Trends in Canadian Mortality By Pension Level: Evidence From the CPP and QPP

Jie Wen[a], Torsten Kleinow[a] and Andrew J.G. Cairns[ab]

This version: October 8, 2019

Abstract

This paper looks at the mortality of Canadian pensioners subdivided by pension level using data from the Canada Pension Plan (CPP) and Québec Pension Plan (QPP). Differing pension levels (11 groups) are found to give rise to significant levels of mortality inequality at age 65, with a declining inequality gap as cohorts get older. We also find that levels of inequality have increased slightly over time, and that the QPP pensioners exhibit greater levels of inequality than CPP. Additionally, we find strong, but indirect, evidence amongst the lowest pension groups for a *healthy-immigrant effect*.

We fit a range of multi-population stochastic mortality models to the CPP and QPP data and find that the Common Age Effect model satisfies best a range of quantitative and qualitative criteria. The model allows us to distill further detail in the underlying mortality data as well as provide a coherent basis for forecasting mortality and assessing uncertainty in these forecasts.

Lastly, we use clustering methods to consider how significant the differences are between the 11 groups.

Keywords: Canadian pensioner mortality; CPP; QPP; pension level; healthy immigrant effect; multi-population stochastic mortality model; clustering.

[a]Maxwell Institute for Mathematical Sciences, and Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK. Internet: `www.macs.hw.ac.uk/~andrewc/ARCresources/`

[b]Corresponding author: A.J.G.Cairns@hw.ac.uk

1 Introduction

A significant element of the work of life and pensions actuaries is the assessment of the historical and future mortality of a portfolio of lives. This paper discusses Canadian mortality with an emphasis on socio-economic differences in the level of mortality and mortality improvement rates. The core objectives are:

- to establish
  - how wide are the differences in mortality between different socio-economic groups?
  - how much do these differences vary with age?
  - have these mortality inequalities increased over time?
- to identify potential reasons for the patterns that we observe in the data;
- to analyse the data within a model-based framework that can be used in future work on the projection of future levels of mortality and assess how much uncertainty there is around central forecasts.

Our discussion of socio-economic differences in the Canadian population is based on data provided by the Canada Pension Plan (CPP) and the Québec Pension Plan (QPP), with data grouped by pension level, and builds on earlier work by Adam (2012a,b, 2016).

Figure 1 shows crude death rates for males and females using data from Statistics Canada over the period 1990 to 2016. The y-axes for each of the six sub-plots are, of course, different, but the ratio from top to bottom is the same in each case, so that a flatter plot indicates lower rates of mortality improvements. Thus, males aged 65-69 have seen the greatest percentage improvements in mortality over the last 25 years, while females aged 45-49 and 85-89 have seen the smallest improvements. In some age groups, we can see a distinct slowdown in mortality improvements after 2011 (e.g. males and females aged 65-69).

As a group of plots, Figure 1 reveals:

- different rates of improvement over different time periods;
- different rates of improvement at different ages;
- different patterns of improvement for males and females of the same age;
- volatility on a year to year basis around a varying underlying trend.

These observations point to uncertainty in both the short and long term when forecasting future mortality rates, as well as rates of improvement at different ages leading to the use of stochastic models in Section 5.

The remainder of the paper is structured as follows. Section 2 begins the focus on subpopulation mortality. Here, we begin our detailed analysis (initially empirical) of the CPP and QPP data subdivided by pension level with a focus on age standardised mortality rates (ASMR's). In order to understand better the CPP/QPP ASMR plots, we then look in more detail in Section 3 at the numbers of individuals attaining different levels of pension and how this has changed over time. Additionally, we consider how the cohort-dependent proportions in each group might have an impact on ASMR's. Section 4 then quantifies how migration in middle age might have an impact on the composition of specific groups by pension level, and we introduce the *healthy immigrant effect*: a factor that plays an important role in the subsequent modelling sections. Stochastic models are considered in Section 5, where we consider a range of potential multi-population stochastic mortality models before settling on a specific model (a special case of the Common Age Effect (CAE) Model of Kleinow, 2015) that best meets a list of desirable criteria. A detailed discussion of the modelling results follows with links to the earlier empirical sections of the paper. We then consider clustering (aggregating small groups with similar levels of mortality) in Section 6 as a further means of reducing noise in the underlying data. Noise reduction then allows us to make additional insights about specific subpopulations. Section 7 concludes.

2 Socio-Economic Differences in Mortality

2.1 Introduction

It is well known that socio-economic status has a significant impact on mortality, with abundant evidence from many countries where data are available in sufficient detail. Typical measures of socio-economic status include affluence (e.g. Chetty et al., 2016 [US], Longevity Science Panel, 2018 [UK], Cairns et al., 2019 [Denmark]) and education (e.g. Mackenbach et al., 2003, 2016).

Here, we explore the extent of mortality inequalities in Canada and investigate how these have evolved over time. Specifically, we consider mortality data for the Canada Pension Plan (CPP) and the Québec Pension Plan (QPP). QPP covers individuals resident in the province of Québec, while CPP covers all other provinces in Canada. A key advantage is that, in combination, CPP and QPP cover almost the entire population of Canada[1], allowing us to compare the pensioners data with national and provincial mortality data.

[1] In particular, CPP and QPP cover Canadians who have participated in the workforce.

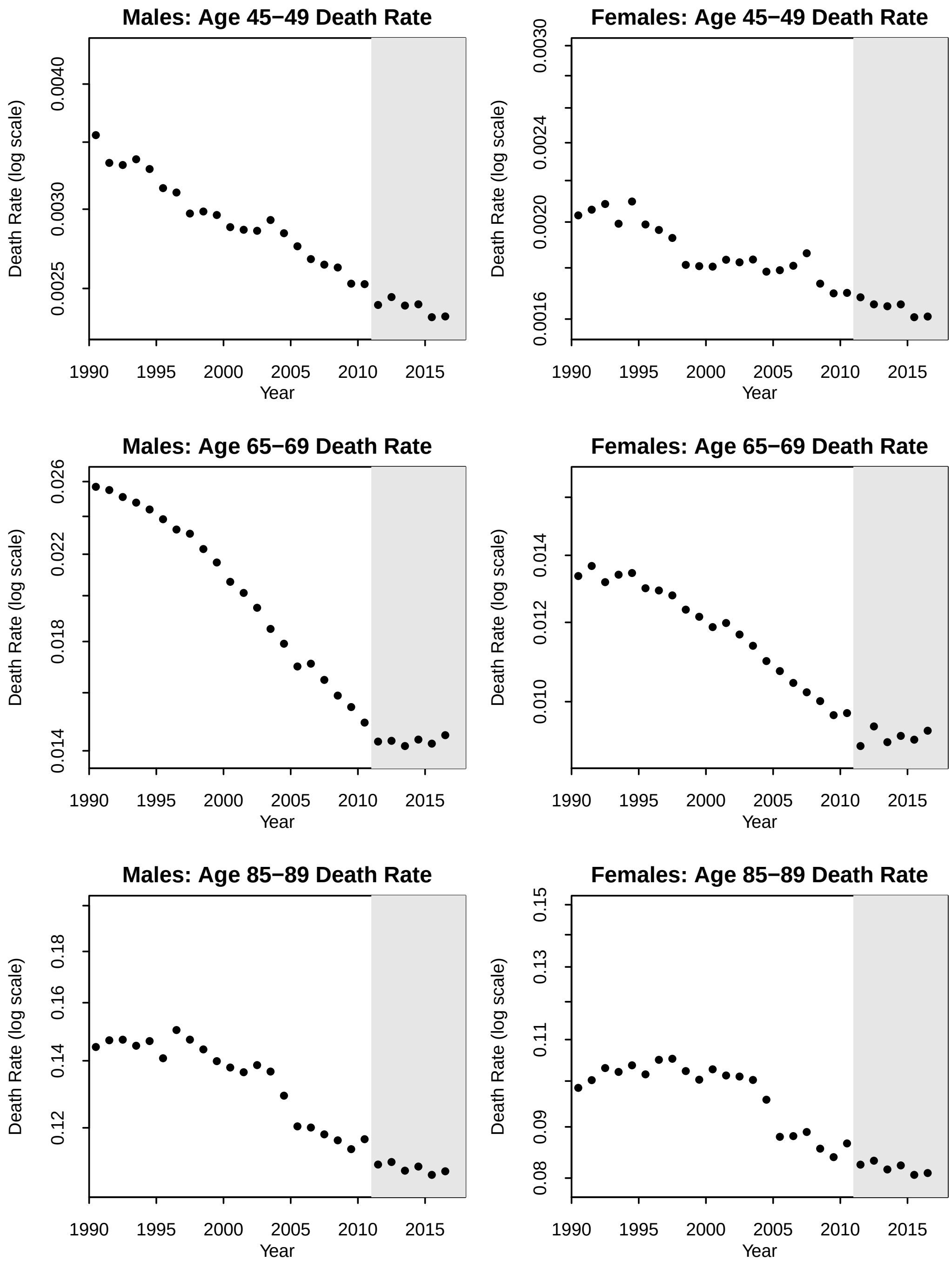

Figure 1: Historical death rates (mid-year to mid-year) for Canadian males and females age groups 45-49, 65-69 and 85-89 from mid-1990 to mid-2016. Source: Statistics Canada, Tables 17-10-0005-01 and 17-10-0006-01. Grey shading highlights the slowdown period after 2011 at some ages.

2.2 CPP and QPP data

Data were provided in an aggregated form for pensioners only (that is, there were no data for pre-retirement members of CPP and QPP).[2] Individual members were grouped according to their pension level expressed as a proportion of the maximum pension achievable by their cohort as follows:

- Group 1: Pension level = $0 - 9\%$ of the cohort maximum pension
- Group 2: Pension level = $10 - 19\%$ of the cohort maximum pension
- $\vdots$
- Group 10: Pension level = $90 - 99\%$ of the cohort maximum pension
- Group 11: Pension level = cohort maximum pension.

Additionally, for the CPP data only, the data exclude (a) individuals who had disability benefits converted into CPP pensions at retirement, and (b) individuals with a pre-existing survivor's pension at the time of retirement and who, as a result, have a pension level in excess of 100%.[3]

Death counts and central exposures[4] in the dataset are denoted by $D(g, i, t, x)$ and $E(g, i, t, x)$ respectively where

- g = gender
- i = pension group
- t = calendar year
- x = age last birthday (at date of death).

The corresponding crude age-specific death rate is then

$$m(g, i, t, x) = \frac{D(g, i, t, x)}{E(g, i, t, x)}.$$

[2]Further details on an earlier CPP/QPP data extract can be found in Section III of Adam (2012a). Adam also discusses the advantages of using CPP/QPP data compared to data from private pension plans and insurers.

[3]It might be possible to make the same exclusions in the QPP data but it was not considered to be necessary for QPP.

[4]Central exposures are the total years of exposure during year t of persons in group (g, i) aged x last birthday. Equivalently, it represents the average number of people during calendar year t aged x last birthday.

The analysis here builds on earlier work by Adam (2012) where groupings are again by pension level as a percentage of the maximum by cohort. Adam initially subdivides into 21 bands (in 5% increments) but then reduces this to three (0-35%, 35-95% and 95-100%) for the purposes of further analysis with an emphasis on the relevance of specific CPP/QPP groups as a proxy for modelling mortality in private pension plans. Here we persist for longer with 11 groups, giving further insights into mortality differentials and time trends.

2.3 Contributions and the cohort maximum pension

During the accumulation phase as a member of CPP/QPP, individuals contribute a defined proportion of their earnings up to the Year's Maximum Pensionable Earnings (YMPE).[5] The YMPE increases each year in line with national average weekly wages, salaries and other earnings, and lies close to average earnings. As a result of the latter, a large proportion of the active membership of CPP/QPP will be contributing at the maximum rate in any given year. However, to achieve the maximum pension, an individual must currently have contributed at the maximum rate in 83% of the eligible working years by cohort,[6] which is much more challenging than contributing the maximum in any single year.[7]

The data cover all pensioners over the period 1967 (CPP) or 1968 (QPP) to 2015. The early years of CPP and QPP did not permit retirement before age 65. QPP allowed early retirement from age 60 from 1984 and CPP from 1987 with a corresponding impact on exposures between ages 60 and 64 from these dates. Late retirement is also permitted, to allow members to accrue additional years (helpful for immigrants) or boost their best 83%. Currently, late retirement is permitted up to age 70 in order to improve the pension amount.[8] Prior to 1989 even later retirement was also possible. Late retirements can be detected in all 11 groups, but it is most obvious in Group 1.[9]

[5]The YMPE was $55,900 in 2018.

[6]83% applies to retirees from 2014 onwards; 84% in 2012 and 2013; 85% up to 2011.

[7] For example, individuals retiring in 2018 were eligible to contribute to CPP from ages 18 to 64 (47 years). So they must have contributed at the maximum rate (i.e. earning above the YMPE) in 39 out of the last 47 years. More generally, the two plans started in 1966 with eligible contributions from the same year, so, for example, an individual retiring at age 65 at the end of 1985 would have had to contribute at the maximum for 17 (85%) out of the last 20 years (1966 to 1985) of their working life: that is, ages 45 to 64. For females the number of eligible years can be reduced from 47 years under the 'child-rearing' provisions (e.g. a female retiring in 2018 with 7 years approved under the child-rearing provision, would have her pension calculated using the best 33 years (83% of $(47-7)=40$) rather than the best 39, and the pension scaled appropriately).

[8]Technically, retirement after 70 is also possible, but there is no financial benefit to taking retirement after age 70.

[9]A late retirement factor is applied for late retirement. Group allocations for late retirees are made by comparing their pension with the age-65 maximum pension scaled up by the late retirement factor. Late retirements can be detected indirectly in the data by comparing decrements

Bearing in mind the maturing nature of CPP and QPP, we used data from 1991-2015 and ages 65 to 89 in our modelling work: a compromise between maturity on the one hand and volume of data on the other.[10] As a related point, the data indicate that persons born before 1895 or 1896 were not eligible to receive a pension from CPP or QPP with resulting zero exposures for these cohorts. However, this would only have an impact on the analysis if we used earlier years than 1991 or higher ages than 89.

2.4 Age standardised mortality rates (ASMR)

The ASMR is a year-t-specific, weighted average of the crude death rates that can be defined over a defined age range $x_0, \ldots, x_1$

$$ASMR(t) = \frac{\sum_{x_0}^{x_1} \hat{m}(t,x)\tilde{E}(x)}{\sum_{x_0}^{x_1} \tilde{E}(x)}$$

where $\hat{m}(t,x)$ is the crude age-specific death rate in year t at age x, and $\tilde{E}(x)$ represents the "standard" exposure at age x (throughout this paper, we use the European Standard Population, 2013; see Eurostat, 2013).[11]

Age standardised mortality rates for ages 65 to 89 based on crude age specific death rates are plotted in Figure 2.[12] A number of observations can be made:

- The broad trends are similar to that for Canada as a whole.
- Although calculation of the ASMR dampens the impact of sampling variation, we can still see that smaller groups (e.g. males Group 1 or females Group 11) produce more volatile ASMR plots compared to larger groups (e.g. males Group 11).
- QPP ASMR's are mostly slightly above those for CPP.
- Significant inequalities are evident between the 11 groups. In particular, the ASMR's for QPP Group 1 males being well over 50% higher than those for Group 11.

in the exposures by cohort after age 65. If the decrement is less than the number of deaths (or even negative) then the difference will equal approximately the number of late retirements.

[10] We limit our investigations to age 65 and above, as the reasons for taking early retirement are varied and can be connected to health.

[11] The use of alternative standard populations might push the ASMR's up or down, but the patterns of improvements that we observe and the relationships between different populations would be largely unaltered.

[12] In the first few years, some ASMR's are missing for high pension QPP groups. This is a result of zero exposures at high ages, meaning death rates required in the calculation of the ASMR are not available.

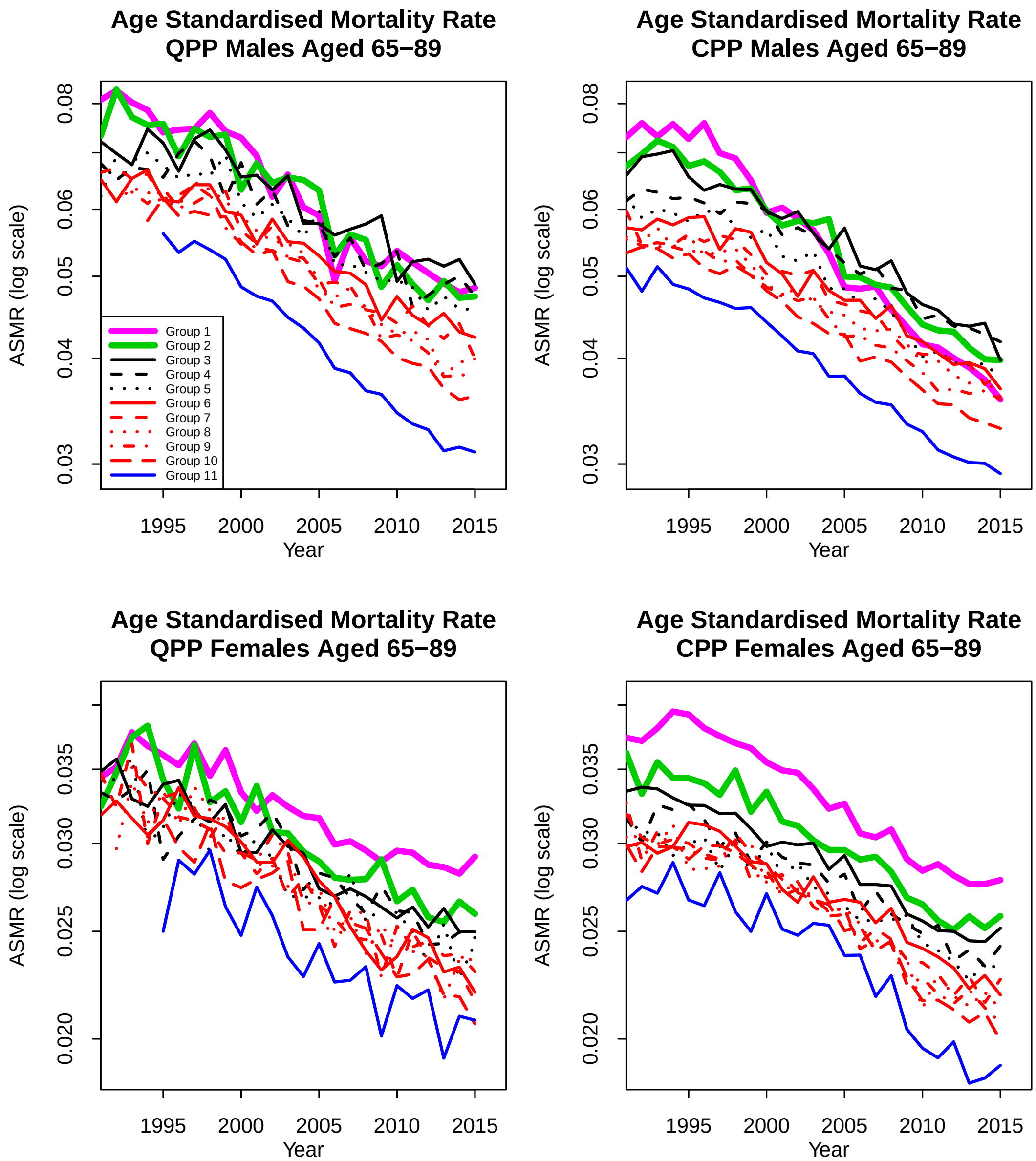

Figure 2: ASMR's for QPP and CPP males and females based on crude age specific death rates for ages 65 to 89. Each plot shows the ASMR for groups 1 (low pension level) to 11 (maximum pension).

- In most years Group 11 stands well below the other groups.
- For females, Group 1 also stands clear of the other Groups.
- Mostly, the ASMR's are ranked approximately in line with the group ordering: high mortality for Group 1 through to low mortality for Group 11. However, in terms of rankings, the data reveal one anomaly that needs some further investigation and discussion:
  - CPP males Groups 1 and 2: these start high, as one would expect, but then gradually drift down and cross over several mid-ranking groups. This is somewhat different from QPP males Groups 1 and 2, which are, first, more noisy but, second, stay about the same levels as Groups 3 to 5.

There are a variety of reasons why Group 11 stands clear of the others. One reason is that within Group 11 there will be a potentially high degree of heterogeneity: some individuals consistently just above the threshold, others much more wealthy; and a mixture of occupation groups. Another reason that we now discuss is *conscientiousness.*

2.5 Conscientiousness

Conscientiousness is one of the five major character traits in the field of psychology. A conscientious individual will: wish to do their work or duty well and thoroughly; they will be careful, hard working, diligent, dedicated and accurate in both their working and personal lives.

We can then *conjecture* that there will be a positive correlation between conscientiousness as a trait and sustained success in employment. In the Canadian context, working hard through one's lifetime and diligence might mean that a conscientious individual attains earnings above the YMPE in a greater number of years than a non-conscientious individual.[13] In particular, amongst, say, second-quartile earners (i.e. just above the YMPE), conscientious individuals are more likely to attain earnings above the YMPE in at least 83% of their working years (see Footnote 7).

Conscientiousness is important because it is the character trait that is most strongly correlated with life expectancy (see, for example, Kern and Friedman, 2008, and

[13] As an example, Egan et al. (2017) provide evidence that conscientious individuals will experience less unemployment in their working lifetimes. In the Canadian context, each period of unemployment makes it less likely that the individual's earnings will exceed the YMPE in a given year.

Deary, Weiss and Batty, 2010).[14] The conjecture that Group 11 in both the CPP and QPP data might contain a greater proportion of conscientious individuals compared to Group 10 would then, in part, explain why Group 11 has significantly lower mortality.

2.6 Impact of group size

In interpreting Figure 2 we also need to be mindful that the proportions of each cohort in each group are changing over time. We discuss this further in Section 3. But, here, we can remark, by way of example, that if Group 1 was shrinking over time, then that might have an impact on mortality rates[15] that interferes with other changes in the level of mortality.

2.7 The slowdown in mortality improvements

We can also look at Figure 2 to investigate if the slowdown observed at the national level affects all groups or some subset. In fact, noise in the ASMR's makes it quite difficult to establish if any of the individual groups has experienced a slowdown, and, certainly, there is no consistent relationship between groups. Group sizes are also changing, making identification of a slowdown at the group level potentially trickier still.

3 Cohort Sizes

We now consider the relative sizes of each group and how these have changed over time. The proportions in each group by cohort are defined as

$$P(g,i,t,x) = E(g,i,t,x) / \sum_j E(g,j,t,x).$$

We focus our discussion on the normal retirement age $x = 65$, but the picture is broadly similar if, say, we used age 70.[16] [17]

[14]For example, conscientious individuals are more likely to: adhere to a healthy diet; visit the doctor early when they develop symptoms related to ill health; and follow doctor's orders when given a diagnosis.

[15]Each group still contains a degree of heterogeneity. Everything else being equal, if Group 1 grows over time, then the average level of deprivation will be reduced with a corresponding lowering of average Group 1 mortality.

[16]Proportions are slightly different at age 70 due to: differential death rates between ages 65 and 70 by group; late retirements between 65 and 70.

[17]Proportions by pension level are also illustrated in a different way in Appendix A of Adam (2012a).

Figures 3 and 4 illustrate in a heat map format for CPP and QPP, males and females, how the proportions, $P(g,i,t,65)$, have changed over time. For example, in the left-hand panel of Figure 4, the proportion in Group 11 at age 65 in 1990 (running up and down the vertical line) was about 33% $(100 \times (1-0.67))$, and 26% $(100 \times (0.67-0.41))$ in Group 10. But by 2015, the cohort aged 65 in Group 11 had fallen to just 8%, while Group 10 had grown to about 31%.

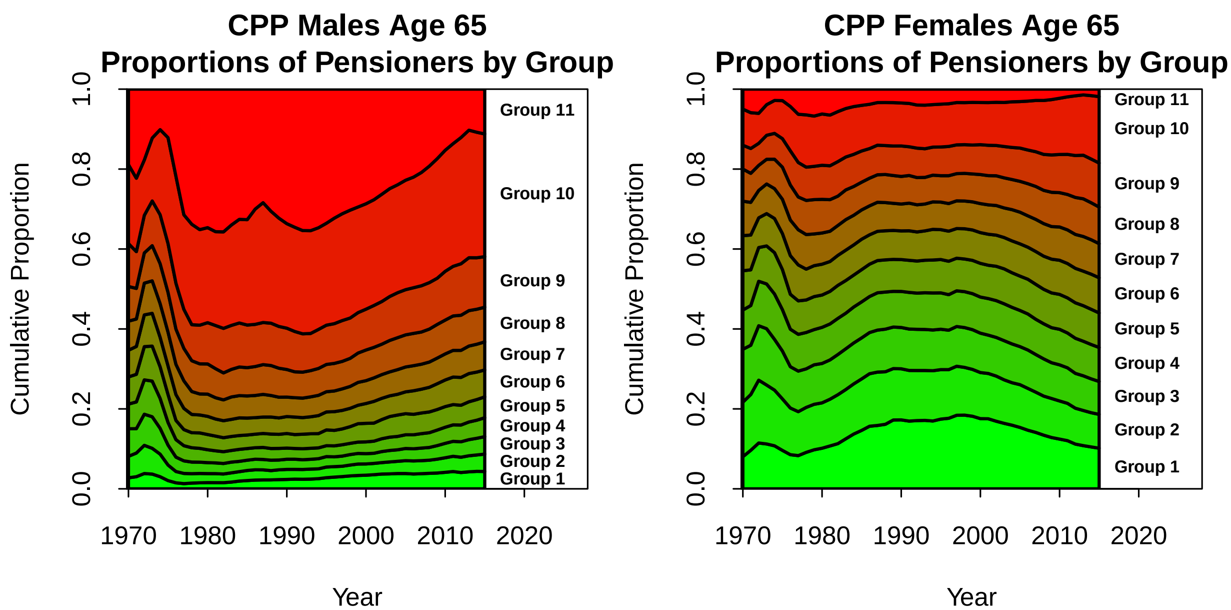

Figure 3: Proportions of pensioners aged 65 in each of Groups 1 to 11 (cumulative) by calendar year. The width of each band gives the proportion in each group. Left: CPP males. Right: CPP females.

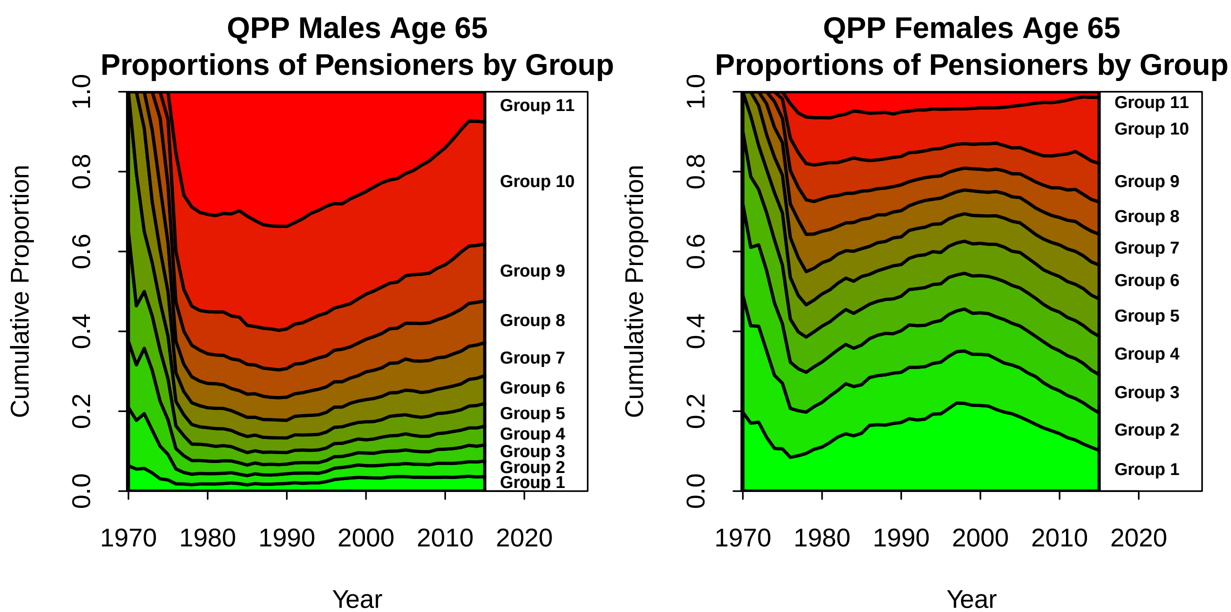

Figure 4: Proportions of pensioners aged 65 in each of Groups 1 to 11 (cumulative) by calendar year. The width of each band gives the proportion in each group. Left: QPP males. Right: QPP females.

We can comment as follows:

- For males, the heat maps for CPP and QPP are broadly similar. Up to about 1977/78 there is considerable distortion, with it being much more difficult in the early years of the plans to attain higher pension levels. The proportions then settle down around 1980 with the largest proportions falling into Groups 10 and 11.

- For females similar comments apply except that a much smaller proportion of females attain higher pensions. In particular Group 1 is generally the largest (although Group 10 has recently exceeded Group 1), reflecting the different work pattern of females compared to males (even after taking account of rules to mitigate the impact of taking some years out to raise a family). Lastly, the same YMPE applies to both males and females, so the extent to which there is a gender pay gap in Canada will be reflected in a lower proportion of females attaining higher pensions in CPP and QPP.

- For males, after 1990 a declining proportion of those reaching their 65th birthday attain the maximum pension (Group 11). The likely reason for this is the number of qualifying number of years to attain the maximum (see Footnote 7) has been changing. In combination with a typical career earnings path (Figure 5) starting low, peaking in middle age and declining slightly towards retirement (see, for example, Blake et al., 2007), this means it would have been easier to attain the maximum pension for someone retiring in 1990 compared to another retiring in 2010 (Figure 5).

- For females, we see a more complex picture. Group 11 declines in size (as for males). Group 1 gradually increases from 1980, peaks just before 2000 and then declines: probably reflecting changes in the underlying work patterns of females.

- For both males and females, the proportions in each group will also reflect historical patterns of immigration. Some individuals retiring at 65 might have lived their entire working lives in Canada, while others might have migrated to Canada during their careers, with a consequent impact on their CPP or QPP pension. For example, an individual who entered Canada at the age of 40 retiring at 65 in 2018 will only have been able to contribute to CPP or QPP in 25 out of the required 47 years. We discuss this further in the next section.

4 Migration and Years of Residency

Given the preceding comments, we need to consider what proportion of each cohort retiring at 65 are immigrants. And how many years have immigrants been contributing to CPP or QPP since they entered the country?

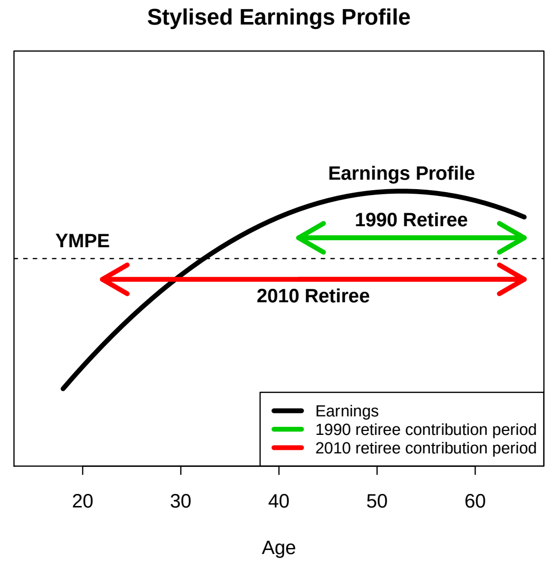

Figure 5: A stylised but typical earnings profile for an individual over their working lifetime. An individual who reaches 65 in 1990 needs to have earned above the YPME in 20 out of 24 years from ages 41 to 64 (1966 to 1989) to get the maximum pension. An individual retiring in 2010 needs 37 out of 44 years (ages 21 to 64) above the YMPE.

We define for each individual

$$\text{Years of Residency (Y)} = \begin{cases} 65 - \text{Age on Arrival} & \text{if age on arrival} > 18 \\ 47 & \text{otherwise.} \end{cases}$$

Out of each cohort we seek to estimate what proportion has $Y = 1, 2, \ldots, 47$. To achieve this we use data from the Canadian Human Mortality Database (CHMD, 2011) for the Canadian provinces (available up to 2011) and adopt a crude set of assumptions. The CHMD data can be used to obtain exposures, $E(i, t, x)$, by year, t, and single age, x, for Québec ($i =$Q) and Canada excluding Québec ($i =$CxQ).

- Following individual cohorts, the change from $E(i, t, x)$ to $E(i, t + 1, x + 1)$ is attributable to deaths and net emigration.
- We assume that:
  - In any year migration is either wholly out of or wholly into Q or CxQ.
  - There is no migration between Q and CxQ.
  - Emigrants do not return to Canada and therefore rejoin CPP or QPP.
  - Plan members retire at 65.

The assumptions are very simplistic and could possibly be improved upon but only with a considerable extra effort. Furthermore, our resulting calculations of years of residency are not used in any further calculations: they are only used to help with qualitative interpretation of the results. More refined calculation of eligible years would be unlikely to change these conclusions.

For each cohort, we use the assumptions above to estimate what proportion of the cohort entered Canada (Q or CxQ) 1 year before age 65, 2 years before 65, ..., 47 or more years before 65. Figure 6 show the results of these proportions as a heatmap. For example, for males, CxQ in 1990 (vertical line): about 75% of the cohort had at least 40 years of residency; 15% had $30 < Y < 40$; 3% had $20 < Y < 30$; 2% had $10 < Y < 20$; and 5% had $0 < Y < 10$.

- The relative sizes of the five groups and the pattern of their changes over time reflects the changing pattern of migration and age profile of immigrants over time.
- A much greater proportion of the Québec population have $Y > 40$ reflecting much lower levels of immigration in Québec at all ages compared to the rest of Canada.
- The black dots in the right-hand panel of Figure 6 pick out places where two boundary lines touch, meaning that there are no individuals within a particular 10-year band for Y in a particular retirement year. The sequence of three dots around 1980, 1990 and 2000 correspond to a sustained period of (net) emigration from Québec in the 1970's at all ages.
- For CxQ we see a more complex plot. However, there is a small but significant flow of immigrants into CxQ above age 55 that contributes to the persistent light green band ($0 < Y < 10$ years).

The size of the light green band ($0 < Y < 10$ years) for CxQ provides us with at least a partial explanation for the anomalous behaviour of male mortality in Groups 1 and 2 in the CPP mortality plots (Figure 2, top right). Groups 1 and 2 cover individuals with a pension of less than 20% of the maximum. Membership of these groups reflects either low earnings or a small number of contributing years. Figure 6 points to CPP having greater numbers of immigrants arriving in their middle ages who would necessarily end up in Groups 1 or 2. Lastly, we have the *healthy-immigrant effect*: immigrants are admitted to Canada on the basis that they are healthy and fit to work, and, therefore healthier than the corresponding established population (see, for example, Vang et al., 2017). Over time, the healthy-immigrant effect fades, so, for our purposes it is strongest amongst the late-middle-age immigrants subsequently reaching age 65. On the other hand, although the selection effect associated with immigration fades, within a particular group by pension level, more recent

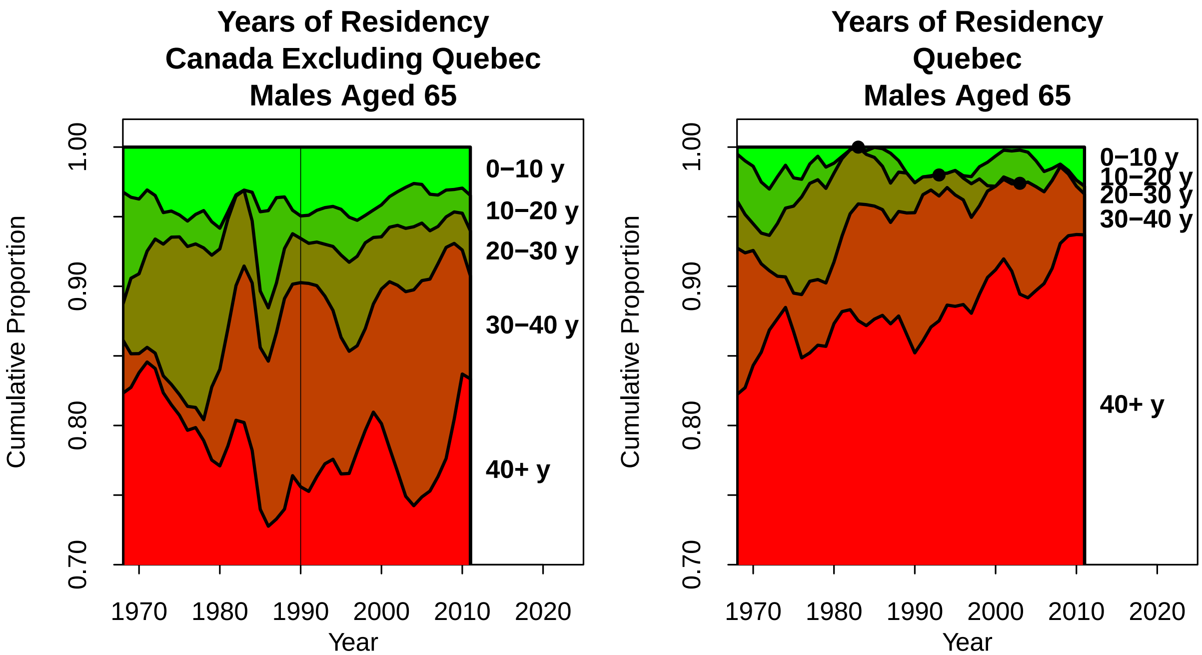

Figure 6: Heat maps showing, for each cohort reaching age 65, estimates of how long people have been Canadian residents and eligible to contribute to CPP/QPP. Proportions are built up cumulatively: 40+ years; $30 - 40$ years; $20 - 30$ years; $10 - 20$ years; < 10 years.

immigrants are likely to be more wealthy and, consequently, more healthy than the 'archetypal' member of Groups 1 or 2 on very low earnings.[18]

So although we would expect Groups 1 and 2 to exhibit high mortality due to people on low earnings, headline mortality rates are reduced because of the presence of healthy immigrants (Group 1 more than Group 2).

As a final point, we can postulate that the proportion of healthy immigrants in Group 1 has been growing over time. Figure 6 shows that the flow of immigrants over age 55 (the 0-10 year band) has been fairly stable over time (no obvious trend). But, for retirement in different years, how low does Y have to be to force an individual into Groups 1 or 2 at age 65? This depends on the number of *eligible* years for a native Canadian: amongst individuals reaching age 65 in 2018, an immigrant with $Y < 8$ is guaranteed to be allocated to Groups 1 or 2; whereas an individual reaching age 65 in 1985 with $Y < 3.4$ will end up for sure in Groups 1 or 2. Correspondingly, everything else being equal, Groups 1 and 2 will have fewer healthy immigrants amongst new retirees in 1985 compared to 2018. (See, also, Appendix B for further discussion.)

[18]Differences between the mortality of immigrants and the established population might also arise for cultural reasons (affecting diet and lifestyle).

For QPP males, Figure 2, there is limited evidence for a healthy immigrant effect: Groups 1, 2 and 3 overlap in an erratic way, with some convergence in the first 10 years.[19]

5 A Family of Multi-Population Stochastic Mortality Models

5.1 Multi-population models

We now move on to consider the use of multi-population stochastic mortality models to capture the historical changes in mortality discussed in the preceding sections. Multi-population models build on the success of single-population and two-population models proposed or discussed by, for example, Lee and Carter (1992), Cairns et al. (2006), Cairns et al. (2011) and Villegas et al. (2017). Existing multi-population models include Li and Lee (2005) and Cairns et al. (2019).

The CPP and QPP data each have 11 populations for each gender, so we seek to model $\log m(i,t,x)$ jointly for $i = 1, \ldots, 11$. Rather than focus on one specific model from the outset, we consider a family of models that have a multi-population version of the Renshaw and Haberman (2003) (RH) model as the most general case. Thus, model M1 is

$$\log m(i,t,x) = \alpha(i,x) + \beta_1(i,x)\kappa_1(i,t) + \beta_2(i,x)\kappa_2(i,t).$$

In this model:

- The age effect $\alpha(i,x)$ can be interpreted as a base table for Group i.
- $\beta_1(i,x)$ and $\beta_2(i,x)$ are group-specific age effects, that we normally anticipate will allow for changes in the level (i.e. $\beta_1(i,x) > 0$ for all x) and slope (i.e. $\beta_2(i,x) > 0$ for lower ages and < 0 for higher ages) respectively of the log-mortality curve.
- $\kappa_1(i,t)$ and $\kappa_2(i,t)$ are period effects that, in combination with the $\beta_1(i,x)$ and $\beta_2(i,x)$ age effects capture the group-specific variation in mortality from the base table over time.

All other models considered were special cases of M1 and are listed in Table 1.

[19] We do not detail this further in Section 5, but model-based estimates of the underlying mortality of QPP males reveal more clearly lower mortality for Groups 1 and 2, particularly at higher ages, less so at lower ages. This might be linked to late immigrants deferring retirement until age 70 or later.

Model	$\log m(i,t,x)$	Comment
M1	$\alpha(i,x) + \beta_1(i,x)\kappa_1(i,t) + \beta_2(i,x)\kappa_2(i,t)$	multi-population RH
M2	$\alpha(i,x) + \beta_1(i,x)\kappa_1(i,t) + \beta_2(x)\kappa_2(i,t)$	
M3	$\alpha(i,x) + \beta_1(x)\kappa_1(t) + \beta_2(i,x)\kappa_2(i,t)$	Li & Lee (2005)
M4	$\alpha(i,x) + \beta_1(i,x)\kappa_1(i,t)$	multi-population LC
M5	$\alpha(i,x) + \beta_1(x)\kappa_1(i,t) + \beta_2(x)\kappa_2(i,t)$	CAE model, Kleinow (2015)
M6	$\alpha(x) + \beta_1(x)\kappa_1(i,t) + \beta_2(x)\kappa_2(i,t)$	CAE model with common $\alpha(x)$
M7	$\alpha(i,x) + \kappa_1(i,t) + (x-\bar{x})\kappa_2(i,t)$	Multi-population Plat
M8	$\alpha(x) + \kappa_1(i,t) + (x-\bar{x})\kappa_2(i,t)$	Plat with common $\alpha(x)$
M9	$\alpha(i,x) + \kappa_1(t) + (x-\bar{x})\kappa_2(i,t)$	Plat, common $\kappa_1(t)$
M10	$\alpha(i,x) + \kappa_1(i,t) + (x-\bar{x})\kappa_2(t)$	Plat, common $\kappa_2(t)$
M11	$\alpha(i,x) + \kappa_1(t) + (x-\bar{x})\kappa_2(t)$	Plat, common $\kappa_1(t)$, $\kappa_2(t)$

Table 1: Stochastic mortality models fitted to CPP and QPP males and females mortality data.

Some models are nested within others (e.g. all models are nested within M1) and the resulting hierarchy is illustrated in Figure 7.

In models M1, M5 and M6, the terms $\beta_1(i,x)\kappa_1(i,t)$ and $\beta_2(i,x)\kappa_2(i,t)$ are interchangeble with no impact on the model fit. The two components can normally be left as they are following model fitting. However, if we prefer that $\beta_1(i,x)\kappa_1(i,t)$ captures changes in the level of mortality across all ages, and $\beta_2(i,x)\kappa_2(i,t)$ captures changes in the slope (or a tilt), then, if required, we can swap round the two components.

A wider range of models is considered and reviewed by Villegas et al. (2017). Their focus is more on choosing an appropriate model for *two* populations (one dominant population and a second sub-population with specific characteristics) compared to 11 here (treated in a more balanced way). The list here also considers some models not considered by Villegas et al. (2017). Adam (2016) analyses several models using male and female CPP and QPP data simultaneously with three income classes and settles upon a model that is close to the Li and Lee (1995) model (here, M3).

We consider a range of quantitative and qualitative criteria to compare models and recommend which model is best for the multiple population dataset over ages 65-89.

From a quantitative perspective we use the Bayes Information Criterion,

$$BIC_M = -2l_M(\hat{\theta}_M) + k_M \log N,$$

where θ_M is the parameter vector for model M and $\hat{\theta}_M$ is its maximum likelihood estimator, $l_M(\theta_M)$ is the log-likelihood function, k_M is the number of parameters in M to be estimated (taking account any identifiability constraints), and N is the number of observations.[20] The term $k_M \log N$ serves to penalise models that

[20]Here $N = 11 \times 25 \times 25 = 6875$ (groups $\times$ years $\times$ ages).

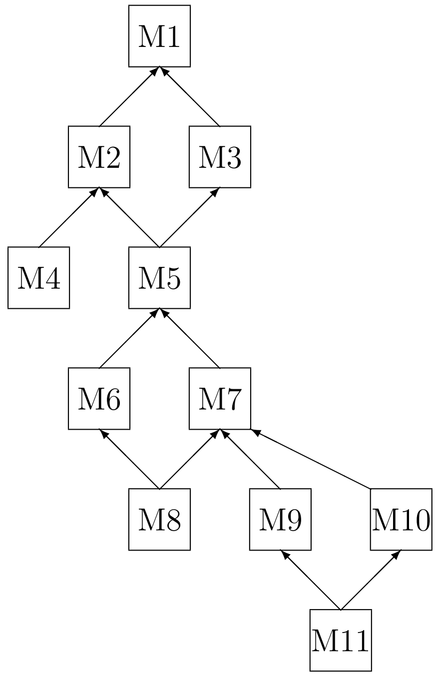

Figure 7: Nested model hierarchy. Arrows indicate nesting: e.g. M2 is nested (i.e. is a special case) of M1.

are overparameterised. In aiming to minimise the BIC, we seek to include greater complexity (e.g. favouring M5 over M6) only if that greater complexity results in a *significant* improvement in the fit. In particular, where models are nested, the more complex model will always achieve a better fit (i.e. the maximum log-likelihood will be higher) but the improvement might be quite marginal and the extra parameters are simply overfitting the data.

We also consider the forward correlation term structure:

$$\rho(t, i, j, x_i, x_j) = \text{cor}\left(\log m(i, t, x_i), \log m(j, t, x_j)\right).$$

Our principal desirable criteria are as follows.

1. The BIC should not be significantly higher than other models.
2. The model should satisfy the principle of coherence. That is, for each (i, j, x), $\log\left[m(i, t, x)/m(j, t, x)\right]$ should not diverge as t gets large (see, for example, Hyndman et al., 2013).
3. The model should avoid significant crossovers in fitted mortality curves where these are not apparent in the raw data. For example, in the historical data and in forecasts, for a given t, is $m(1, t, x) > m(11, t, x)$ over all ages x?
4. Correlations between future mortality rates in different populations should be less than 1.

5. Correlations between future mortality rates at different ages should be less than 1.

6. Does the model produce a plausible forward correlation term structure? For example, for a given (t, i, j, x_i), does the shape of $\rho(t, i, j, x_i, x_j)$ as a function of x_j look reasonable: e.g. unimodal with a peak close to x_i? And is $\rho(t, i, j, x, y) < \rho(t, i, i, x, y)$ (i.e. the correlation between two ages in the same population is likely to be higher than the same ages in different populations)?[21]

Additional relevant criteria can be found in Cairns et al. (2009) and Villegas et al. (2017).

5.2 Model selection outcome: M6 – CAE with common $\alpha(x)$

There was no single model that satisfied all criteria better than all other models. In particular, the model with the lowest BIC did not completely satisfy some of the qualitative criteria.

On balance, we selected model M6 (CAE with common $\alpha(x)$) as being the most suitable for both the CPP and QPP males and females datasets: this model had one of the lowest BIC values (but not *the* lowest) as well as satisfying the qualitative criteria. For a full discussion of the model selection process, see Appendix A.

Maximum likelihood estimates of the age and period effects for CPP and QPP males and females are presented in Figures 8 to 11. In each of the four cases (e.g. CPP males), the model is fitted to the 11 groups jointly, resulting (for M6) in estimates for the common age effects, $\alpha(x)$, $\beta_1(x)$ and $\beta_2(x)$, and the group-specific period effects, $\kappa_1(i, t)$ and $\kappa_2(i, t)$.

We can comment on the model and results as follows:

- $\beta_1(x)$ is linked (in combination with the $\kappa_1(i, t)$) to changes in the level of mortality. Specifically, since $\beta_1(x)$ is positive at all ages, if $\kappa_1(i, t)$ falls, then death rates fall at all ages and, given the shape of $\beta_1(x)$ falls by a larger percentage at younger ages than older ages.

- Since $\beta_2(x)$ is positive at younger ages and negative at older ages, it is linked to tilts. If $\kappa_2(i, t)$ falls then death rates *fall* at younger ages and *rise* at older ages: that is the death rate curve tilts around, approximately, age 77 (where $\beta_2(x)$ crosses from positive to negative).

- The broad shapes of $\beta_1(x)$ and $\beta_2(x)$ are similar for all four main populations. The $\beta_2(x)$ curves, in particular, are very similar, with a characteristic shape

[21] See, Cairns et al. (2019) for further discussion of plausible forward correlation term structures.

that is fairly steep and then *gradually* flattens off when it turns negative. This specifically links $\beta_2(x)$ to the curvature that we observe in many log mortality curves, which are less steep in middle age and become more steep in older age, converging to a Gompertz type of mortality curve.

- The similarity of the $\beta_1(x)$ and $\beta_2(x)$ for the four higher-level populations provides evidence that the underlying model is robust.
- For some groups, estimated curves for the period effects are relatively smooth (e.g. QPP males, Group 11) while others are quite volatile (e.g. QPP females, Group 11): both consequences of sampling variation in the observed deaths.
- Since M6 assumes a common base table, $\alpha(x)$, differences in the level of mortality between the 11 groups is modelled through the $\kappa_1(i,t)$ (in particular) and $\kappa_2(i,t)$. In each figure, therefore, we see that the $\kappa_1(i,t)$ estimates for high-pension groups tend to be lower (hence lower mortality) than lower-pension groups.
- The trend, shape and local volatility of the group-specific $\kappa_1(i,t)$ very closely match the shape of the corresponding ASMR's plotted in Figure 2. This indicates that the $\kappa_1(i,t)$ are the main drivers of headline mortality.
- The $\kappa_2(i,t)$ are generally more variable and consequently trends are less easy to detect. However, in some cases, there is an upwards trend in the $\kappa_2(i,t)$ after the late 1990s. This, in combination with decreasing values for the corresponding $\kappa_1(i,t)$,[22] is consistent with increased mortality improvement rates at high ages.
- In each figure we can also see that the $\kappa_2(i,t)$ are typically quite low for high-pension groups (e.g. $i = 10, 11$) and high for low-pension groups. Consequently, log-mortality curves tend to be fairly linear for high-pension groups and more convex for low-pension groups.
- A widening gap between the $\kappa_1(i,t)$ reflects generally increasing levels of mortality inequality (e.g. QPP males and females).[23]
- A widening gap between the $\kappa_2(i,t)$ normally means growing levels of inequality at the younger ages much more than higher ages (e.g. CPP females, Groups 1 and 10).

Model fitting outputs also include estimates of the underlying death rates:

$$\log m(i,t,x) = \alpha(x) + \beta_1(x)\kappa_1(i,t) + \beta_2(x)\kappa_2(i,t).$$

[22] Note that the trend in the $\kappa_1(i,t)$ also changes around the same time.

[23] Note that the assumption of coherence in the model (Section 5.1) does not prevent some degree of divergence in the short and medium term. So the slight widening observed here is not inconsistent with the coherence assumption.

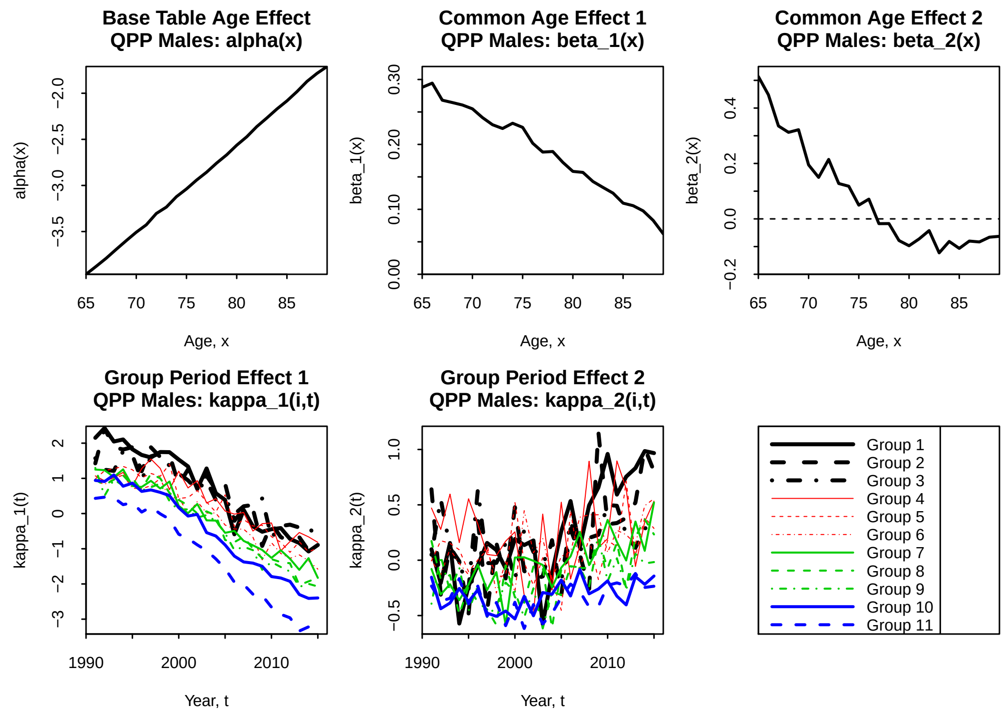

Figure 8: M6 fitted age and period effects for QPP males.

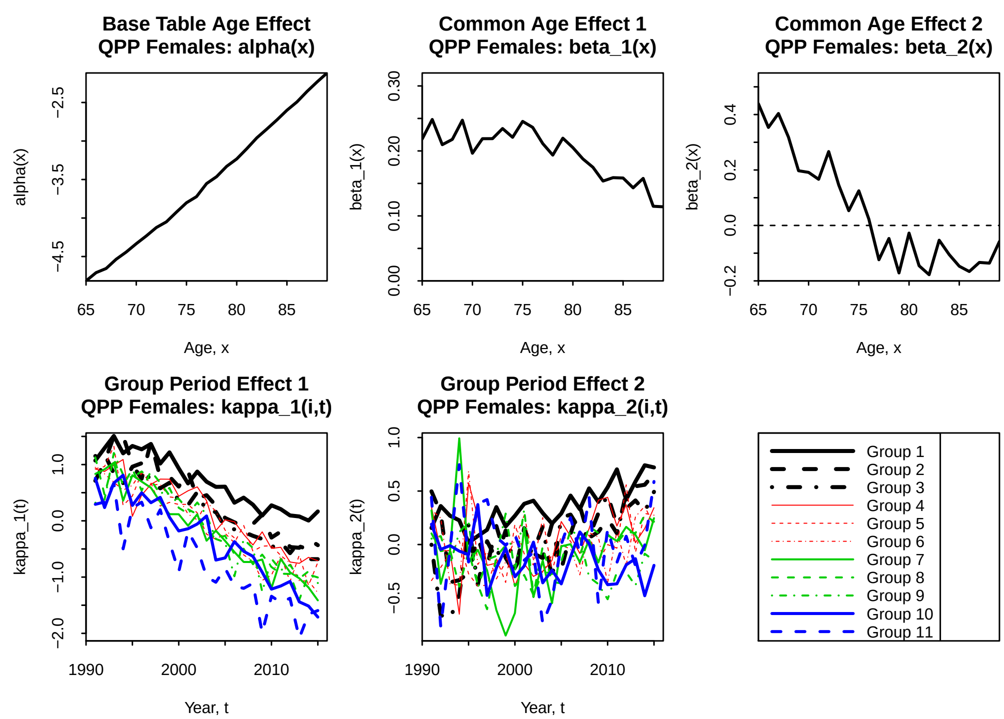

Figure 9: M6 fitted age and period effects for QPP females.

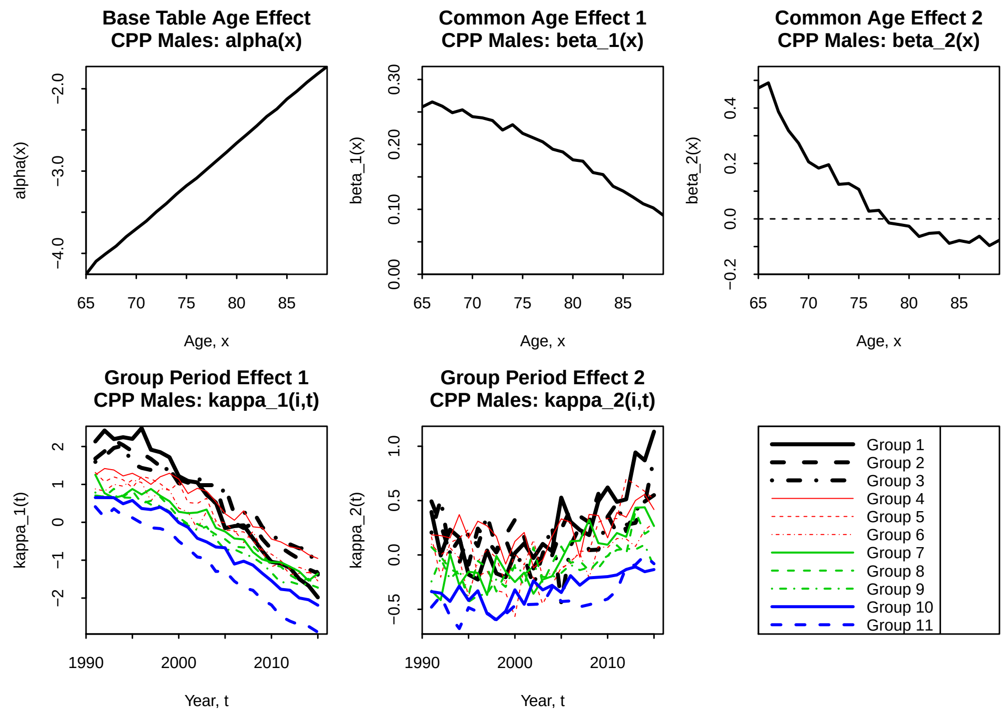

Figure 10: M6 fitted age and period effects for CPP males.

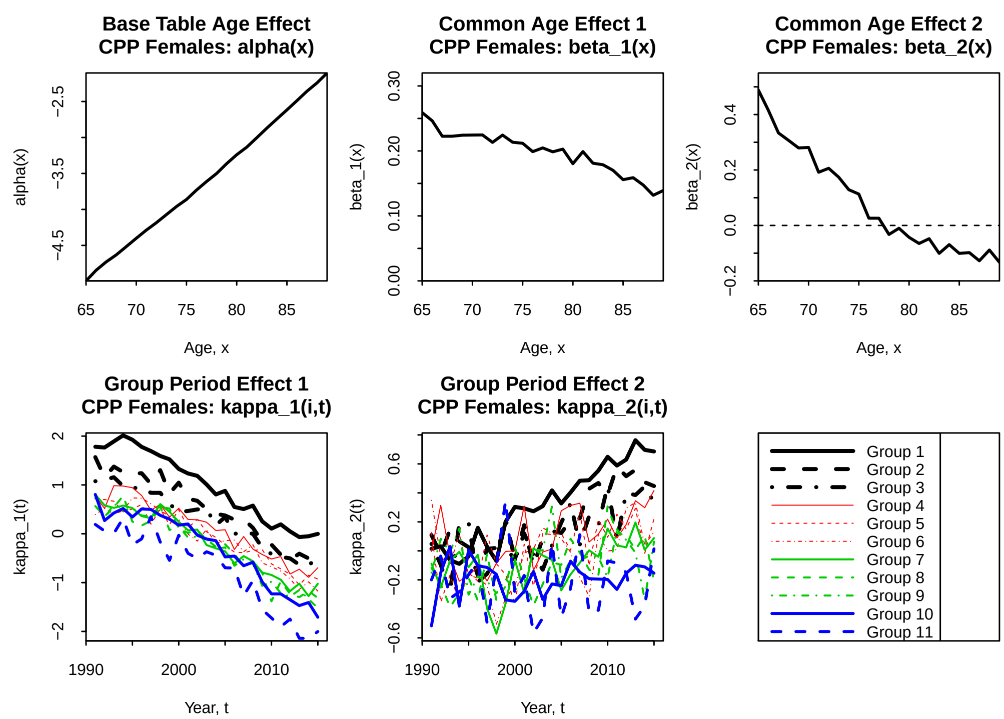

Figure 11: M6 fitted age and period effects for CPP females.

Examples of these for 1995 and 2010 are plotted in Figures 12 and 13.

For the CPP males and females data (Figure 12) some aspects of the plots are as we might expect: higher pension groups experiencing lower mortality; a wider gap at lower ages converging at higher ages. But there are a number of non-standard features or anomalies that were not evident in the earlier plots of the ASMR's (Figure 2).

- For CPP males, the 1995 mortality curves look fairly standard. On a more detailed level, we note that Groups 1-6 are tightly clustered at age 65, and then a spread develops by age 70 and beyond, perhaps connected to the healthy-immigrant effect.

  By 2010, the inequality gap has widened at age 65, but not at high ages. More importantly, we can see the anomalous Groups 1 and 2 in more detail than the previous summary ASMR's. At age 65, Groups 1 and 2 have amongst the highest mortality, but the mortality curves then drift well below Groups 3 and 4, ending up close to Groups 10 and 11. This behaviour is consistent with the healthy-immigrant effect and the heterogeneity that this creates within each group. Although this is a period rather than a cohort mortality curve, Group 1 will consist of a mixture of low-paid, long-stay/native Canadians and higher-paid, newer healthy immigrants. The former will die off at a much faster rate leaving a much higher proportion of late immigrants in their 80's in Group 1 with low mortality.

For the QPP males and females data (Figure 13) some aspects of the plots are as we might expect: higher pension groups experiencing lower mortality; a wider gap at lower ages converging at higher ages. But there are a number of non-standard features or anomalies that were not evident in the earlier plots of the ASMR's (Figure 2).

- For QPP males, 1995 looks reasonably 'standard' with the exception of Group 4 where older individuals have relatively low mortality. This might be consistent with the healthy-immigrant effect: older pensioners in 1995 would have retired in the early years of QPP when the number of eligible years would have been quite short allowing healthy, middle-aged immigrants entering the country in the 1970's to accumulate a reasonable pension (e.g. in the 30-40% band) by age 65.

  By 2010 we see a significant widening of the inequality gap at age 65. We also see a potential healthy-immigrant effect has emerged in Group 1 which is similar to, but smaller than, the effect in CPP Group 1 that we discuss further below.

- For QPP females we see some similar features to males: a widening inequality gap; Group 4 having low mortality in 1995.

Figures 12 and 13 can be compared to, and are consistent with, Adam (2016, Slides 12 and 13; 2009-2011 mortality). But, with the more detailed groupings here, we see how mortality inequalities continue to accumulate as we move right into the upper and lower tails of the income spectrum.

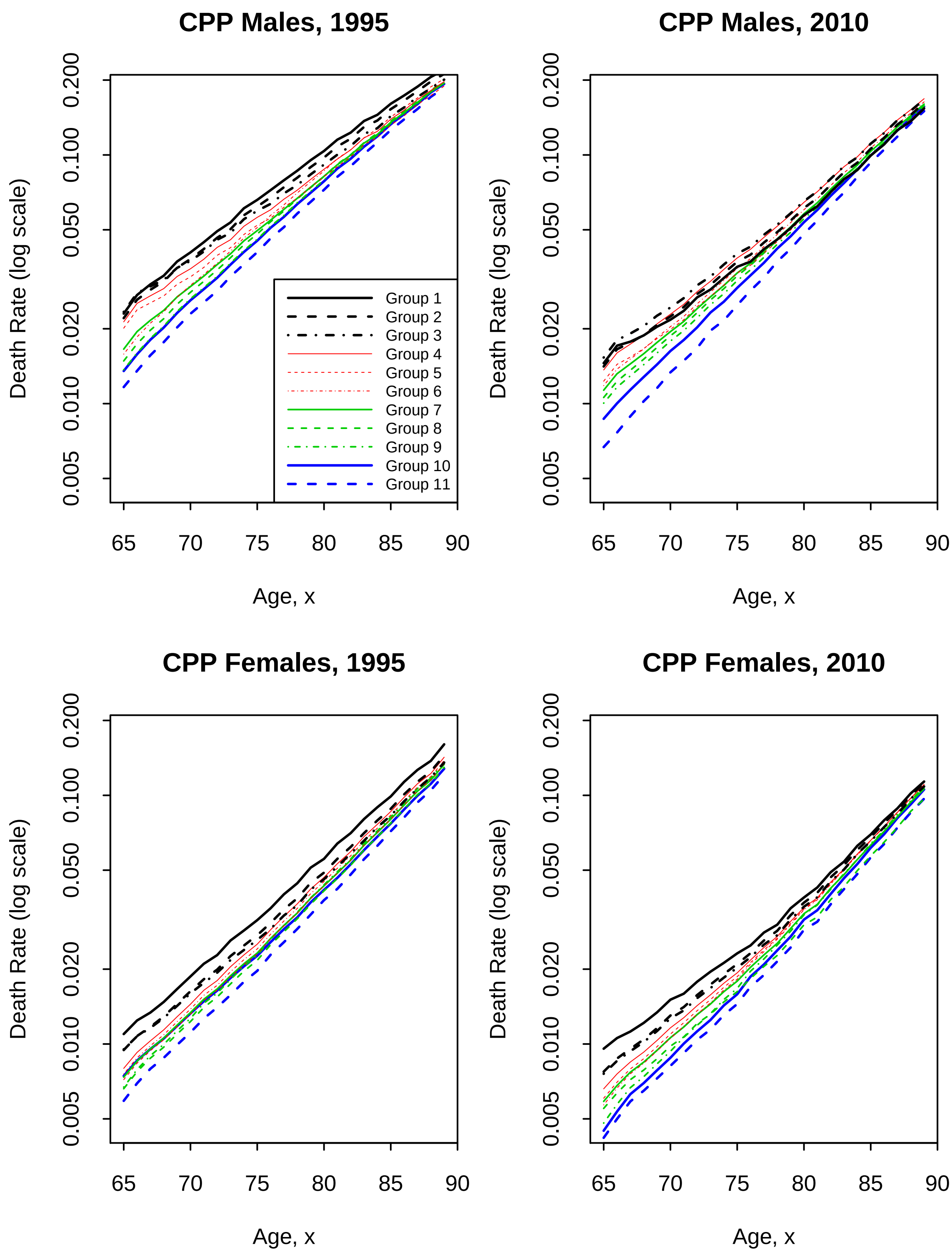

Figure 12: Fitted death rates by group for CPP males and females in 1995 and 2010.

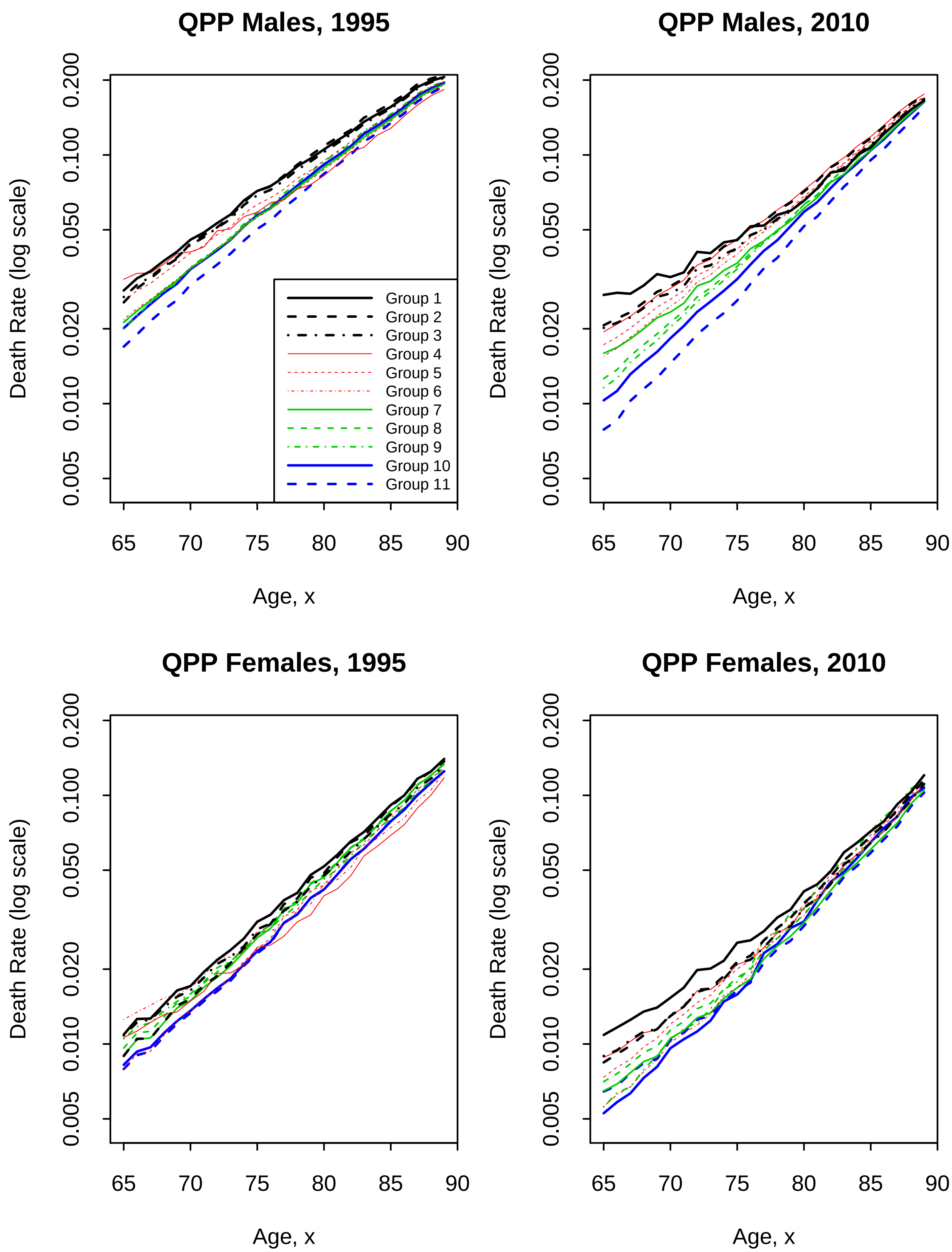

Figure 13: Fitted death rates by group for QPP males and females in 1995 and 2010.

6 Clustering

In the last section it was evident that large groups (e.g. QPP and CPP males, Groups 10 and 11) have relatively smooth estimates for the period effects (Figures 8 and 10, $\kappa_1(i,t)$, Groups 10 and 11). In contrast, some groups (see, for example, Figure 2, QPP males, Groups 1 to 4) typically represent only 2-4% of each cohort with more volatile estimates of the period effects (correspondingly, Figure 8, Groups 1 to 4). The reason for this is that, even after model fitting, greater sampling variation in death counts in small groups results in noticeably greater sampling variation in fitted period effects.[24] A consequence of this additional noise in fitted period effects is that forecast levels of uncertainty in future mortality can be artificially high for these small groups (Chen et al., 2017; Villegas et al., 2017).

At the same time, we can note that some adjacent groups typically have quite similar levels of mortality. Combining this with the slightly artificial choice of group boundaries (10%, 20% etc. of the maximum pension), it would seem appropriate to consider combining adjacent groups into *clusters*. Clusters will typically improve the results if groups are small and have similar levels of mortality.

A systematic approach was taken to consider all possible clusters of the 11 groups.[25] For QPP males using model M6, for example, we found that four clusters were optimal (optimal BIC) as detailed in Table 2. For the same population, other models typically, but not always, found the same four clusters to be optimal. Optimal clustering differed more for other populations, especially females (partly because of their very different group sizes).

Fitted ASMR's for QPP males under M6 with and without clustering are plotted in Figure 14. Without clustering, we can clearly see how the smaller groups 1 to 5 produce high levels of volatility in the ASMR from year to year. With clustering we can see that the ASMR's for the four clusters all now exhibit similar levels of volatility: partly due to the larger sizes generally, partly because the clusters are all more similar in size (Table 2).

With clustering, the smoother ASMR's also now allow us to see more clearly the different trends experienced by the different QPP clusters. In particular, we see a widening of the gap between the different clusters of QPP males. The reasons for this are not clear, although it might, in part, be due to the changing cluster sizes by cohort (Figure 4).

Figure 15 shows the fitted age effects, $\alpha(x)$, $\beta_1(x)$ and $\beta_2(x)$ without and with clustering. Importantly, the shift from 11 groups to four clusters has relatively little

[24] For further discussion of the impact of group size, see Chen et al., 2017.

[25] The systematic approach took advantage of the natural ordering of the Groups by assuming that clusters can only consist of adjacent groups (e.g. Groups 1 to 4 being a permissible cluster, but not Groups 1, 3 and 6). There are then $2^{10} = 1024$ possible sets of clusters to consider.

	Cluster	Groups	Exposures
QPP Males	1	1-5	1.799 M
	2	6-8	2.124 M
	3	9-10	3.956 M
	4	11	2.826 M
QPP Females	1	1-2	3.138 M
	2	3-11	6.572 M
CPP Males	1	1-4	3.699 M
	2	5-8	6.326 M
	3	9-10	10.740 M
	4	11	8.277 M
CPP Females	1	1	4.832 M
	2	2-3	5.877 M
	3	4-6	6.813 M
	4	7-11	10.287

Table 2: Optimal clustering for QPP and CPP males and females aged 65-89 and years 1991-2015.

impact on estimates of $\beta_1(x)$ and $\beta_2(x)$: further evidence that model M6 is robust.[26]

Corresponding results for QPP females and CPP males and females are plotted in Figures 16 to 21.

For CPP males (Figures 18 and 19), the picture is much clearer with clusters compared to 11 groups, now that ASMR's have been smoothed out. Cluster 1 (Groups 1-4) shows the conjectured impact of the healthy immigrant effect. Perhaps because of this, there is little evidence for a widening inequality gap. It is tempting to compare cluster 4 for QPP and CPP males: the latter being lower. However, we need to recall that the QPP data include disability pensioners while the CPP data do not.

Both QPP males and CPP males exhibit a slight widening of the gap between clusters 3 and 4, potentially linked to the gradual shrinkage of Group 11 over time (Figures 3 and 4).

We can also look for evidence of a slowdown in mortality improvements. The ASMR's based on clusters are generally smoother than the 11 groups making it easier to detect trend changes. Clustering has reduced noise in the plots. Despite this, any evidence for more of a slowdown in one group than another is mixed. Additionally, in assessing a slowdown, this should take account of differentials in improvement rates prior to 2010. For example, in Figure 16, Cluster 1 appears to exhibit a bigger slowdown than Cluster 2. However, when improvement rates after

[26]The small shift in $\alpha(x)$ in Figure 15 is due to the application of identifiability constraints in the model fitting process.

2010 are compared with improvement rates before 2010, the amount of the slowdown in the two clusters is not as significant as it appears to be.

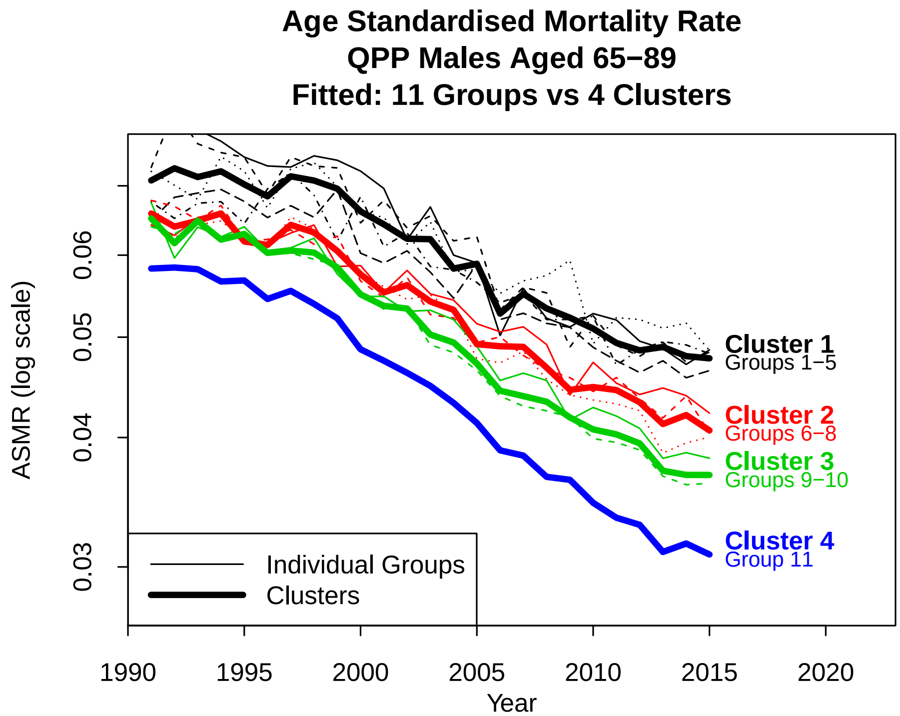

Figure 14: ASMR's based on fitted mortality using model M6 for QPP males from 1991 to 2015 for the original groups (thin lines) and the optimal clusters (thick lines).

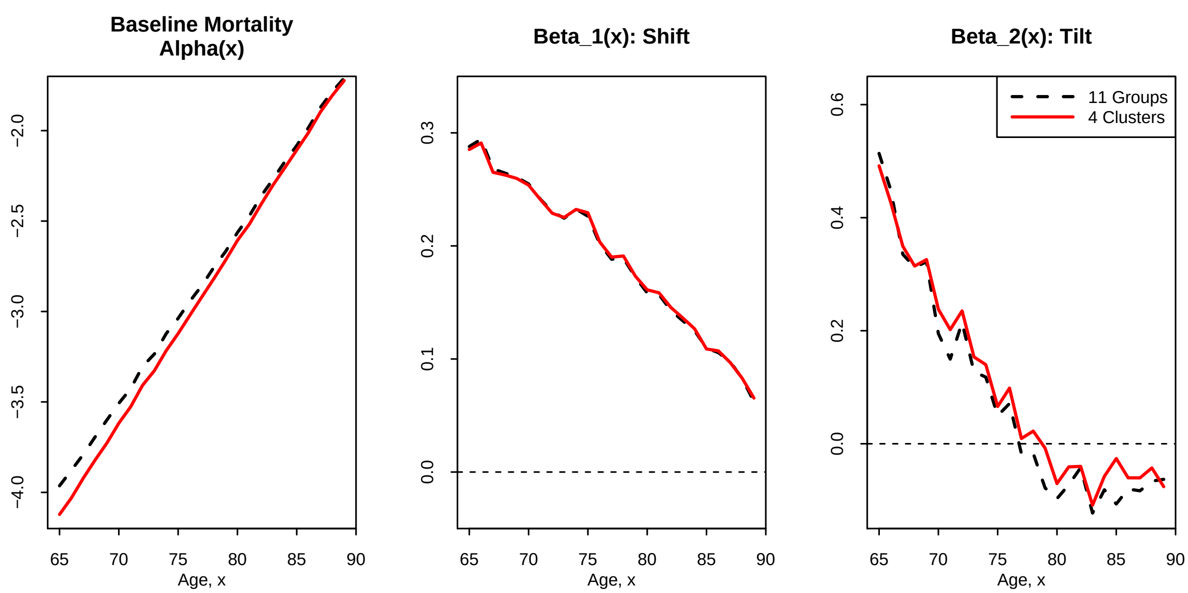

Figure 15: Common age effects, $\alpha(x)$, $\beta_1(x)$ and $\beta_2(x)$, for QPP males without (11 Groups) and with (4 clusters) clustering.

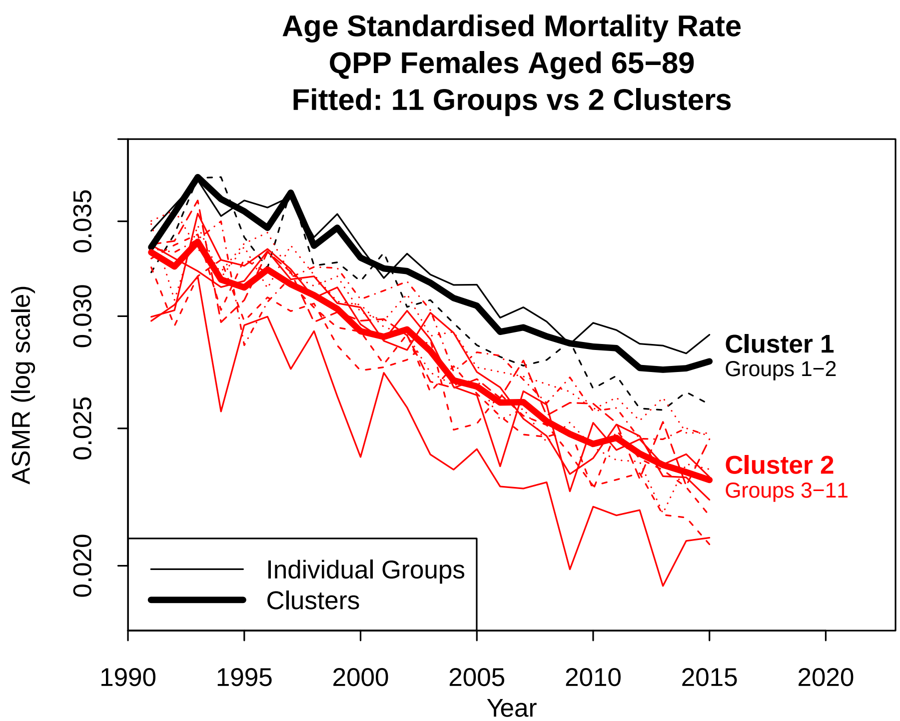

Figure 16: ASMR's based on fitted mortality using model M6 for QPP females from 1991 to 2015 for the original groups (thin lines) and the optimal clusters (thick lines).

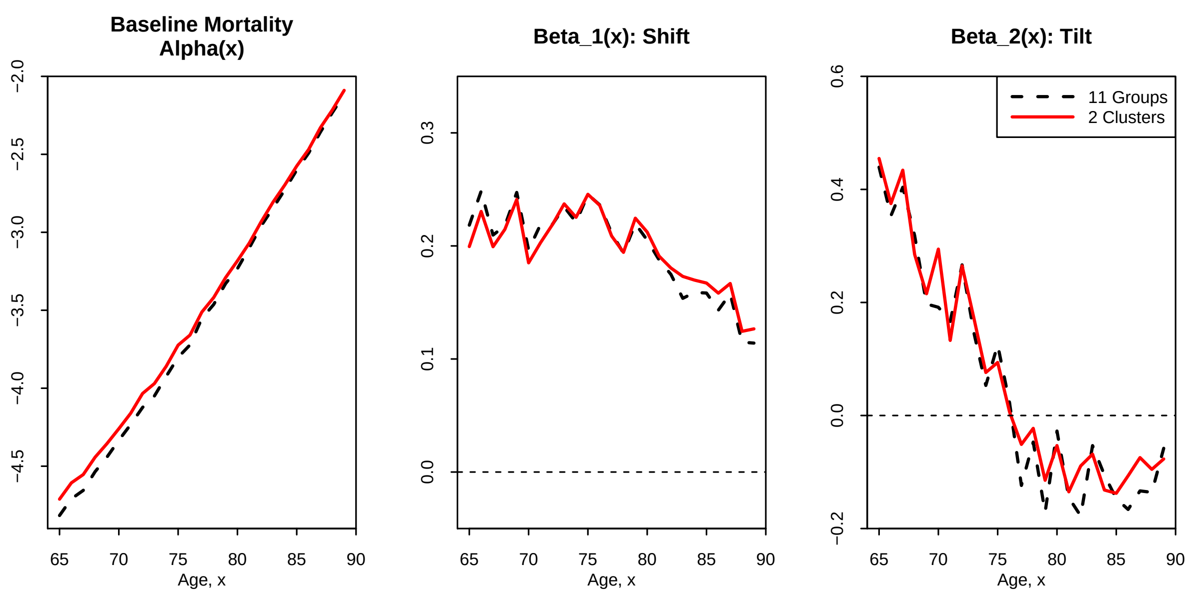

Figure 17: Common age effects, $\alpha(x)$, $\beta_1(x)$ and $\beta_2(x)$, for QPP females without (11 Groups) and with (2 clusters) clustering.

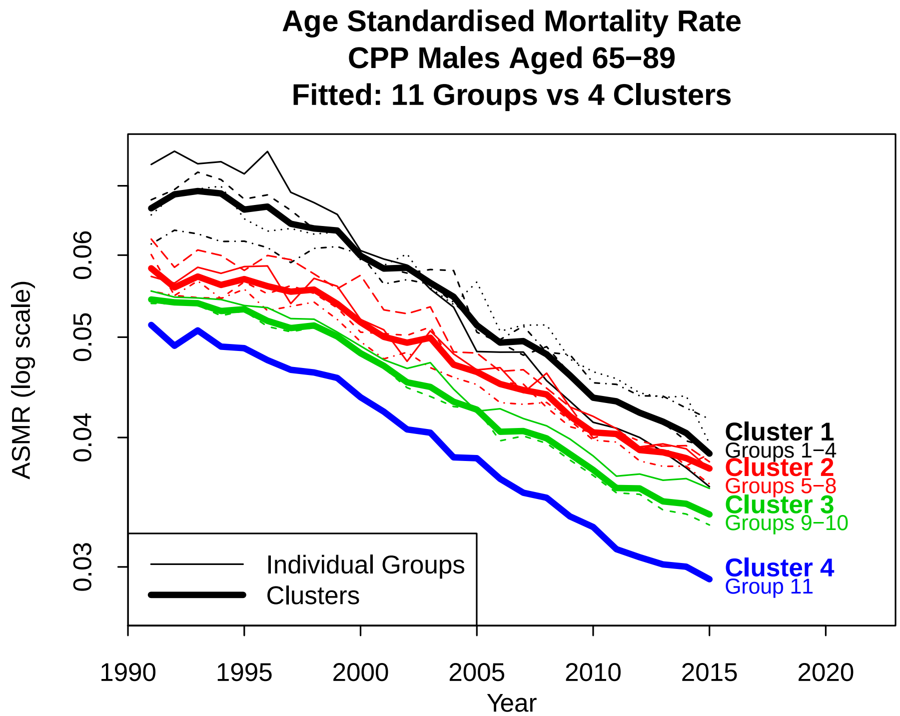

Figure 18: ASMR's based on fitted mortality using model M6 for CPP males from 1991 to 2015 for the original groups (thin lines) and the optimal clusters (thick lines).

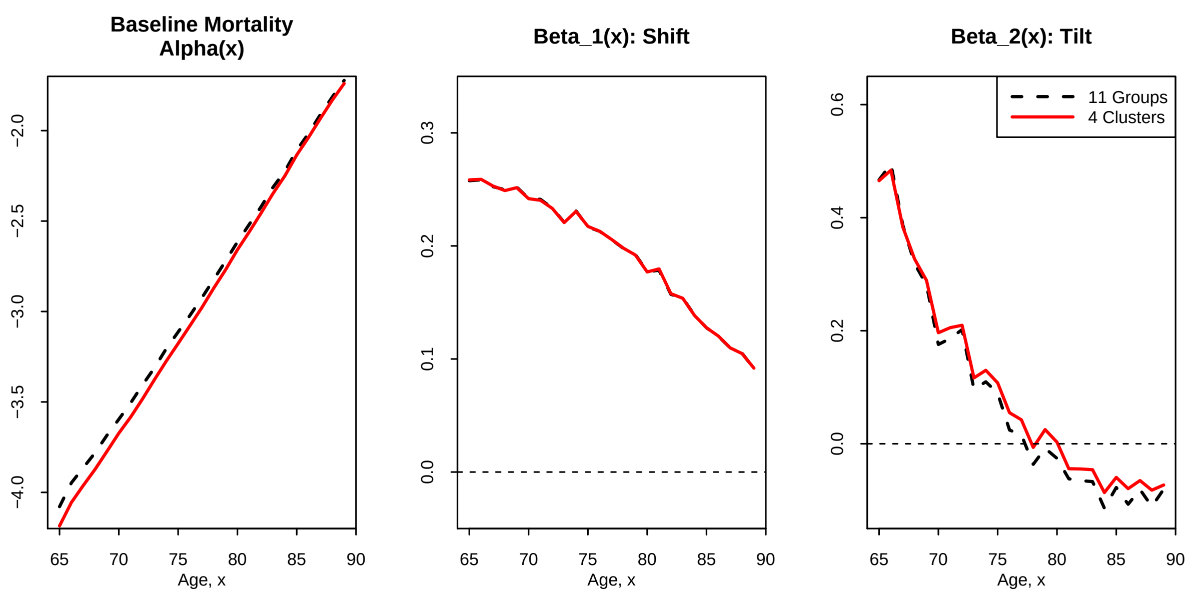

Figure 19: Common age effects, $\alpha(x)$, $\beta_1(x)$ and $\beta_2(x)$, for CPP males without (11 Groups) and with (4 clusters) clustering.

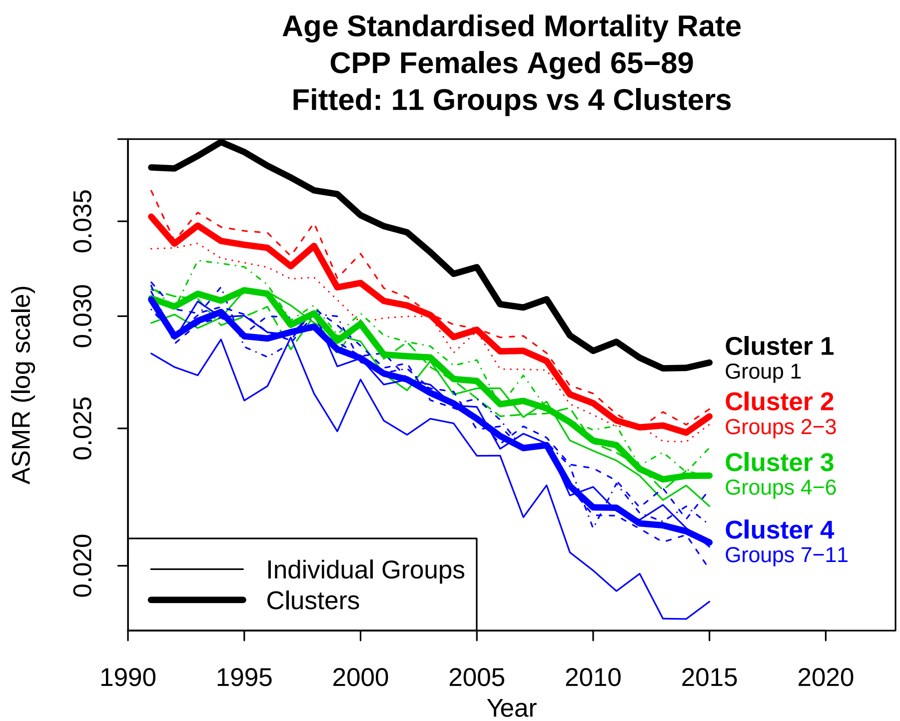

Figure 20: ASMR's based on fitted mortality using model M6 for CPP females from 1991 to 2015 for the original groups (thin lines) and the optimal clusters (thick lines).

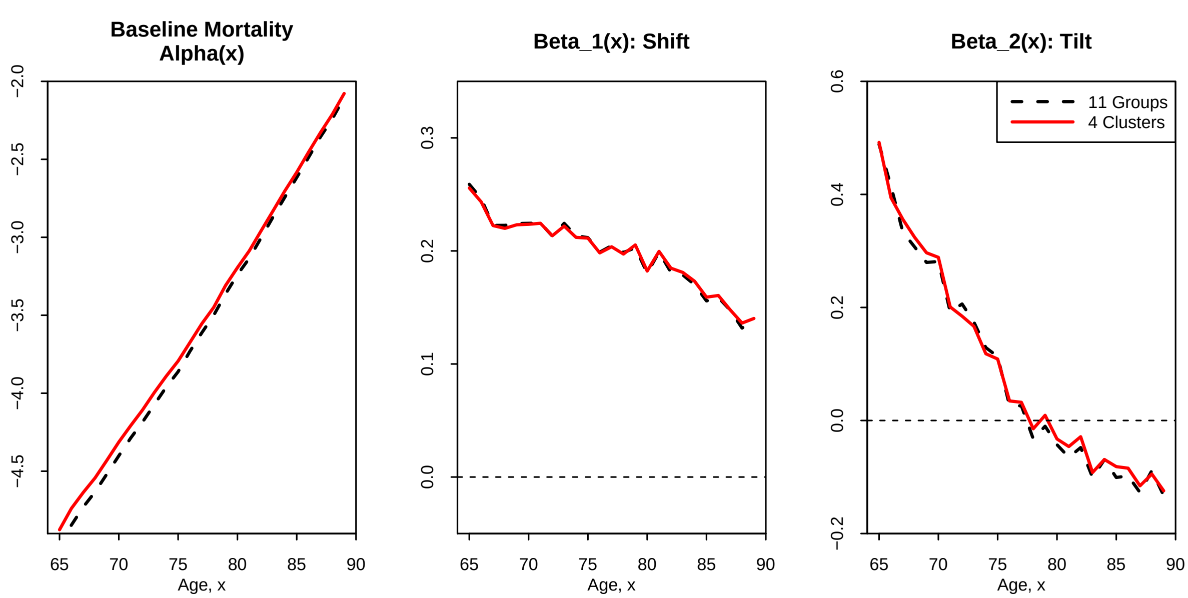

Figure 21: Common age effects, $\alpha(x)$, $\beta_1(x)$ and $\beta_2(x)$, for CPP females without (11 Groups) and with (4 clusters) clustering.

7 Conclusions

We have carried out a detailed analysis of the mortality of CPP and QPP pensioners subdivided by pension level. Analysis has been backed up by a careful look at how the profile of each cohort by pension level has changed over time as well as the impact of migration on mortality levels.

Headline conclusions were consistent with what we observe in other countries:

- significant variation in the level of mortality by pension level between all pension levels, especially at younger ages;
- the inequality gap narrows with age.

Other important conclusions could not have been immediately anticipated at the outset:

- different patterns of inequality for males and females across the 11 groups;
- greater levels of inequality in QPP than CPP;
- a prominent healthy-immigrant effect that has a significant impact on the observed mortality of low pension groups in the CPP;
- a widening inequality gap between ages 65 and 75.

(The reasons for differing levels of inequality and the widening inequality gap are not clear and are likely to be complex.)

The second half of the paper considered how stochastic mortality models can be used to enhance our analysis of historical mortality as well as provide a stepping stone towards projections. A wide variety of multi-population mortality models were considered and, through consideration of a mixture of quantitative and qualitative criteria, we found that the Common Age Effect model (M6) was best suited to the data being considered. The model indirectly provides a smoothing mechanism by pooling data over many more years than a traditional actuarial graduation. The model outputs then reveal further detail in the data not previously apparent including an indication of the strength of the healthy immigrant effect. This smoothing, if unchecked, might inadvertently introduce model and parameter risk. It is, therefore, important to check that the preferred models do not distort the underlying data to any significant extent beyond simply smoothing out sampling variation.

Evidence that the slowdown in Canadian mortality improvements might be greater at one end of the socio-economic spectrum than the other was mixed, especially when differences in improvement rates prior to 2010 were taken into account.

Acknowledgements

The authors would like to thank the Project Oversight Group (POG) at the Canadian Institute of Actuaries for their support, input and feedback throughout the project; Alain Guimond and Michel Montambeault for preparing the CPP data; and Mario Pépin for preparing the QPP data. The authors are also grateful to actuaries at the Office of the Superintendent of Financial Institutions and Retraite Québec for their feedback on earlier versions of this report.

The authors gratefully acknowledge funding from the Actuarial Research Centre of the Institute and Faculty of Actuaries, the Canadian Institute of Actuaries and the Society of Actuaries through the "Modelling Measurement and Management of Longevity and Morbidity Risk" research programme (see `www.actuaries.org.uk/arc`).

References

Adam, L. (2012a) The Canadian Pensioners Mortality Table: Information on mortality for the triennial period ending December 31, 2007, with data as at December 31, 2008 Université Laval, Québec City, QC, Canada May 31, 2012.

Adam, L. (2012b) The Canadian Pensioners Mortality Table: Historical Trends in Mortality Improvement and a Proposed Projection Model Based on CPP/QPP Data as at December 31, 2007 Université Laval, Québec City, QC, Canada December 3, 2012.

Adam, L. (2016) A comparative analysis of cohort, region and income in the measurement of mortality for Canadian Pensioners as at December 31st, 2012. Presentation, 2016 SOA Life and Annuity Symposium, Nashville.

Blake, D., Cairns, A.J.G., and Dowd, K. (2007) The impact of occupation and gender on the pensions from defined contribution plans, *Geneva Papers on Risk and Insurance*, 32: 458-82.

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

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, 13: 1-35.

Cairns, A.J.G., Kallestrup-Lamb, M., Rosenskjold, C.P.T., Blake, D., and Dowd, K., (2019) Modelling Socio-Economic Differences in the Mortality of Danish Males

Using a New Affluence Index. To appear in ASTIN Bulletin.

Chen, L., Cairns, A.J.G., and Kleinow, T. (2017) Small population bias and sampling effects in stochastic mortality modelling *European Actuarial Journal*, 7: 193-230.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A, and Cutler, D. (2016) The association between income and life expectancy in the United States, 2001-2014. *Journal of the American Medical Association,* 315: 1750-1766.

CHMD (2011) Canadian Human Mortality Database. `www.bdlc.umontreal.ca/chmd/`

Deary, I.J., Weiss, A., and Batty, G.D. (2010) *Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities*, Monograph, University of Edinburgh.

Egan, M., Daly, M., Delaney, L., Boyce, C. J., and Wood, A. M. (2017). Adolescent conscientiousness predicts lower lifetime unemployment. *Journal of Applied Psychology*, 102: 700-709.

Eurostat (2013) *Revision of the European standard population. Report of Eurostat's task force. 2013 edition.* Luxembourg: Publications Office of the European Union.

HMD (2018) Human Mortality Database: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de

Hyndman, R., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography*, 50: 261-283.

Kern, M.L., and Friedman, H.S. (2008) Do conscientious individuals live longer? A quantitative review *Health Psychology* 27: 505-512.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63: 147-152.

Lee, R.D., and Carter, L.R. (1992) Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association*, 87: 659-675.

Li, N., and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42: 575-594.

Longevity Science Panel (2018) Life expectancy: Is the socio-economic gap narrowing?

`www.longevitypanel.co.uk/viewpoint/life-expectancy-is-the-socio-economic-gap-narrowing/`

Macdonald, A.S., Richards, S.J., and Currie, I.D. (2018) *Modelling mortality with actuarial applications*. Cambridge University Press, Cambridge.

Mackenbach, J.P., Bos, V., Andersen, O., Cardano, M., Costa, G., Harding, S., Reid, A., Hemström, Ö., Valkonen, T., and Kunst, A.E. (2003). Widening socio-economic inequalities in mortality in six Western European countries. *International Journal of Epidemiology,* 32: 830-837.

Mackenbach, J.P., et al. (2016) Trends in inequalities in premature mortality: a study of 3.2 million deaths in 13 European countries. *Journal of Epidemiology and Community Health,* 69: 207–217.

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

Renshaw, A.E., and Haberman, S. (2003) Lee-Carter Mortality Forecasting with Age-Specific Enhancement. *Insurance: Mathematics and Economics*, 33: 255–72.

Vang, Z.M., Sigouin, J., Flenon, A., and Gagnon, A. (2017) Are immigrants healthier than native-born Canadians? A systematic review of the healthy immigrant effect in Canada. *Ethnicity and Health*, 22: 209-241.

A Model Selection Criteria Assessment

Desirable criteria for a model to satisfy are:

- BIC: Model has a low BIC score, although not necessarily the best.
- GD: Graphical diagnostic tests[27] all or mostly satisfactory. Conclusions will depend on the fit of each model to the specific datasets.
- Coh: Multi-population model should satisfy the principle of coherence. Assessment considers whether or not the model prevents mortality rates in two populations from diverging over time and is not dependent on the datasets used.
- Cross: Model should not impose in-year mortality curve crossovers where these are not apparent in the raw data. Assessment is partly dependent on the fit of the model to the specific data, and partly based on the potential for crossovers to arise in future scenarios.

[27] Graphical diagnostic tests (GD) offer a more informal alternative to more formal hypothesis tests: Figure 22 gives some examples. A GD will typically have an underlying hypothesis. If the hypothesis turns out to be valid then the GD should exhibit certain characteristics. For example, in the top left panel of Figure 22 the hypothesis is that the residuals should be independent of each other leading to a GD that should exhibit a random pattern of reds and blues. If the GD does not exhibit the anticipated characteristics (e.g. if the top left panel of Figure 22 had clear clusters of reds and blues) then it is likely that the underlying hypothesis is not true. Additionally, the characteristics that we do observe can point to how the model might be improved.

	BIC				Criterion				
	CPP Males	CPP Females	QPP Males	QPP Females	BIC	GD	Coh	Cross	Corr
M1	63936	61415	56265	52401	✗	✓	✗	✗	✓
M2	62106	59638	54389	50598	✗	✓	✗	✗	✓
M3	60466	57854	52588	48733	✗	✓	✗	✓	✗
M4	60411	57737	52515	48658	✗	✓?	✗	✗	✗
M5	60598	58051	52775	48957	✗	✓	✓	✓	✓
M6	58823	56277	51030	47185	✓	✓	✓	✓	✓
M7	60502	57770	52526	48750	✗	✗	✓	✗	✓
M8	58764	56000	50798	46921	✓	✗	✓	?	✓
M9	59484	56244	51168	46965	✓	✗	✓	✗?	✓
M10	58774	56113	50835	46939	✓	✗	✓	✗?	✓
M11	57747	54546	49493	45139	✓	✗	✓	✓	✗

Table 3: Assessment of model selection criteria for each model and dataset. BIC values should only be compared within the same column (same dataset). Ticks indicate that a specific model satisfies a particular criterion or scores well relative to other models.

- Corr: Does the model produce a plausible forward correlation term structure, $\rho(t, i, j, x_i, x_j)$, within a population, non-trivial correlations between ages; between populations, non-trivial correlations; lower correlations between a pair of ages in different populations than within the same population; lower correlations for ages that are further apart. Assessment is not dependent on the datasets used.

Our assessment of the models against these criteria, along with BIC values are presented in Table 3 with further detail on graphical diagnostics in Table 4. The nesting of models means that, for example, model M1 has a higher maximum log likelihood than other models, but it is heavily penalised for being over-parameterised leading to a poorer BIC than all of the other models.

Examples of graphical diagnostics are given in Figures 22 to 25. Standardised residuals are defined as

$$Z(i,t,x) = \left(D(i,t,x) - m(i,t,x)E(i,t,x)\right) / \sqrt{m(i,t,x)E(i,t,x)},$$

and should be approximately independent and identically distributed standard normal random variables if we have a good model. Figure 22 is very typical for the great majority of groups, populations and models. Figure 23 highlights a potential cohort effect (see, e.g., Cairns et al., 2009) that arises with most models for CPP males in Group 11 only.[28] Comparison of Figure 23 panels (c) and (d) reveals some

[28]In this context, cohort effects can arise when the individual groups still contain some residual

Model	QPP-males	QPP-females	CPP-males	CPP-females
M1	✓	✓	✓	✓
M2	✓	✓	✓	✓
M3	✓	✓	✓	✓
M4	?	✓	✓	✓
M5	✓	✓	✓	✓
M6	✓	✓	✓	✓
M7	✓	✓	✓✗	✗
M8	✓	✓	✓✗	✗
M9	✗	✓	✗	✗
M10	✗	✓	✗	✗
M11	✗	✓	✗	✗

Table 4: Graphical diagnostic results for all underlying models. For CPP datasets group 11 demonstrates significant non-random pattern in standardized residuals under all models (see, for example, Figure 23). Therefore any model that behaves well for group 1 to 10 and not obviously worse than other models for group 11 is marked as ✓.

similarity in the pattern from about 1918 onwards. Panel (c) is a standard plot of the residuals while panel (d) shows the proportion of each cohort by year of birth in Group 11 (dots at different ages, red line at age 65 only for younger cohorts). In particular, given that Group 11 is the top group, if it is smaller (so more concentrated on the most sustained high earners) then we might expect (even) lower mortality than would be the case if the proportion in Group 11 remained the same from year to year.

For these datasets (especially the age range of 65-89) we could equally opt for M8 rather than M6. However, our reasons for preferring M6 over, say, M8 are:

- M6 extends more easily to younger ages whereas the linear age-period effects in M8 lead to a poorer fit over a wider age range.
- The linear age-period effects also lead to minor cross-over problems in some years. This is implicit in a comparison of M6 and M8 for Group 1 in Figures 24 (M6) versus 25 (M8). In Figure 25(a) there are are clusters of red cells in the later years at high and low ages. At the high ages this causes a cross over that is not evident in the data in these later years.

We can also compare M6 with M5. The difference between the two is that M5 has group-specific base tables, $\alpha(i, x)$, rather than the country-specific $\alpha(x)$. The significant difference in BIC's indicates that the additional complexity in M5 does

degree of heterogeneity. If the balance between "sub-groups" changes by cohort then this has an impact on levels of mortality by cohort.

not result in a sufficiently large improvement in the fit. For M6, the $\kappa(i,t)$ then dictate group-specific deviations from the national base table, while in M5 they dictate deviations from the group-specific base table.

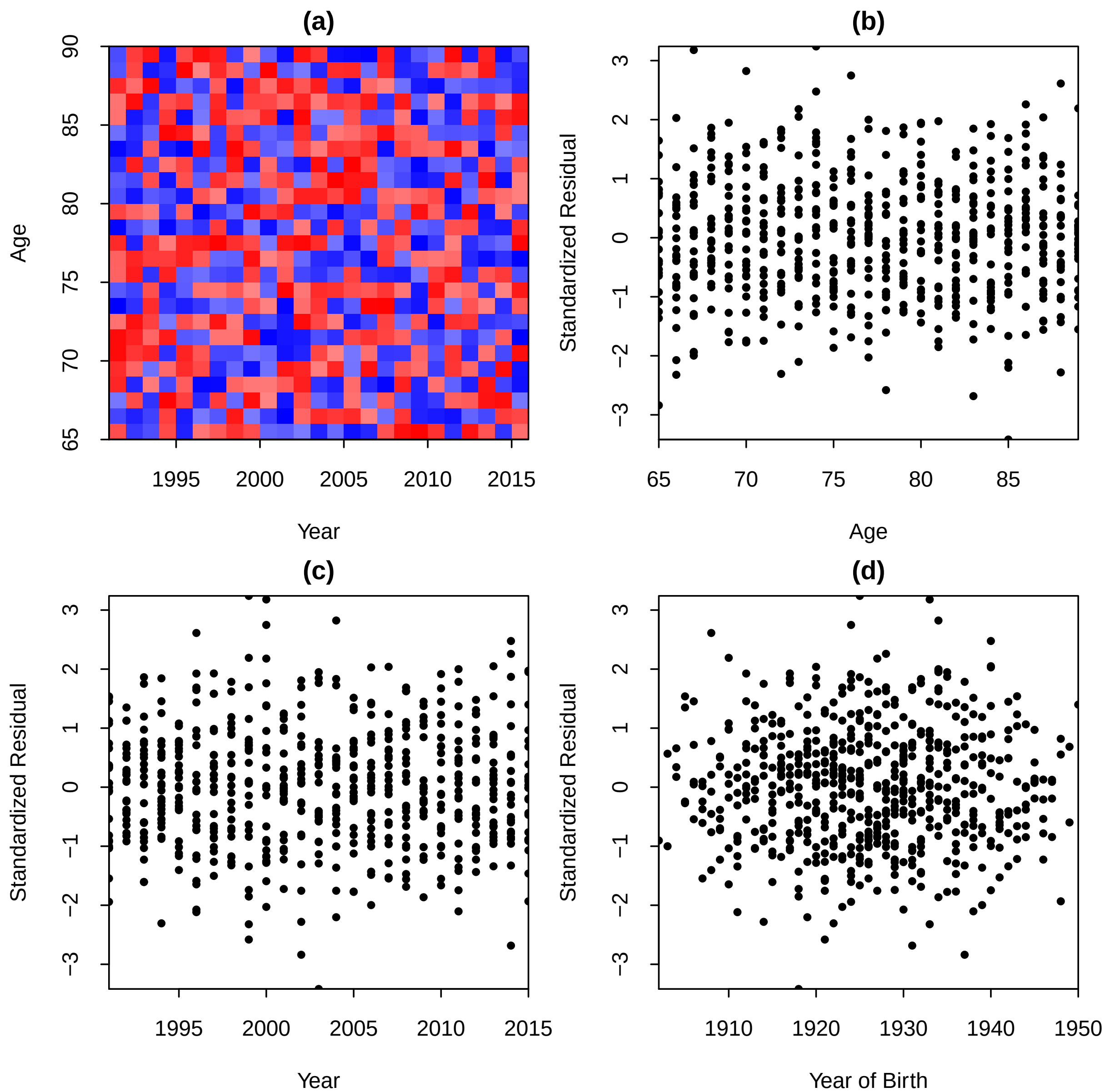

Figure 22: Model M6 graphical diagnostics for CPP males Group 6 using standardised residuals, $Z(i, t, x)$. (a) heat map of the $Z(i, t, x)$ (red - positive; blue - negative). (b) residuals by age. (c) residuals by calendar year. (d) residuals by year of birth.

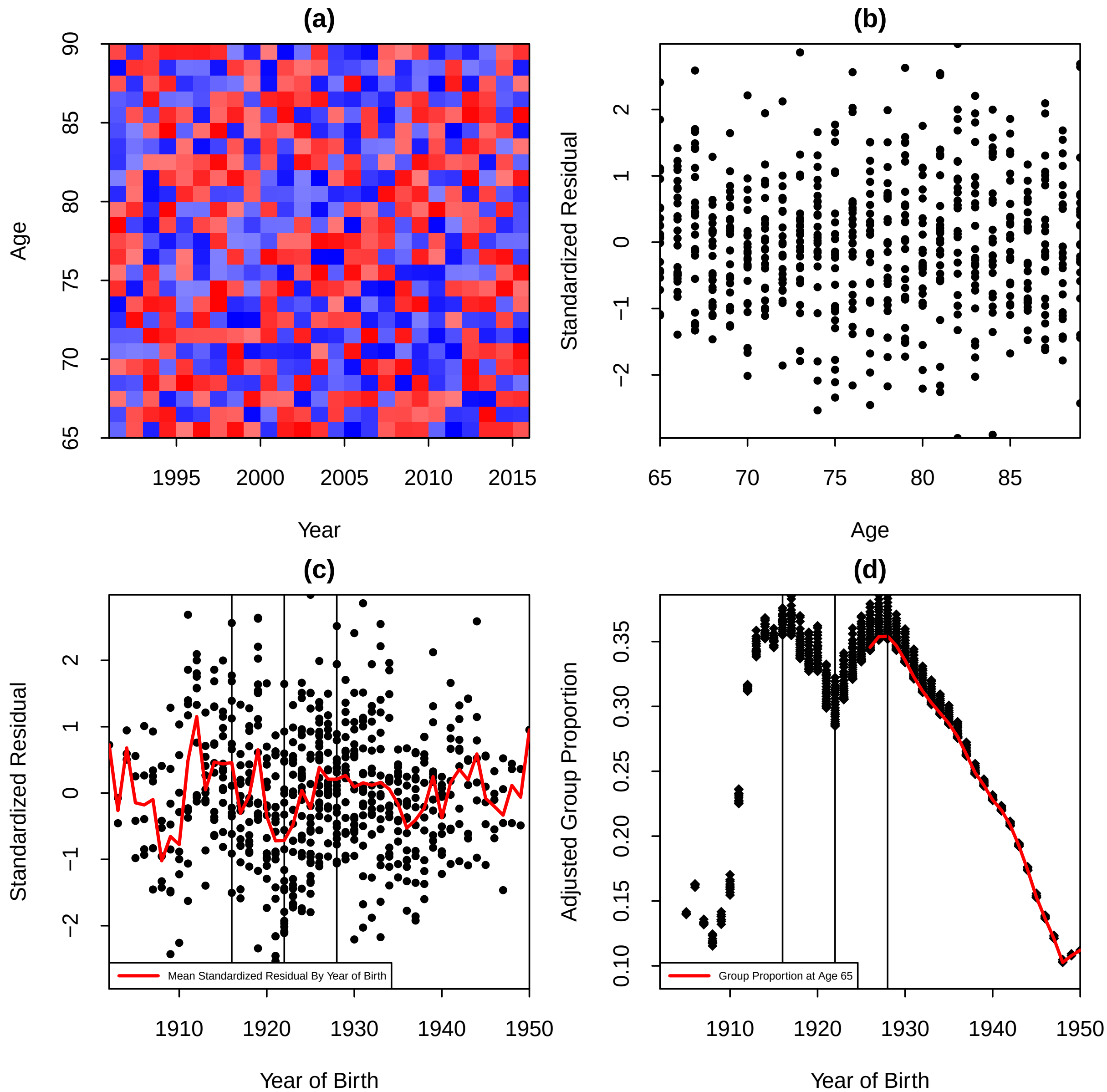

Figure 23: Model M6 graphical diagnostics for CPP males Group 11 using standardised residuals, $Z(i,t,x)$. (a) heat map of the $Z(i,t,x)$ (red - positive; blue - negative). (b) residuals by age. (c) residuals by year of birth. (d) proportion of total (t,x) exposures in Group 11 by cohort (red line: proportions at age 65 by cohort) with linear adjustment.

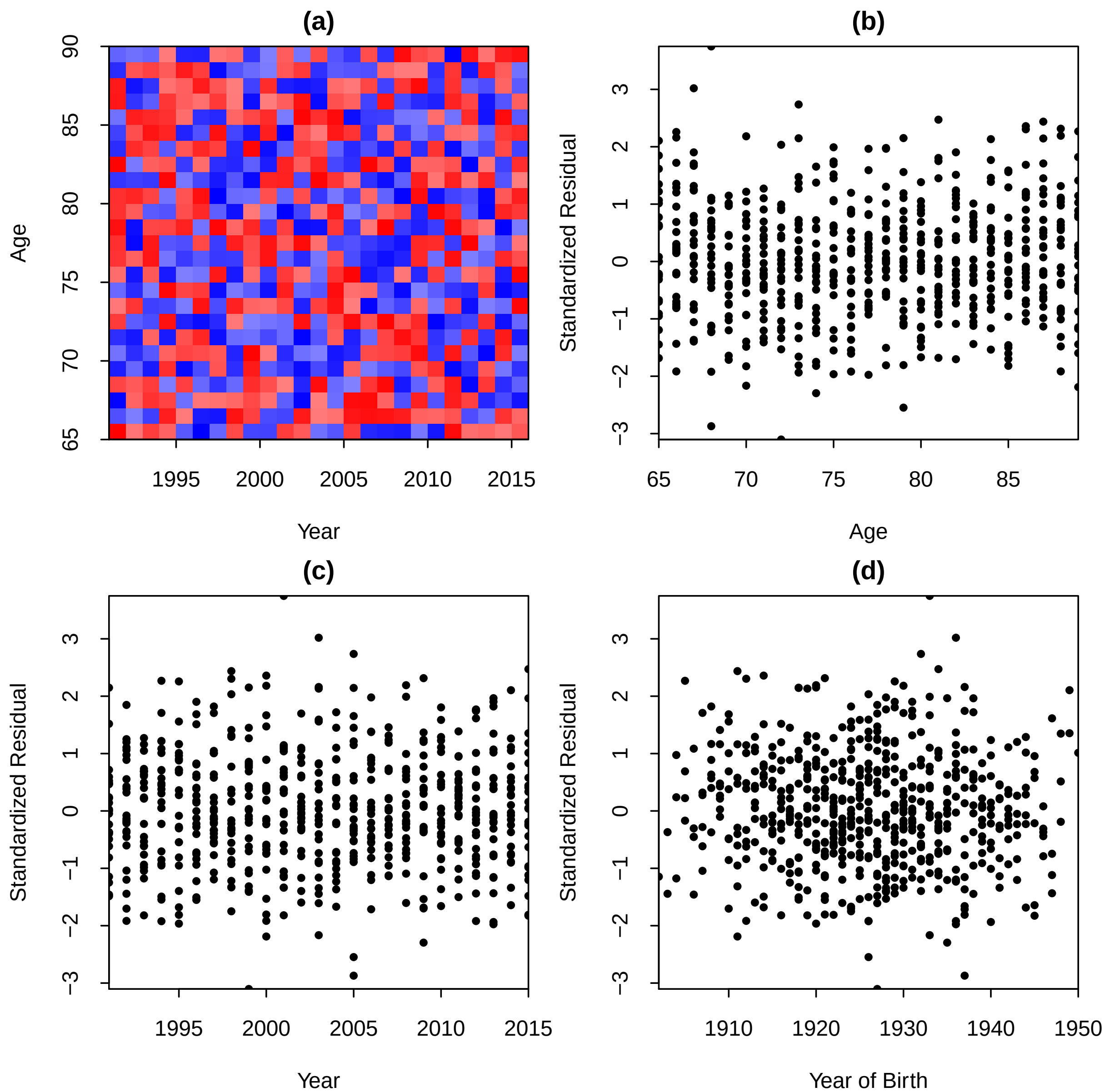

Figure 24: Model M6 graphical diagnostics for CPP males Group 1 using standardised residuals, $Z(i, t, x)$. (a) heat map of the $Z(i, t, x)$ (red - positive; blue - negative). (b) residuals by age. (c) residuals by calendar year. (d) residuals by year of birth.

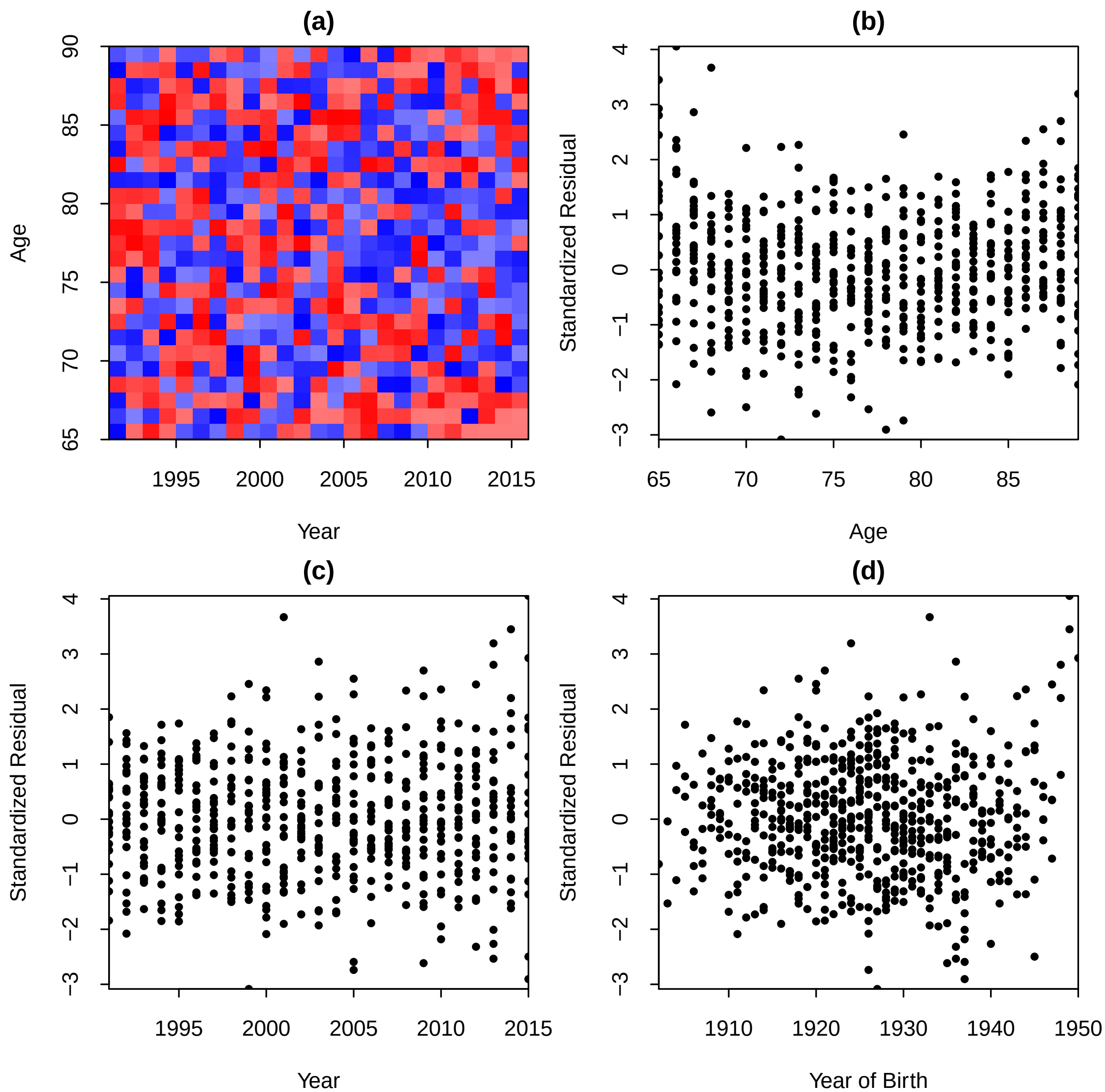

Figure 25: Model M8 graphical diagnostics for CPP males Group 1 using standardised residuals, $Z(i,t,x)$. (a) heat map of the $Z(i,t,x)$ (red - positive; blue - negative). (b) residuals by age. (c) residuals by calendar year. (d) residuals by year of birth.

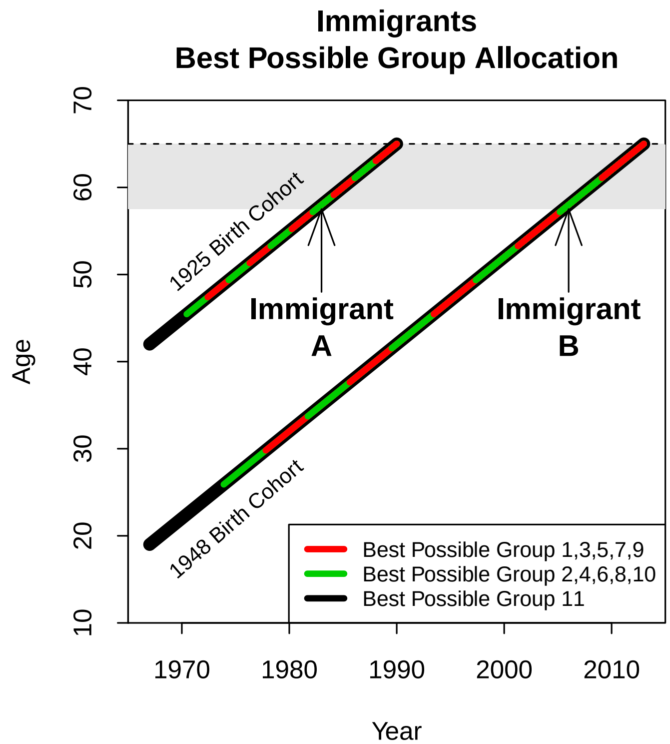

Figure 26: Attainability of different pension levels by cohort for immigrants at different ages. Immigrants A and B both arrive in Canada at age 58. If immigrant A (1925 birth cohort) earns consistently above the YPME then he/she will still only be in Group 4 on retirement at 65. Immigrant B (1948 birth cohort) cannot do better than Group 2.

B Growing Proportion of Immigrants in Low Groups

Figure 26 shows how, as the plan has matured, the balance of immigrants in the pensioners population will have shifted towards lower groups. Consider two immigrants A and B who both migrated to Canada at age 58. A belongs to the 1925 birth cohort, B to the 1948 cohort. If A earns above the YMPE from age 58 to 64 then he/she can retire at age 65 with a CPP pension that will be just under 40% of the maximum (Group 4) but no higher without deferring retirement. In contrast, immigrant B cannot do better than Group 2.

From the 1925 cohort, migration to Canada in their early 40's would be sufficient to make attainment of the maximum pension possible. From the 1948 cohort, migration to Canada would have had to be in their early 20's.