x)m(1,t,x)$ for all t and x (i.e., mortality rate improvements for USS are the same at those for IMD-10), versus
- H1: $m(2,t,x) \neq A(x)m(1,t,x)$ for all t and x (i.e., mortality rate improvements for USS differ from those for IMD-10),

where $A(x)$ is an unsmoothed, non-parametric age effect.

We again apply a likelihood ratio test across ages 65–89 and years 2005–2016; there are $\nu = 12 \times 25 - 25 = 275$ degrees of freedom. Table 8 shows the test results. We conclude that mortality rate improvements in the USS (ages 65–89) are indistinguishable from those of the corresponding IMD-10 data. Importantly, the IMD-10 dataset is much larger—and so has much less small-population noise—and has more years. This implies that we can reasonably forecast USS mortality rates using forecasts for IMD-10 and then applying the age adjustment, $A(x)$.

5.5 Test 5: a CBDX model comparison of USS and IMD-10

We next fit the CBDX model (Eq. (1)) to both USS and IMD-10 populations using the following algorithm:

- fit the model to IMD-10 (population 1) to get estimates of $\alpha(1,x)$, $\kappa_1(1,t)$, and $\kappa_2(1,t)$;
- fit the model to USS (population 2) to get estimates of $\alpha(2,x)$, $\kappa_1(2,t)$, and $\kappa_2(2,t)$; or
- fit the model to USS, but estimate $\alpha(2,x)$ only. Import the period effects from IMD-10: that is, fix $\kappa_1(2,t) = \kappa_1(1,t)$ and $\kappa_2(2,t) = \kappa_2(1,t)$.

This leads to the following two hypotheses:

- H0: CBDX model for USS with $\kappa_1(2,t) = \kappa_1(1,t)$ and $\kappa_2(2,t) = \kappa_2(1,t)$ (i.e., mortality rate improvements for USS derived from the CBDX model are the same at those for IMD-10), versus
- H1: CBDX model for USS with $\kappa_1(2,t)$ and $\kappa_2(2,t)$ to be estimated from the USS data (i.e., mortality rate improvements for USS derived from the CBDX model differ from those for IMD-10).

Table 9 Results of Test 5: a CBDX model comparison of USS versus IMD-10. CBDX smooths results across ages, but leaves year-to-year variation

	Test statistic, S	95% critical threshold	p-value	Conclusion
Males	27.10	36.42	0.300	Accept H0
Females	18.29	36.42	0.789	Accept H0

For the data under consideration, H1 has 24 more parameters. We again use the likelihood ratio test for primary pensioners and spouses, across ages 65–89 and years 2005–2016. Table 9 shows the test results. Under the stronger assumptions of the CBDX model, we still find that improvement rates in fitted USS mortality over the age range 65–89 are not significantly different from improvement rates in fitted IMD-10 mortality.

The importance of Test 5, as a model-based variant of Test 4, lies in the potential usefulness of the CBDX model for forecasting future mortality rates.

6 A first-of-the-month effect

We now expand upon a remark in Sect. 2 concerning the day of the month in which a pensioner is recorded as having retired and its connection with pension amount.

For primary retirees, Fig. 5 shows the relationship between an individual's age at retirement and the date during a calendar year when that individual retired. We observe a number of striking features. For both males and females, we see concentrations of dots along specific horizontals and verticals: verticals at the key ages of 50, 55, 60, 63.5 and 65;[16] horizontals correspond to the first of a month as the date of retirement. Relatively few individuals retired both later in the month and at an uncommon age. Furthermore, we found that pension amount had a clear link to the day of the month on which an individual is recorded as having retired.

Figure 6 shows the cumulative distribution of pension amounts for those who retired on the first of a month and for those who retired later in the month. The figure shows a striking difference between the two groups, with those *not* retiring on the first of a month receiving pensions that are approximately 40% on average of

[16] Over most of the relevant data period, USS constituent universities had the following common retirement ages: the lowest age for early retirement for a scheme member not in ill health (officially known as the normal minimum pension age, introduced in 2006 and set at 50, rising to 55 in 2010, following the Finance Act 2004), the normal age of female retirement (60), the normal age of male retirement (65) and a gender-neutral normal retirement age used by some universities (63.5) that was part way between the male and female normal retirement ages. The Pensions Act 1995 (later amended by the Pensions Act 2011) gradually raised the normal retirement age for women from 60 to equal that of men (65) between 2010 and 2018, and for the normal retirement age for both men and women to rise to 66 in 2020. The Equality Act 2010 removed an upper limit on the retirement age on the grounds of age discrimination (some universities had previously included in their contract of employment a specific maximum retirement age, e.g., 67 or 75; this became illegal under the Act, although employers can still require employees to retire on the grounds of incapacity).

those who retired on the first of a month. On further investigation, first-of-the-month retirees were found to be mostly pensioners who had retired from active service, with a significant peak in retirements at the beginning of October, the day after the end of the previous academic year. These individuals typically have long service and, consequently, relatively large pensions.[17] In contrast, later-in-the-month retirees were found to be mainly previously deferred pensioners. These will typically have shorter periods of service and, consequently, smaller pensions.

7 Mortality rate analysis II: does pension amount matter?

We have already seen some indication that USS mortality rates might be dependent on pension amount in Figs. 1 and 2. However, it is worth mentioning that the differences between lives- and amounts-based mortality rates are rather smaller with USS than is typical of standard tables (e.g. [9]).[18,19] We have already seen in the previous section that there is a link between the day of the month when a pensioner retired, membership status pre-retirement and pension amount (Figs. 5 and 6).

We now seek to investigate pension amount as an predictive variable in more depth.

7.1 Subdivision of first-of-month primary retirees

Our empirical observations based on Figs. 1, 2, 5 and 6 prompted us to divide the primary retirees as follows:

- Group 1: those recorded as having retired on the first of the month (mostly retirement from active service) with further subdivision into tertiles by pension amount (revalued to 1 April 2017)
  - 1A: revalued pension less than a threshold P_1;
  - 1B: revalued pension between thresholds P_1 and P_2;
  - 1C: revalued pension above P_2.
- Group 2: those who retired between the 2nd and last day of the month (mostly deferred members).

[17] USS covers a large proportion of UK university academic staff. Individuals might move from one university to another, but their accrued and prospective benefits in USS are not affected by this move, other than as a consequence of any change in salary.

[18] For example, the difference between the CBDX lives and amounts base tables, $\alpha(x)$ (Eq. 1), lies between 0.01 and 0.06 depending upon age and gender.

[19] Comparisons between lives- and amounts-based graduations need to be treated with considerable caution. Specifically, a proper comparison requires additional knowledge of the distribution of pension amounts around the mean at each age. Often, we do not have this information for standard tables. Everything else being equal, the difference between lives- and amounts-based graduations will be larger if the underlying pension amounts are more widely dispersed.

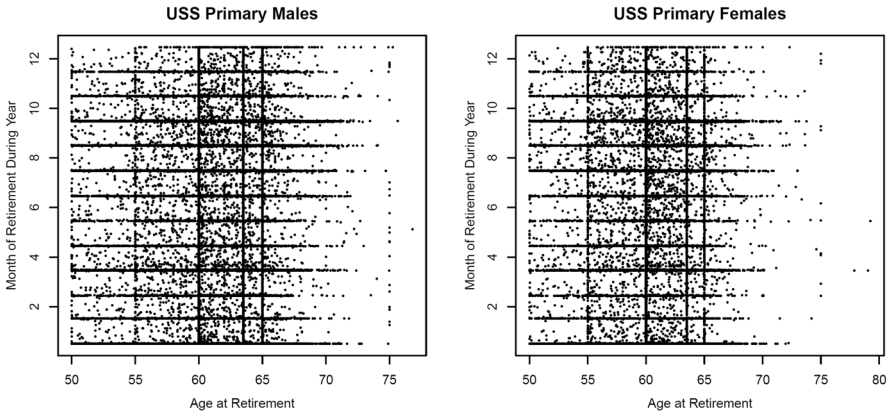

Fig. 5 Primary retirees: scatterplot of age at retirement versus exact time of year of retirement. Each dot represents one individual

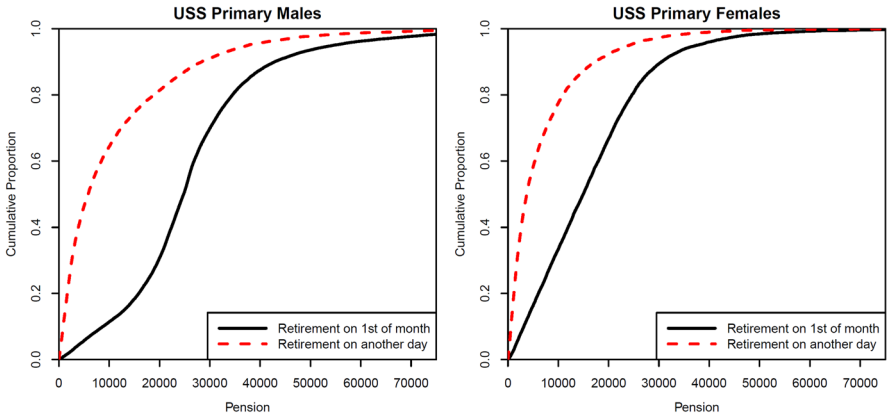

Fig. 6 Primary retirees: cumulative distribution of the pension amount for those who retire on the first of a month and for those who retire on the second of a month or later

The Group 1 thresholds P_1 and P_2 are set at the 1/3 and 2/3 quantiles of the empirical distribution of the pension amounts for all first-of-month primary retirees, calculated separately for males and females. Values are given in Table 10.

Various graphical diagnostics were considered to explore the impact of both the first-of-the-month effect and the pension amount. Figure 7 for males shows how the ASMR for ages 65–89 varies over time for different groups. Specifically, we can see that Group 1B is slightly below Group 1A and Group 1C is more clearly below 1B. All three lie below IMD-10. On the other hand, small-population noise in the smaller Group 2 prevents us from detecting any obvious deviation from Groups 1A to 1C. Although Group 2 USS pensions are much lower, Group 2 is also mostly deferred pensioners. It is therefore likely that they also have non-USS pensions which reduce the gap between Group 2 and Group 1.

Table 10 Pension amount (in pounds) thresholds (revalued to 2017) for low, medium and high pensions for USS male and female primary pensioners

	P_1	P_2
Males	20,760	28,845
Females	10,001	19,993

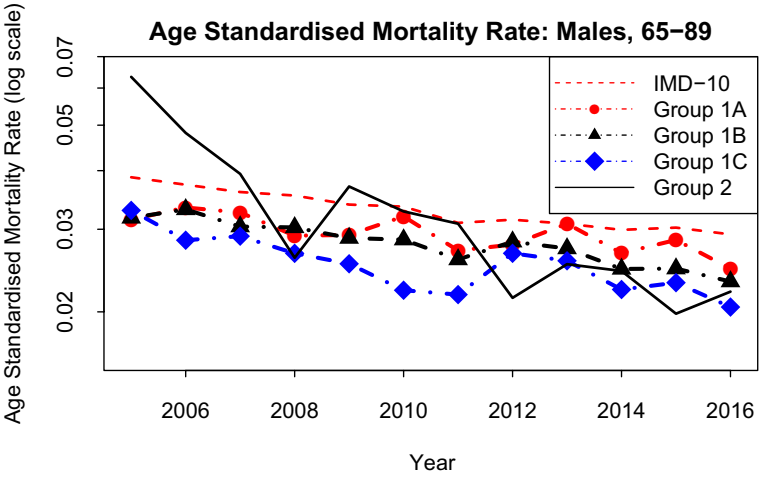

Fig. 7 Age standardised mortality rates based on crude death rates for USS males Groups 1A, 1B, 1C and 2 versus IMD-10 based on the age range 65–89

Figure 8 considers baseline mortality under the fitted CBDX model for males in IMD-10, Groups 1A, 1B and 1C and Group 1 (i.e., the aggregation of Groups 1A,

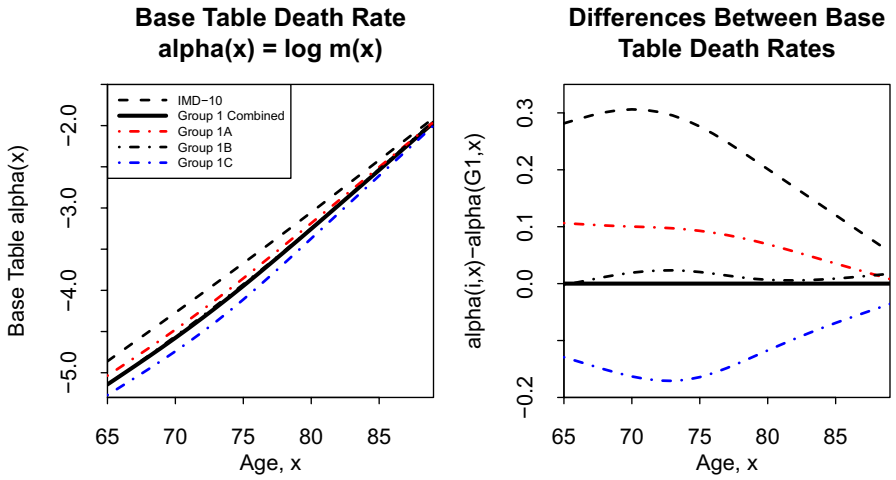

Fig. 8 Left: base mortality, $\alpha(i,x)$ (Eq. 1), for groups i = IMD-10, 1 (Groups 1A, 1B, and 1C aggregated), 1A, 1B and 1C. Right: log of the ratio of base mortality to aggregated group 1 (G1) base mortality: $\alpha(i,x) - \alpha(\text{ G1 },x)$

1B and 1C). The left-hand plot shows the fitted $\alpha(x)$ (Eq. 1) for each group under the assumption that all follow the IMD-10 improvements as in Test 5 (Sect. 5). The right-hand plot shows the difference between the $\alpha(x)$ for each group and the $\alpha(x)$ for Group 1. This shows more clearly the differences between the five groups, reinforcing the empirical differences observed in Fig. 7.

So a key finding from these two figures is that, for males, the pension amount might be predictive of the level of mortality, but only for first-of-the-month retirees. This implies that caution must be exercised when simply using pension amount without also taking the day of retirement into account. For females, the number of lives is sufficiently small that the results are inconclusive.

Our earlier tests 1–5 point to a strong occupational effect. As a graphical diagnostic, Fig. 7 suggests that, additionally, *former* academics (i.e., most of those in Group 2) have just as low mortality rates as those who stay in academia until retirement. Amongst other factors, they will have the same exceptionally high levels of formal education.

7.2 Test 6: explanatory power of pension amount

In this sub-section, we carried out a statistical test—similar to Tests 1 and 2 (in Sect. 5)—of the null hypothesis that Groups 1A, 1B and 1C have the same death rates versus the alternative hypothesis that Groups 1A, 1B and 1C have significantly different death rates. Formally,

- H0: $m(1A,t,x) = m(1B,t,x) = m(1C,t,x)$ for all $(t,\ x)$ (i.e., the groups have the same death rates), versus
- H1: $m(1A,t,x) \neq m(1B,t,x) \neq m(1C,t,x)$ for all $(t,\ x)$ (i.e., the groups have different death rates).

The likelihood ratio test in this case had a p-value of 0.0003 (Table 11), so we reject the null hypothesis and conclude that there are sufficient differences between the three groups that they should be modelled separately.

Finally, the difference between Groups 1A and 1C was compared with the CMI experience analysis of Self Administered Pension Scheme (SAPS) data [11]. Chart 6.1 of CMI [11] offers some insight into the differences in the SAPS data for males between pension bands. Although a direct comparison is difficult, the CMI chart points to much greater differences between the various CMI-SAPS pension amounts than we see in the USS data: potentially the result of greater heterogeneity in the SAPS data which comprise a mixture of industry sectors.

Table 11 Results of Test 6: explanatory power of pension amount for first-of-month primary retirees

	Test statistic, S	95% critical threshold	p-value	Conclusion
Males	728.0	658.1	0.0003	Reject H0

For females, there was no graphical evidence supporting a difference in mortality rates between Groups 1A, 1B and 1C; in part, this is due to small-population noise.

8 Discussion of future short- and long-term mortality rate improvement assumptions

In this section, we discuss how our findings in the previous sections can be used to help set assumptions about future mortality rate improvements for the USS. We have noted that the different IMD groups have experienced different rates of improvement over the last 15 years. However, the principle of coherence—see Li and Lee [22] and Hyndman et al. [19]—suggests that these differences are unlikely to be sustainable in the long term. In other words, it is implausible to assume that mortality rates in the different IMD groups will continue to diverge in the long run. In setting long-term improvement rates, therefore, we should not look at IMD-10 in isolation. Instead, we should focus, on additionally, (a) national mortality trends and (b) international mortality trends. For the same reason, we should assume the same long-term improvement rate for females and males, with potential differences only in the short and medium term.

The framework embedded in CMI [12] requires several inputs including age-dependent short- and long-term improvement rates. The rationale for having a short-term rate allows users to assume that recent improvement rates (high or low) might continue in the short term before reverting over time to some long-term rate.

We do not present here specific recommendations for the short and long-term improvement rates. Instead, we discuss what we see as a set of relevant factors, including those specific to USS that should be considered carefully before these assumptions are set.

8.1 National mortality trends

As a starting point for this discussion, Table 12 presents historical improvement rates using the overlapping periods available for USS, IMD and English national

Table 12 Summary of historical improvement rates over the period 2005–2016 for different groups

2005–2016 improvement rates (p.a.)	Age 65 (%)	Average (%)	Age 89 (%)
USS males	5.5	3.8	2.1
English IMD-10 males	3.1	2.7	2.2
English males	2.8	2.3	1.8
USS females*	3.5	2.6	1.7
English IMD-10 females	2.5	2.5	2.4
English females	2.4	1.9	1.4

Values stated are based on fitted death rates using the CBDX model. The average improvement rate is the mean of the improvement rates at single ages from 65–89 over the period 2005–2016. *USS females: smaller numbers of deaths resulted in much greater volatility in fitted mortality rates

data. We highlight IMD-10 in this table in response to our conclusions in Tests 4 and 5 that USS mortality rate improvements are indistinguishable from IMD-10.

Using English mortality rates from 2001–2016 (i.e., by aggregating the 10 IMD groups) and following the approach of Cairns et al. [4], a point estimate for the drift of $\kappa_1(t)$ (males) in the CBDX model is -0.025 with a 95% confidence interval of $(-0.035, -0.016)$. $\kappa_1(t)$ dictates changes in the level of mortality at all ages, while $\kappa_2(t)$ affects the slope only. Thus, the negative of the drift of $\kappa_1(t)$ can be interpreted as the headline improvement rate: that is, *on the basis of the CBDX model and English national population data for the period 2001–2016,* a point estimate for the headline improvement rate is 2.5% per annum with a 95% credibility interval of (1.6, 3.5)%.[20]

More detailed credibility intervals for selected individual ages are presented in Table 13. This clearly shows that, under the assumptions of the CBDX model, higher ages are predicted to have slower improvement rates than lower ages. If a longer run of data is considered appropriate, then the credibility intervals would probably narrow slightly reflecting the greater number of observations.

In the short run, consistent with the recent past, a projection of IMD-10 mortality might have a higher rate of improvement than the national population if one takes the view that there is a certain amount of persistence in mortality rate improvements in specific subgroups.[21]

Counterbalancing this, short-term mortality rate improvement assumptions might reflect the recent (since around 2011) downturn in UK mortality rate improvements. Quantifying the extent of this downturn in the USS data is very difficult because of small-population noise. As a substitute, we present improvement rates over two time periods (2001–2011 and 2011–2016) for the IMD-10 group in Table 14.

Table 13 Credibility intervals for long-term improvement rates under the CBDX model at different ages based on English males data from 2001–2016

Credibility level (%)	2.5	50	97.5
Age 65 (%)	4.1	3.2	2.3
Age 77 (%)	3.5	2.5	1.6
Age 89 (%)	3.3	1.9	0.4

Table 14 Improvement rates for IMD-10 males over two time periods

Improvement rate per annum	2001–2011	2011–2016
Age 65 (%)	4.5	1.5
Age 77 (%)	4.0	1.2
Age 89 (%)	3.4	0.9

CBDX smooths results across ages, but leaves year-to-year variation

[20] The higher mean of 2.5% compared with Table 12 is the result of using a longer period: 2001–2016. The associated credibility interval is for the central improvement rate only, and does not include the stochastic component of the random walk, so the actual improvement rate over a given time horizon would be more uncertain.

[21] *Persistence* would mean that if one subgroup has experienced, for example, 1% higher improvement rates over the last few years than another subgroup, then this difference might persist in the future, at least in the short term.

Consideration of the improvement rate might also take account of information from other sources, as well as international comparisons (discussed below). For example, cause-of-death analysis or a downturn in health spending might help identify the reasons for the slow down in UK mortality rate improvements. We could then assess whether these factors can be expected to persist to the extent that short- or even long-term improvement rates will be lower than those experienced in the countries that are 'world leading'—in the sense of Oeppen and Vaupel [26] and Kleinow and Vellekoop [20]—for higher-age mortality rate improvements over the last 30 years.

8.2 International mortality trends

It is prudent to consider how patterns of mortality rate improvements vary in other countries to check whether they are in line with those in England. And, if they are out of line, should we alter our assumption for England to ensure that all countries (and hence the USS) have a common long-term improvement rate?

There has been considerable discussion of the slow down since around 2011 in mortality rate improvements in the UK. In Fig. 2, we saw that although different IMD groups have experienced different rates of mortality improvement, all 10 groups have, arguably, experienced a slow down.

On the other hand, compared with other countries, the UK has experienced higher average rates of improvements over a longer timespan. In particular, a comparison with Sweden and Japan is useful. Over the last 30 years, English and Welsh male mortality has improved at the rate of 2.6% per annum (ages 65–89), compared with 2.1% in Sweden and 1.7% in Japan. However, Sweden and Japan historically, *and still do*, have lower rates of mortality than England and Wales at higher ages, but have also experienced more steady rates of improvement over this period.

In Fig. 9, we illustrate how England and Wales (EW)[22] males mortality has fared relative to other developed countries over ages 65–89. We see different patterns of improvement in the six countries. Canada exhibited a similar slow down to EW. But Sweden and Japan, which already had low mortality, have not exhibited a slow down in recent years. This suggests that the EW slow down could be a local and temporary phenomenon. On the other hand, Sweden and Japan might be overdue for a slowdown, e.g., if the factors affecting EW are not the result of local drivers, but apply internationally, but with different lead times. Figure 9 also shows age standardised mortality rates for the English centiles (grey lines from 2001). This allows us to see that the spread between countries is much narrower than the spread between the centiles.

9 Summary of findings and recommendations

In this section, we summarise the findings for the USS from the previous sections:

[22] We have used the larger EW population here as a longer run of mortality data are available for the combined population.

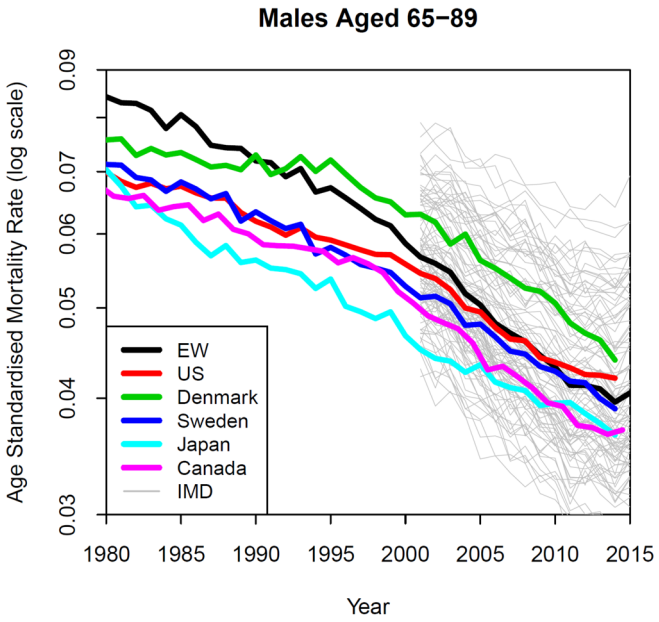

Fig. 9 Age standardised mortality rates for six countries for the age range 65–89. Grey lines (2001–2016) show the equivalent English IMD centiles. Sources: Human Mortality Database, Office for National Statistics, Statistics Canada

- The mortality rates of both male and female USS pensioners were found to be *significantly lower* than those for the least deprived (and longest-lived) of the national IMD deciles (IMD-10). On average across all post-retirement ages, USS male mortality rates are lower than the average of the 1% least deprived areas in England, while for females it is around the level of the 1–2% least deprived. In short, the USS has a distinctively different base mortality table from both the English national and IMD-10 populations and this observation should be useful for the purpose of valuing pension liabilities.
- Mortality rate improvements for USS males and females were found not to be *significantly different* from the corresponding IMD-10 improvement rates. This is very useful when it comes to forecasting, since IMD-10 has the advantage of a greater volume of data (implying less small-population noise) and a longer run of data.
- Many common longevity covariates which in different datasets have significant explanatory power turn out not to do so, or to be so strong, when it comes to USS members. A key example is post code or place of residence which can be associated with both region (e.g., N.E. England, London, etc.) and an IMD decile. We found that neither an individual's IMD decile nor the region where they live has any explanatory power, although there was a significant difference between the north and south of England. The implication of this is that the *very strong homogeneity* of the occupational group to which USS pensioners belong (i.e., university lecturers and senior administrators) overrides any other potential

covariate available in the USS database as a predictor of an individual's mortality.[23]
- There were *significant differences* between the mortality of male pensioners by pension amount, although only for those who retired on the first of the month (mostly persons retiring directly from active service).

Based on the tests carried out in this paper, our recommendations are as follows:

- The USS should use its own base mortality tables for both males and females.
- The short-term future mortality rate improvement for USS pensioners should be the same as that for IMD-10 (using the full experience of IMD-10 from 2001 onwards) and this should be different from (i.e., higher than) the short-term mortality rate improvement for the English national population.
- The principle of coherence suggests that the long-term improvement assumption should be compatible with related populations including IMD-10, the English national population and other similar populations internationally.
- The day of the month on which the member retires and the pension amount might help to predict the level of mortality for male USS pensioners.

CMI [12] also requires, as an input, an assumption about how quickly the improvement rate should shift between the short and long-term assumptions. However, this is beyond the scope of this paper and we leave it for future research.

10 Conclusions

In the process of examining the mortality experience of the UK Universities Superannuation Scheme, we have developed a general framework for analysing the mortality experience of a large portfolio of lives, such as those in any large pension scheme or annuity book.

The framework has the following features and requirements:

- It makes extensive use of a wide variety of graphical diagnostics. These should be seen as an essential part of the process for several reasons:
  - The graphics help to identify distinctive characteristics of the portfolio, such as, here, the day and month of retirement (and, in particular, the first-of-the-month effect), the swings in mean age at retirement, and concentrations of early retirements in specific years or in a narrow band of years.
  - Knowledge of these characteristics can be used to formulate hypotheses about the data or to group the data in sensible ways. For example, we were able to relate increases in retirement or concentrations of early retirements to specific legislative measures or changes in government policy.

[23] We are unable to comment on geodemographic profiling as these data have not been collected by USS.

  - Graphical diagnostics also help to identify errors in the data (e.g., an error in the date of birth).
  - Some of the diagnostics plot the same data in slightly different ways, and, although, in some sense, there is nothing specifically new, the different perspectives can reveal additional nuances that we might not have identified previously.

- The framework makes use of a stochastic mortality model, such as the CBDX model, for the purpose of smoothing the mortality data and for forecasting future mortality rates.
- The framework is designed to work with a variety of larger benchmark datasets that can be helpful, again using graphical diagnostics, for determining the base mortality table and for forecasting future mortality trends both in the short and long term:

  - A key requirement is to find a stable relationship between the mortality rates of the portfolio of lives of interest and those of one or more of these benchmark datasets.[24] We were able to find a stable relationship between the mortality data on USS pensioners and that of English mortality for one of the Index of Multiple Deprivation (IMD) deciles. The principal benefit of this is that data for IMD deciles, and in particular IMD-10, are much more comprehensive than USS in having a greater number of years of observations and a significantly larger number of lives (resulting in substantially lower small-population noise). This can be exploited to help improve predictions about future USS mortality rates.
  - Of particular importance is the ability to map a member of the dataset to a particular socio-economic group or region whose mortality can also be modelled independently. As the previous bullet point makes clear, we were able to do this with the USS dataset by mapping a member's post code to a Lower Super Output Area and thence to the Index of Multiple Deprivation. This index is specific to England. However, there are other geodemographic classification techniques based on the notion of 'linking people to places' that are available, such as:

    - the Neighborhood Socioeconomic Status (NSES) Index in the US (see, e.g., [24]),
    - geodemographic profiling (see, e.g., [28]); this includes commercial organisations that collate multi-source information to produce socio-demographic measures (such as Experian's Mosaic consumer classification system, see, e.g., [30]),
    - customised socio-economic mortality indices (e.g., the Longevity Index for England (LIFE; [7]),[25] and
    - spatio-temporal datasets constructed from cell phone usage that can reveal human behavioural traits related to specific socio-economic characteris-

[24] In other words, there should be no significant changes in the underlying characteristics of the portfolio of interest that would, everything else being equal, cause the mortality rates of the group to drift away from those of an even larger reference population being used for comparison.

[25] The associated LIFE App is available at https://andrewcairns.shinyapps.io/LIFEapp_Version3/.

tics, such as spending patterns in specific locations, which then need to be mapped to individuals in the data set, via post/zip code (see, e.g., [18]).

- Finally, it is important to recognise that different datasets will have their own idiosyncratic features which need to be teased out using the graphical tools of the general framework and then exploited to help set the base mortality table and to improve mortality forecasts. For example, our discovery of the exceptionally strong homogeneity of USS members combined with their high longevity can be fully exploited to improve the robustness of future mortality rate projections. Similarly, our finding that the day of the month on which a member retired had a significant impact on pension amount and life expectancy. Other data sets are likely to have a more heterogeneous membership, but if the socio-economic composition can be reliably determined then it may still be possible to get a good fix on the most appropriate base mortality table and to generate reliable forecasts of future mortality trends both in the short and long term.

Appendix: General characteristics of the pensioners dataset

Figure 10 shows scatterplots of retirement date versus age at retirement for the financially important groups of primary pensioners (i.e., those who have retired in normal health) and widowed spouses (and partners). In the earlier years, for primary pensioners, we see concentrations of retirements at the beginning of October of each year, the day after the end of the previous academic year, whereas, more recently, primary retirements have become more spread out.

There is a trend towards later retirement, probably in response to increasing healthy life expectancy and to changes in both scheme rules (including less favourable early retirement options) and legislation which raised both the normal minimum pension age from 50 to 55 (the Finance Act 2004) and the normal retirement age for both males and females to 66 (the Pensions Acts of 1995 and 2011) and outlawed age discrimination by removing age limits on retirement (the Equality Act 2010). The plots are denser for cohorts born after around 1946, probably reflecting the expansion of the university system in the 1960s and the recruitment of large numbers of young staff at that time.

This increase is perhaps clearer in Fig. 11, where we plot age at retirement versus date of birth. For pensioners who retired before 2005, the data are censored (i.e., records only exist for those who survived to 2005) and survivors are identified by grey dots. The lower left of each plot is, therefore, mostly blank because these early cohorts retired long before 2005 and mostly also died before 2005, so they are not included in the USS pensioners dataset. The upper right part of the plot is blank, as members have not yet attained the given age by 1 January 2017. The prominent diagonals correspond to retirements at the beginning of October each year.

In Fig. 11, we can identify the following features:

- greater numbers of retirees in cohorts born after around 1946;

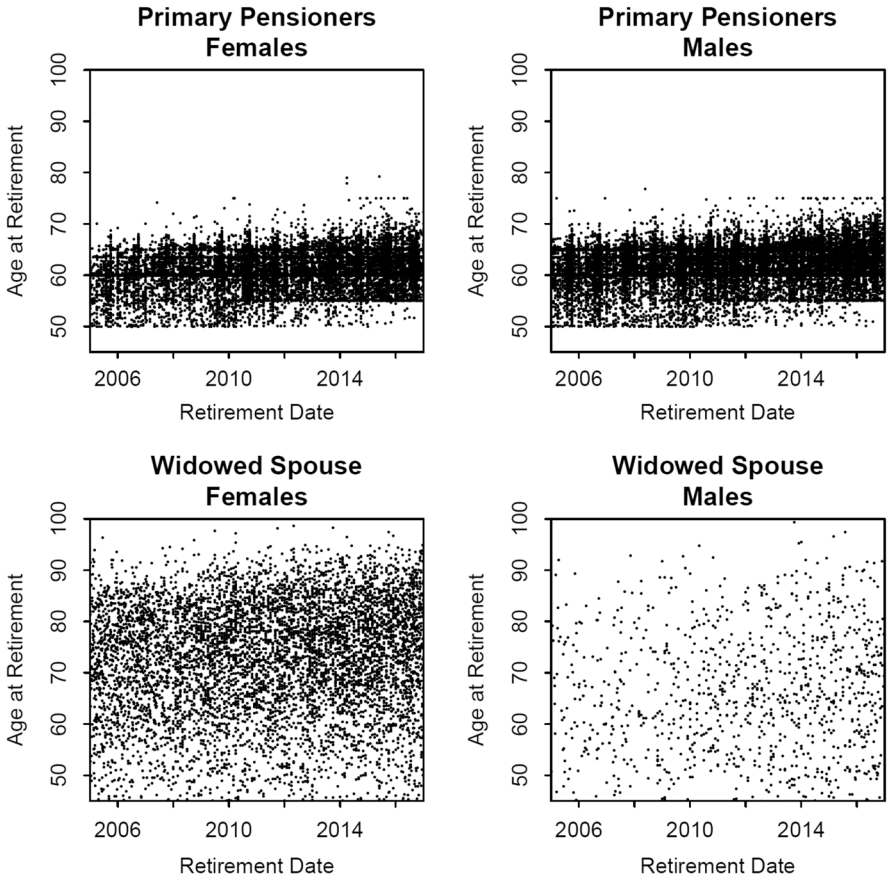

Fig. 10 Retirement patterns for males and females split by pensioner type: date of retirement versus age at retirement. Top: primary pensioners. Bottom: widowed spouse's age at the date of commencement of the spouse's pension; the retirement date is the commencement date of the widowed spouse's pension

- generally heavier rates of retirement around 1982–1984, 1988–1990, 1997 and 2010 (diagonal lines labelled DOR) corresponding to periods of significant reductions in staffing through early retirement and redundancy programmes. In July 1981, as part of the Thatcher government's austerity programme to reduce the national debt, the University Grants Committee reduced funding to UK universities over the following 3 years by 17% [17], while the Education Reform Act 1988 allowed universities to dismiss academics in permanent posts on the ground of redundancy, whereas previously they could only be dismissed for 'good cause', such as an inability or unwillingness to perform their duties, or gross misconduct. The increased retirements in 1997 and 2010 were the consequence of the introduction of market-driven forces in the university system, first, as a result of the Dearing Report [14] which recommended the abolition of student maintenance grants and the introduction of student

fees, and, second, following the Browne Report [3] which recommended the removal of the £3290 p.a. cap on fees, allowing universities to raise their fees to £9000, which reduced student demand for a range of courses.

Figure 12 illustrates the changing patterns of retirement in a different way by plotting the mean age at retirement for primary pensioners. For pensioners who retired before 2005, the data are left-censored: only those who survived to the start of 2005 are included in the data. This introduces a bias in the (conditional) mean age at retirement which will be lower than the true or unconditional mean (younger retirees from a given cohort are more likely to survive to 2005). This bias will gradually increase as we move towards the left of the plot. In spite of this bias, there are three noteworthy features in the data:

- the mean retirement age was initially relatively high (at around 63), with the male age being only slightly higher than the female age;
- there were significant falls in the mean age around 1983 and 1989 corresponding to phases of increased early retirement as part of redundancy programmes (see the remarks above on Fig. 11);
- there has been a steady increase in the mean age from 1997 onwards, with the male retirement age being around 2 years higher than the female age.

By the end of the sample period, the male retirement age was back to where it was at the start of the sample period, while the female age was still a little lower. It is hard to be sure why the retirement age was so high between 1975–1980 compared with the rest of the period, but we offer the following possible explanations. USS was still a relatively young scheme so pensions had not built up to adequate levels causing

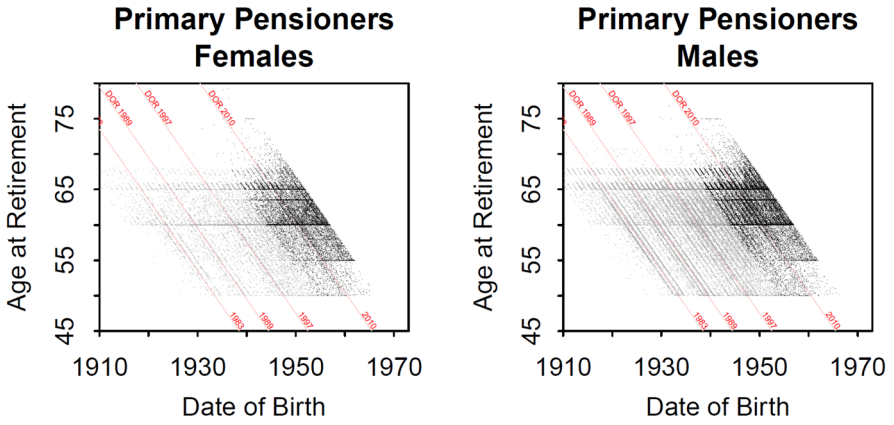

Fig. 11 Retirement patterns for males and females split by pensioner type: date of birth versus age at retirement. Each dot represents one individual. Top: primary pensioners. Bottom: Widowed spouses. Grey dots: survivors to 2005 with a date of retirement prior to 2005. Black dots: all retirees retiring from 2005 onwards. Additional diagonal lines highlight retirements (DOR: Date of Retirement) in 1983, 1989, 1997 and 2010

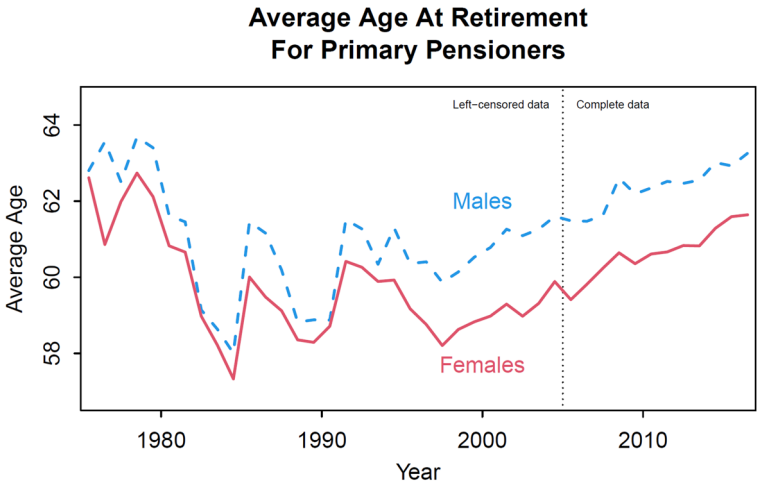

Fig. 12 Mean age at retirement by calendar year for primary pensioners and males and females. Data for retirements prior to 2005 are left censored: mean age at retirement is conditional on survival to the start of 2005

people to have to work longer, and there was also a much stronger sense of having a nation-wide common retirement age (65 for men and 60 for women) and people expected to work until these ages. It is also likely that the first early retirement programme of 1982–1984 opened academics eyes to early retirement being a new possibility for those tired of their jobs. Further, USS had a big surplus in the 1980s and 1990s allowing universities to offer generous early retirement packages.

The changing patterns in age at retirement are illustrated further in Figs. 13 and 14, for females and males, respectively. For retirements in years 1981–2004, the input data are subject to left censoring: that is, only those who have survived to 2005 are included. In 1981 and 1985, retirements are concentrated around ages 60 and 65. In contrast, retirements in 1982 to 1984 are more spread out with a substantial increase in the total number of retirements. There is a similar pattern (especially for males) over the period 1988–1990 (spread similar to the 1982–1984 histograms) contrasted with 1987 and 1991 (concentrated around 60 and 65). These changes in the distribution of ages at retirement are reflected in the mean age at retirement plotted in Fig. 12. More generally, they show how patterns of retirement have changed over time. Although the normal retirement age is 65 for both males and females throughout the period covered by the data (it was not raised to 66 until 2020), the peak retirement ages are still 60 (females) and 65 (males), but these are less prominent than they used to be. The shift to later retirement is clear.

Figure 15 reveals that average pensions for males are nearly double those of females, with relatively little change between the two 6-year periods. Figure 15 also shows that the shape of the distribution has changed quite considerably for males from one that was clearly bimodal—reflecting the different outcomes for short stayers and long stayers—to one that is more skewed towards small pensions.

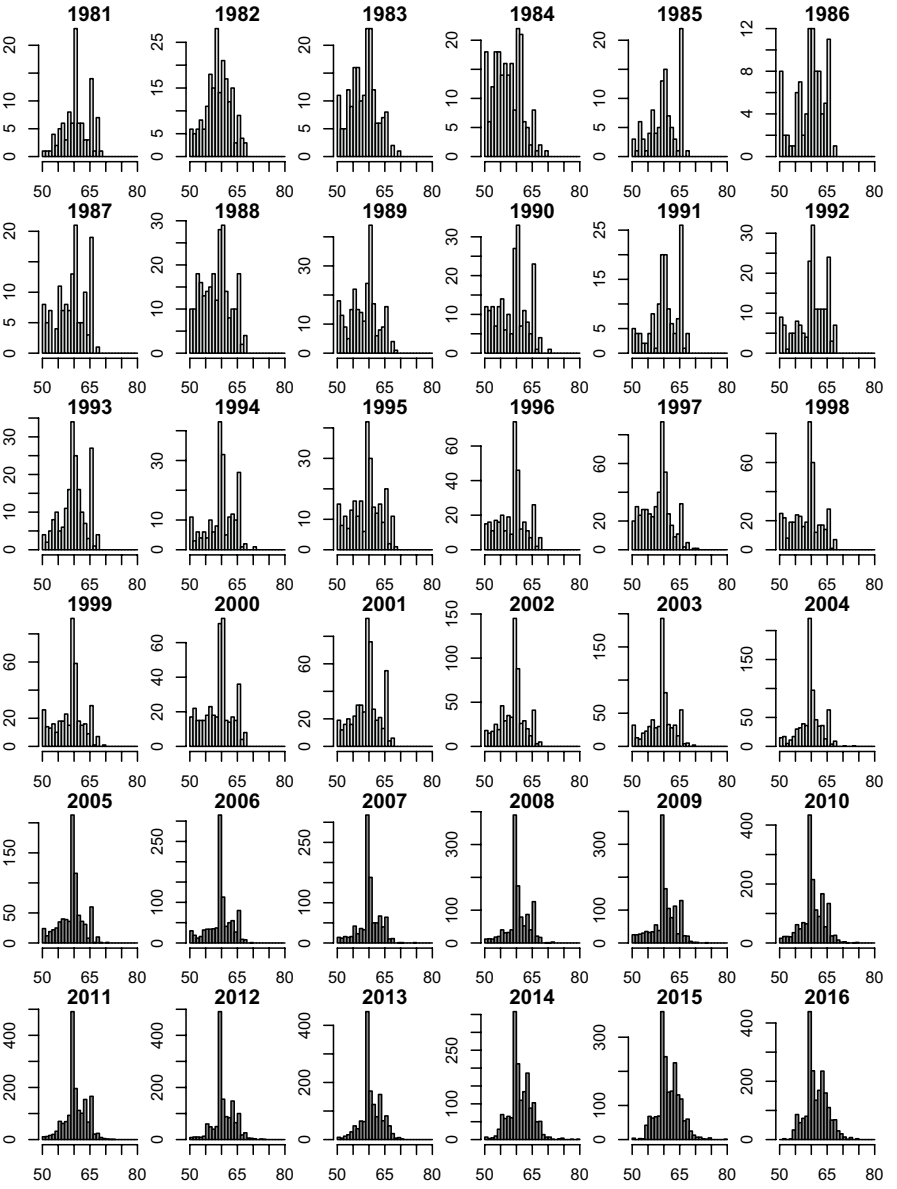

Fig. 13 Histograms of age at retirement for female primary pensioners. Light grey bars (1981–2004) record age at retirement only for those who have survived to 2005. Dark bars (2005–2016) display complete data

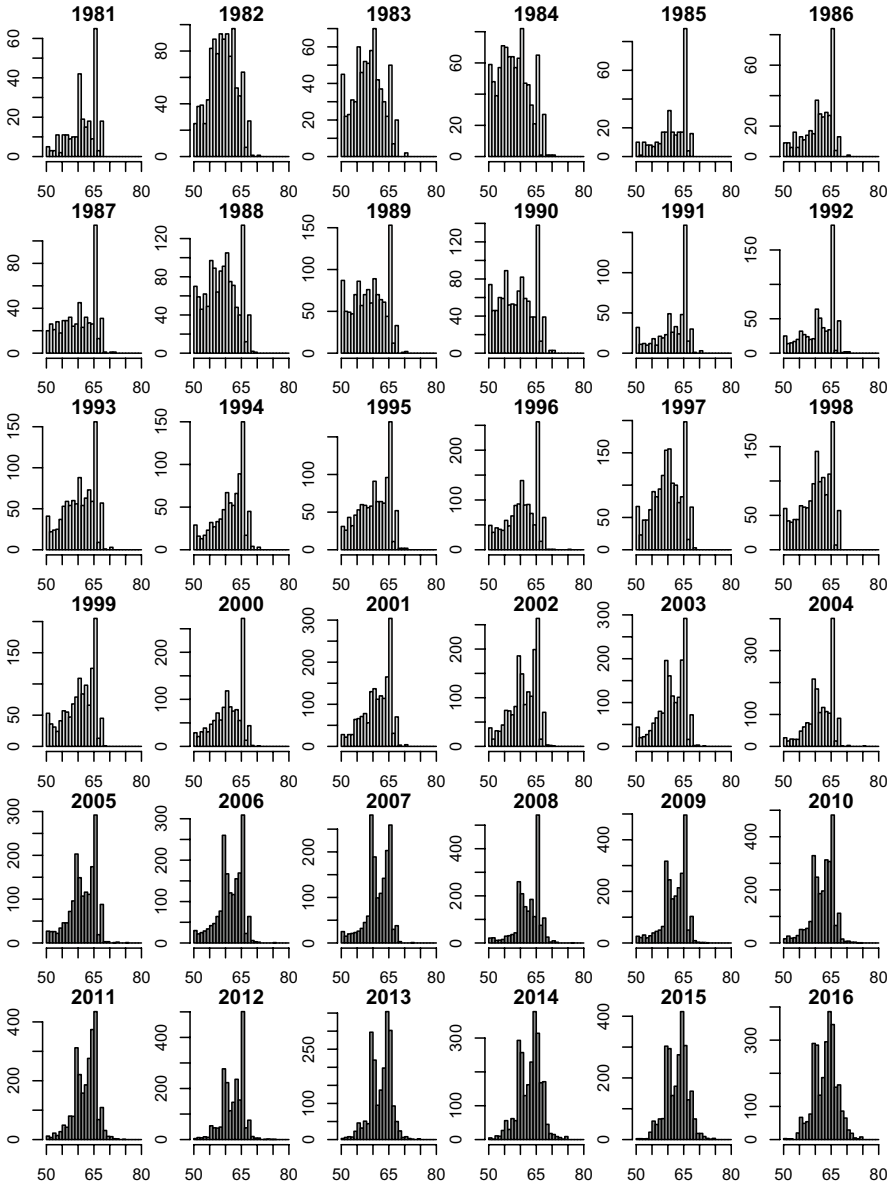

Fig. 14 Histograms of age at retirement for male primary pensioners. Light grey bars (1981–2004) record age at retirement only for those who have survived to 2005. Dark bars (2005–2016) display complete data

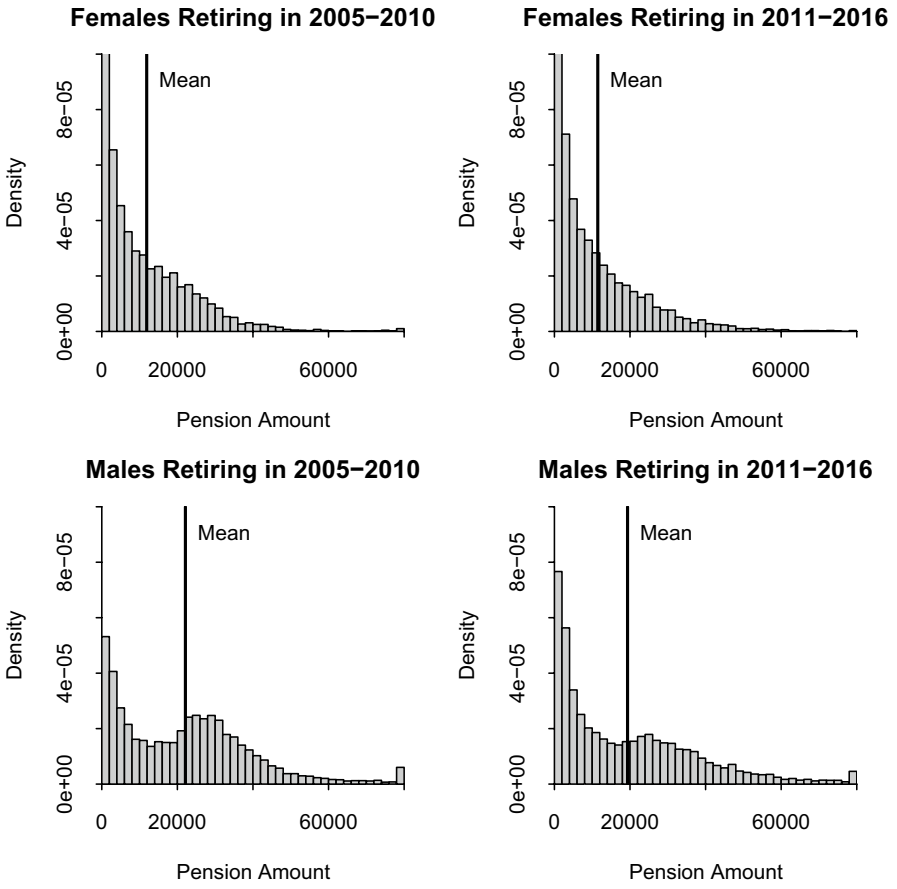

Fig. 15 Histograms of pension amounts (revalued to 1 April 2017) for specific groups of pensioners. In each plot, the right-hand bar includes all pensions in excess of £80,000 per annum. Vertical line: mean pension amount

Acknowledgements AJGC acknowledges funding from the Actuarial Research Centre of the Institute and Faculty of Actuaries.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Bradley G, Hanlon S (2017) The Scotland effect: is there one and why does it matter? Presentation, Institute and Faculty of Actuaries Life Conference, Birmingham
2. Brønnum-Hansen H, Baadsgaard M (2012) Widening social inequality in life expectancy in Denmark. A register-based study on social composition and mortality trends for the Danish population. BMC Public Health 12:994
3. Browne Report (2010) Securing a sustainable future for higher education: an independent review of higher education funding and student finance. Department for Business, Innovation and Skills. https://www.gov.uk/government/publications/the-browne-report-higher-education-funding-and-student-finance
4. Cairns AJG, Blake D, Dowd K (2006) A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. J Risk Insur 73:687–718
5. Cairns AJG, Blake D, Dowd K, Kessler AR (2016) Phantoms never die: living with unreliable population data. J R Stat Soc Ser A 179:975–1005
6. Cairns AJG, Kallestrup-Lamb M, Rosenskjold CPT, Blake D, Dowd K (2019) Modelling socio-economic differences in mortality using a new affluence index. ASTIN Bull 49:555–590
7. Cairns AJG, Wen J, Kleinow T (2021) Drivers of mortality: risk factors and inequality. Working paper, Heriot-Watt University. www.macs.hw.ac.uk/~andrewc/ARCresources/
8. Chen L, Cairns AJG, Kleinow T (2017) Small population bias and sampling effects in stochastic mortality modelling. Eur Actuar J 7:193–230
9. CMI (2014) Graduations of the CMI SAPS 2004–2011 mortality experience based on data collected by 30 June 2012: Final "S2" Series of Mortality Tables. Continuous Mortality Investigation, Working Paper 71
10. CMI (2015) Mortality experience of pensioners by industry classification, 2006–2013. Continuous Mortality Investigation, Working Paper 86
11. CMI (2018) Mortality experience of pensioners for the period 2009 to 2016. Continuous Mortality Investigation, Working Paper 104
12. CMI (2018) CMI Mortality Projections Model: CMI_2017 Continuous Mortality Investigation, Working Paper 105
13. Coughlan GD (2015) Longevity: it's academic, presentation at l11: the eleventh international longevity risk and capital markets solutions conference, Lyons. www.cass.city.ac.uk/__data/assets/pdf_file/0019/293311/COUGHLAN-Guy-L11.pdf
14. Dearing Report (1997) Higher Education the Learning Society: Report of the National Committee of Inquiry into Higher Education. Her Majesty's Stationery Office, London
15. Department of Communities and Local Government (2015) English indices of multiple deprivation. www.gov.uk/government/statistics/english-indices-of-deprivation-2015
16. Dowd K, Cairns AJG, Blake D (2020) CBDX: a workhorse mortality model from the Cairns–Blake–Dowd family. Ann Actuar Sci 14:445–460
17. Hayward J (1986) Responses to contraction: the University of Hull, 1979–1984. Minerva 24:74–97
18. Hong L, Frias-Martinez E, Frias-Martinez V (2016) Topic models to infer socio-economic maps. In: Proceedings of the thirtieth AAAI conference on artificial intelligence (AAAI-16). Association for the Advancement of Artificial Intelligence, pp 3835–3841
19. Hyndman R, Booth H, Yasmeen F (2013) Coherent mortality forecasting: the product-ratio method with functional time series models. Demography 50:261–283
20. Kleinow T, Vellekoop M (2018) Minimum reversion in mortality models for multiple populations. Presentation at the fourteenth international longevity risk and capital markets solutions conference, Amsterdam
21. Lee RD, Carter LR (1992) Modeling and forecasting U.S. mortality. J Am Stat Assoc 87:659–675
22. Li N, Lee R (2005) Coherent mortality forecasts for a group of populations: an extension of the Lee–Carter method. Demography 42:575–594
23. Mackenbach JP, Bos V, Andersen O, Cardano M, Costa G, Harding S, Reid A, Hemström Ö, Valkonen T, Kunst AE (2003) Widening socioeconomic inequalities in mortality in six Western European countries. Int J Epidemiol 32:830–837
24. Miles J, Weden M, Lavery D, Escarce J, Cagney K, Shih R (2016) Constructing a time-invariant measure of the socio-economic status of U.S. Census Tracts. J Urban Health 93:213–232

25. Office for National Statistics (2020) Changing trends in mortality by national indices of deprivation, England and Wales: 2001 to 2018. www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/changingtrendsinmortalitybynationalindicesofdeprivationenglandandwales/2001to2018
26. Oeppen J, Vaupel J (2002) Broken limits to life expectancy. Science 296:1029–1030
27. Plat R (2009) On stochastic mortality modelling. Insur Math Econ 45:393–404
28. Richards S (2008) Applying survival models to pensioner mortality data. Br Actuar J 14:257–303
29. Waldron H (2013) Mortality differentials by lifetime earnings decile: implications for evaluations of proposed social security law changes. Soc Secur Bull 73(1):1–37
30. Wami WM, Dundas R, Molaodi OR, Tranter M, Leyland AH, Katikiredd SV (2019) Assessing the potential utility of commercial 'big data' for health research: enhancing small-area deprivation measures with Experian™ Mosaic groups. Health Place 57:238–246
31. Wen J, Kleinow T, Cairns AJG (2020) Trends in Canadian mortality by pension level: evidence from the CPP and QPP. North Am Actuar J 24:533–561
32. Wen J, Cairns AJG, Kleinow T (2021) Fitting multi-population mortality models to socio-economic groups. Ann Actuar Sci 15:144–172

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Andrew J. G. Cairns[1] · David Blake[2] · Kevin Dowd[3] · Guy D. Coughlan[4] · Owen Jones[5] · Jeffrey Rowney[4]

[1] Department of Actuarial Mathematics and Statistics, School of Mathematical Sciences, Heriot-Watt University and Maxwell Institute, Edinburgh EH14 4AS, UK

[2] Pensions Institute, City University of London, London, UK

[3] Durham University, Durham, UK

[4] Universities Superannuation Scheme, London, UK

[5] Universities Superannuation Scheme, Liverpool, UK