Members' Paper

Key Factors for Explaining Differences in Canadian Pensioner Baseline Mortality

By

Seyed Saeed Ahmadi, PhD

Richard Brown, FCIA, FSA, CFA

Any opinions expressed in this paper are those of the authors and are not necessarily the opinion or position of the Canadian Institute of Actuaries.

Les opinions exprimées dans cette communication sont celles des auteurs et ne sont pas nécessairement les opinions ou politiques de l'Institut canadien des actuaires.

May 2018

Document 218049

Key Factors for Explaining Differences in Canadian Pensioner Baseline Mortality

By

Seyed Saeed Ahmadi, PhD

Richard Brown, FCIA, FSA, CFA

Abstract	**Résumé**
In this paper, we further the research published by the Canadian Institute of Actuaries in the Final Report of Canadian Pensioners' Mortality study, CPM, (2014), and consider new factors that explain differences in Canadian pensioner baseline mortality. We explore many of the factors that were found to explain UK pensioner baseline mortality in Madrigal, et al., (2011). The factors we explore include the following: health condition at retirement, affluence, geographical information together with socio-economic factors (i.e., geodemographics), occupation, industry, and public versus private sector, among others. We analyzed mortality data for defined benefit pensioners collected by Club Vita Canada directly from pension plan and post-retirement benefit plan sponsors or their administrators. The data analyzed covers approximately 1.4 million exposure years and 38,000 deaths for pensioners and survivors during the calibration period of 2012–2014. We found that retirement health and geodemographic information (via postal code) were some of the most important drivers of variations in baseline mortality, along with pensioner type (i.e., pensioners versus surviving spouse), affluence (via	Dans le présent document, nous poursuivons la recherche publiée par l'Institut canadien des actuaires dans le Rapport final de l'étude de la mortalité des retraités canadiens, CPM, (2014), et nous analysons de nouveaux facteurs qui expliquent les différences au chapitre de la mortalité de base des retraités canadiens. Nous examinons bon nombre des facteurs réputés expliquer la mortalité de base des retraités du Royaume-Uni, dans Madrigal, et al., (2011). Les facteurs que nous étudions comprennent, entre autres : l'état de santé à la retraite, la richesse, les données géographiques jumelées à des facteurs socioéconomiques (c.-à-d. les caractéristiques géodémographiques), la profession, le secteur d'activité, et le secteur public ou privé. Nous avons analysé les données sur la mortalité des retraités touchant des prestations déterminées. Ces données sont recueillies par le Club Vita Canada directement auprès des promoteurs de régimes de retraite et des régimes de prestations à la retraite, ou de leurs administrateurs. Les données analysées portent sur environ 1,4 million d'années d'exposition et 38 000 décès de retraités et de survivants pendant la période d'étalonnage comprise entre 2012 et 2014. Nous avons constaté que l'information sur la santé à la retraite et les données géodémographiques (par code postal) comptaient parmi les vecteurs les plus importants de la variation de la mortalité de

salary or pension amount), and occupation. In contrast to current Canadian pension industry practices, whether pensioners were part of a public sector or private sector plan was found not to be a significant factor in the context of other available factors.

After determining the key factors, we model Canadian baseline post-retirement mortality for defined benefit pension plan members by applying a generalized linear modelling framework. The models created explain variations in period life expectancy at age 65, ranging nine years for male pensioners compared to CPM, (2014) which explained just under four years when accounting for sector and pension size.

We found that by capturing the baseline mortality characteristics of individual plan members, plans could develop more accurate baseline mortality assumptions, as supported by closer alignment with their own experience. A focus on individual plan member factors also allows longevity characteristics to be uncovered that cannot be identified through actual-over-expected mortality experience analysis (e.g., due to differences in the profile of current pensioners and active members).

The results presented in this paper are directly applicable to Canadian defined benefit pension and post-retirement benefit plans in the assessment of baseline mortality and longevity risk, and the creation of assumptions for actuarial valuations for funding and financial statement purposes, as well as pricing and valuing of insured group annuitant portfolios by insurance and reinsurance companies.

base, tout comme le type de retraité (c.-à-d. retraité ou conjoint survivant), la richesse (au moyen du salaire ou du montant de la rente) et la profession. Nous avons constaté que contrairement aux pratiques actuelles du secteur canadien des régimes de retraite, la question de savoir si les retraités participaient à un régime public ou privé ne constitue pas un facteur important dans le contexte des autres facteurs disponibles.

Après avoir déterminé les facteurs clés, nous modélisons la mortalité canadienne de de base postérieure à la retraite pour les participants de régimes de retraite à prestations déterminées en appliquant un cadre de modélisation linéaire généralisée. Les modèles créés expliquent les variations de l'espérance de vie à 65 ans, soit neuf ans pour les hommes retraités comparativement à CPM, (2014), qui est ramenée à un peu moins de quatre ans si l'on tient compte du secteur et de l'ampleur de la rente.

Nous avons également constaté qu'en saisissant les caractéristiques de base de la mortalité des participants individuels, les régimes pouvaient élaborer des hypothèses plus précises de la mortalité de base, comme l'indique l'harmonisation plus étroite avec leur propre expérience. L'accent mis sur les facteurs propres à chaque participant à un régime permet également de découvrir les caractéristiques de la longévité qui ne peuvent être décelées au moyen de l'analyse des résultats de mortalité réelle par rapport à prévue (p. ex., en raison des différences dans le profil des retraités actuels et des participants actifs).

Les résultats affichés dans le présent document s'appliquent directement aux régimes de retraite à prestations déterminées et aux régimes canadiens de prestations postérieures à la retraite dans le cadre de l'évaluation du risque de mortalité et de longévité de base, à la

	création d'hypothèses pour les évaluations actuarielles aux fins du provisionnement et de la préparation des états financiers, de même qu'à la tarification et à l'évaluation des portefeuilles de rentes collectives des assurés par les sociétés d'assurances et de réassurance.

Table of Contents

1. Introduction

Organizations that sponsor pension and post-retirement plans for their current and former employees have become increasingly aware of the risk and increased costs posed by their members continuing to outlive previous longevity expectations. The uncertainty about how long plan members will live exposes pension and post-retirement benefit plan sponsors to longevity risk. The key components of that longevity risk can be deconstructed as follows:

- Idiosyncratic risk – may also be known as individual risk or binomial risk, and relates to the variations in mortality experience within a specific population compared to mortality expectations. Idiosyncratic risk is reduced as the size of the population increases because the mortality patterns of the plan population become more stable.
- Measurement risk – stems from the difficulty in measuring baseline mortality (i.e., current mortality rates) given the large amounts of data required, and the uncertainty about how well baseline mortality expectations represent the expectations of different populations. Large plans will often perform mortality studies to assess how their plan's mortality experience compares to that of a published mortality table, and may apply adjustment factors to attempt to mitigate measurement risk.
- Trend risk – represents the fact that mortality rates are expected to change in the future, and particularly that they will likely continue to decrease; however, there is significant uncertainty regarding the pace at which mortality rates will decrease. The only way for a plan to truly mitigate trend risk is to hedge that risk, such as through a longevity insurance or swap transaction, or a buy-in or buyout annuity.

Measurement risk can be mitigated through better measurement of baseline mortality. In this paper, we focus on identifying the factors that best explain variations in baseline mortality and then build models that allow baseline mortality assumptions to be tailored to the mortality characteristics of specific groups of individuals (e.g., pension plan or group annuitants). We do this by collecting and analyzing a large volume of recent pension plan mortality experience data that includes information regarding a wide range of mortality factors. We employ comprehensive statistical modelling to identify those factors that are most predictive of Canadian pensioner baseline mortality, and use generalized linear modelling (GLM) to create baseline mortality assumptions that capture the influence of the different factors.

When valuing pension and post-retirement obligations, most Canadian plans currently use one of the baseline mortality tables published in CPM, (2014). The CPM study published three sets of gender-specific baseline mortality tables—namely the combined, private, and public—and included a set of pension amount size adjustment factors to allow users to adjust the published tables for differences in baseline mortality expectations by pension size. The CPM study provided significant improvements upon past published baseline mortality studies, in particular:

- It was the first comprehensive published study covering Canadian defined benefit pensioners (i.e., past studies had predominantly consisted of US pensioners); and
- The introduction of public versus private sector tables, and pension size adjustment factors provided the ability for plans to begin to customize baseline mortality assumptions to their plan members' specific characteristics.

The CPM study was not without its challenges and the final report notes that while the researchers would have liked to investigate other factors (e.g., socio-economic ones), limitations of the data provided prevented such an investigation.

This paper therefore aims to build upon the findings of the CPM study through the investigation of a wider range of mortality factors. To do this we follow Madrigal, et al., (2011) and consider a comprehensive series of mortality rating factors including age, gender, health condition at retirement, affluence (via pension or salary), geographical information together with socio-economic factors (geo-demographics), occupation, industry, and public versus private, among others. We apply GLM in a multivariate framework to fit the baseline mortality. The GLM approach provides substantial flexibility not only to analyze the effect of each factor individually and assess its importance on baseline mortality, but also evaluate their interaction (e.g., with age) simultaneously. We believe our approach of constructing baseline mortality rates using a range of mortality factors in a multivariate framework

- Enables plans to develop baseline mortality assumptions that are much more representative of their plan members by accounting for individual member characteristics.
- Reduces the reliance on plan-level actual over expected adjustments (which are generally only possible for very large plans) when standard mortality tables are suspected not to be representative based on analysis of credible historical plan-specific mortality experience.
- Allows plans without credible mortality experience the ability to estimate baseline mortality for their plan membership more accurately.
- Accounts for how the mortality profile of a plan's membership influences its projected benefit payments (i.e., how the shape of mortality with age varies based on different mortality factors), enabling better matching to projected asset cash flows in a liability-driven investment strategy.
- Allows mortality assumptions to capture changes in the mortality characteristics of different generations of plan members (e.g., a plan's active members may have very different characteristics, and therefore life expectancies, than the pensioner population that worked 30 years ago).

The main contributions of this paper can be summarized as follows:

- We have analyzed pensioner and survivor mortality data between 2012 and 2014 which provides an update compared to the CPM study's collection period of 1999 to 2008.
- We have examined a wide set of mortality rating factors in the modelling of baseline defined benefit pensioner mortality in Canada, which—to the best of our knowledge—have not been investigated previously in published research.
- We determined the importance of postal code based lifestyle grouping in our analysis.
- We found that the differentiation of public sector versus private sector employment is not statistically significant when considering other explanatory rating factors.
- Rather than the traditional plan-specific approach that adjusts standard tables for plan experience, we are providing an alternative member-focused approach by creating sets of mortality rates for subgroups of plan members based on key rating factors that show statistically significant differences in baseline mortality. These mortality tables can be readily

used to map each plan member independently (regardless of being in a large or small plan) based on their individual characteristics. This approach elevates the ability to customize baseline mortality assumptions for specific Canadian pension plan membership populations compared to the current status quo, which primarily considered only public vs. private sector and pension size.

This paper is organized as follows:

- Section 2 provides a summary of the data analyzed including the factors investigated, definitions of age, exposure and death, data validations performed, and how we classified data.
- Section 3 provides a brief review of the generalized linear model and methodology that we use to develop models for baseline mortality.
- Sections 4 and 5 illustrate our process for developing our baseline mortality models, including a detailed review of all the statistical and actuarial tests performed.
- Sections 6 and 7 present adjustments and extensions of the final models.
- Sections 8 and 9 investigate the performance of the final models.
- Section 10 compares our results to those of the CPM study.
- Section 11 summarizes our key findings.

2. Data Set and Data Preparation

The data used to determine the key rating factors that predict Canadian pensioner baseline mortality and to perform the modelling outlined in this paper were compiled by Club Vita Canada, a subsidiary of Eckler. The data set consists of a cross section of Canadian defined benefit pension plan members across the country and industries, and includes both private and public sector plans. In total, the data set analyzed consisted of data contributed from 34 plans.

The death experience data focuses on post-retirement ages only, as pre-retirement mortality is typically of a much lesser concern due to the nature of pre-retirement death benefits. We have analyzed post-retirement mortality separately for males and females, and also differentiate between pensioners and survivors—this is in contrast to the CPM study, which excluded survivors. Given that gender and pensioner type (i.e., original plan member pensioner versus surviving spouse) represent distinct groups, these factors were used to stratify the data into the following four strata:

- Male pensioners;
- Female pensioners;
- Male survivors; and
- Female survivors

We further stratify male and female pensioners by retirement health type, with disabled pensioners being referred to as “ill health” and non-disabled pensioners as “non-ill health”.

The data set covers mortality experience over the calendar years 2012–2014. While many Club Vita Canada member plans have provided historical mortality experience prior to 2012 and after 2014, we have used the three-year period for the following reasons:

- Baseline mortality can and should be measured objectively, whereas future mortality trends (i.e., how mortality varies over time) are inherently subjective due to the uncertainty of how future mortality trends will compare to the past. Given that baseline mortality represents the rates of death that are currently being experienced, ideally only a single calendar year would be used. However, even with very large data sets, mortality experience can vary materially year-to-year across ages and factors. Expanding the range beyond a single year then begins to combine the impact of baseline mortality and mortality trends.
- The longer the period used to assess baseline mortality, the more difficult it will be to identify changes in baseline mortality over time.
- 2014 was selected as the final year for our analysis since it was the latest year where complete and reliable mortality experience was available for all plans.

Considering the points above, we follow Madrigal, et al., (2011) and use a three-year range which we believe provides an appropriate balance. The middle year of the mortality experience analyzed is 2013, and therefore this can be viewed as the base year for the resulting mortality models.

The remainder of this section

- Outlines the primary factors we have investigated;
- Describes how age, deaths, and exposures have been calculated;
- Provides more information on our data validation and quality assessment process;
- Outlines how certain data fields have been classified for modelling purposes; and
- Shows summary statistics for the data set.

2.1 Covariates

The following table outlines all the factors (i.e., covariates) that were investigated and their rationale for inclusion. The covariates investigated were largely selected because of the plan administrators' ability to extract the information from their administration systems and because of evidence of their influence on mortality based on existing research or observed differences in mortality from available data. While we believe the list is comprehensive for the type of data available in pension administration systems, other factors exist but are mainly unattainable (e.g., smoking status).

Table 1

Investigated covariates and their rational for inclusion	
Covariate	**Rationale for inclusion**
Age	Death rates are highly dependent on age.
Gender	Gender has been shown by numerous studies to be one of the most significant longevity differentiators.
Pensioner type	Various research and studies have shown that survivors exhibit different mortality patterns than pensioners, with this behaviour being attributed to the so-called grieving widow(er)s effect. For example, Sullivan & Fenelon, (2014) found that "Becoming widowed is associated with a 48%

Investigated covariates and their rational for inclusion	
Covariate	**Rationale for inclusion**
	increase in risk of mortality".
Retirement health type	The Society of Actuaries' (SOA) RP-2014 Mortality Tables Report, (2014) found that disabled retirements had higher rates of mortality compared to non-disabled retirements.
Pension amount	The CPM study found that the relative level of mortality decreased with pension size.
Pre-retirement earnings (i.e., salary amount)	Madrigal, et al., (2011) found that pre-retirement earnings generally outperform pension amount at differentiating mortality.
Occupation	The SOA's RP-2014 study found that pensioners formerly employed in blue collar occupations exhibited higher levels of mortality than those who had white collar occupations.
Sector	The CPM study found a meaningful difference in mortality for public sector versus private sector pensioners.
Industry	The CPM study found that relative levels of mortality varied by industry.
Province	The Canadian Human Mortality Database highlights that mortality has varied by province.
Postal district	Further to mortality differences by province, Canada can be divided into smaller geographic regions by looking at the first character of postal codes (i.e., the postal district).
Urban versus rural	Another potential geographic mortality differentiator is whether pensioners live in an urban versus rural setting, as access to healthcare and lifestyle behaviour may vary between urban and rural populations.
Postal code	There is much research supporting variations in mortality by geographic regions smaller than province. Further, postal code can be used to proxy socio-economics factors such as level of education and household income, or a combination of factors via geodemographic segments.
Year of exposure	Given that mortality varies with time, it is natural to consider the impact of year of exposure.
Month or season of birth	The motivation for including month or season of birth stems from capturing any potential advantage (or disadvantage) gained from being born during a particular part of a year. For example, developmental advantages from being born early in the year.

2.2 Age, death, and exposure

Age is obviously a critical factor in any mortality analysis. Similarly, the definitions of deaths and exposures are also very important. The remainder of this section explains our approach to the measurement of age, death, and exposure.

For practical reasons, we convert continuous ages to integer ages for modelling purposes. Different age definitions are available to do this—that is, age last birthday, age nearest birthday, or age next birthday. We have selected the age nearest birthday definition (i.e., rounding the age of each member up or down to an integer number depending on his/her closest birthday), as we believe it is the most common definition pension actuaries use.

Since we are analyzing mortality experience over three years, individual pensioners and survivors could have up to three records (i.e., one for each of the exposure years). For each record, the proportion of a year that an individual i, age x is exposed to the risk of death, noted as ETR_{xi}, is determined based on the following:

1. The later of January 1 of the exposure year and the date benefits commenced; and
2. The earlier of December 31 of the exposure year and the date benefits ceased.

The ETR_{xi} is calculated as the number of days in the exposure year between (2) and (1) as defined above, divided by the total number of days in the year. If the member died during the year of exposure, the ETR_{xi} is set to 1.

At the beginning of each exposure year, a member can have already been exposed to the risk of death (i.e., benefits commenced in a prior year), become exposed during that calendar year, or not become exposed until a later year. By the end of each exposure year, a member can still be under the risk of death or have already exited the plan. Note that while pensioners and survivors will primarily exit due to death, it is also possible to see exits due to reasons other than death (e.g., pension benefits transferred to an insurer).

2.3 Data validation and quality assessment

Ideally, we would have complete and reliable data for every member for each plan. However, data is generally not complete for all plans and contains some degree of missing and/or suspicious data. Further, the period for which mortality experience data is available will vary by plan. To assess the completeness and reliability of the data, Club Vita Canada has performed an extensive data validation and quality assessment process for each plan. This process culminates in measures of data quality for each record (referred to as quality flags) and each plan.

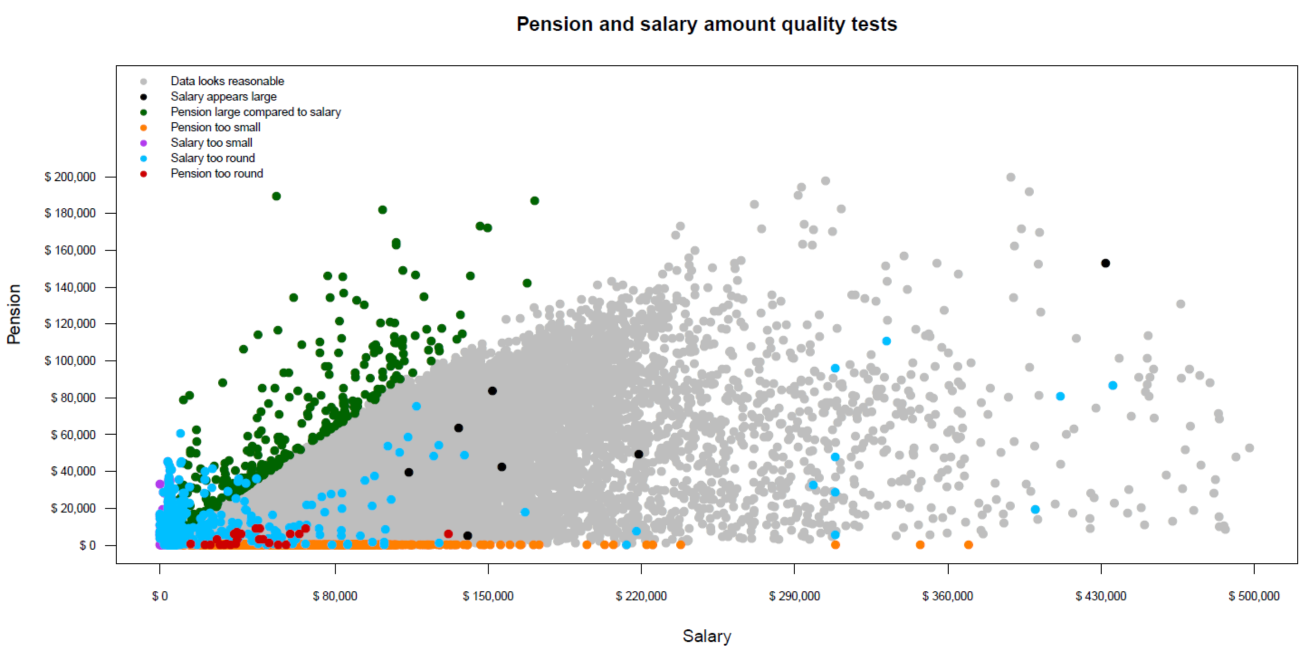
Our full data validation and quality assessment process is beyond the scope of this paper; however, we briefly outline our process for assigning quality flags due to its importance in the context of our modelling. Our process consists of three main phases as follows:

1. Individual tests: Approximately 100 data validation tests are performed on each record to identify data issues. Tests primarily consist of those assessing the overall reliability of the record (i.e., missing or inconsistent dates) or the reliability of covariates (e.g., missing, suspiciously high, or suspiciously low pension amount).

2. Individual-level quality flags: Individual records are assigned quality flags based on the validation tests in phase 1, with the quality flag values being either Good or Bad. Different individual-level quality flags are determined at the overall record level and for different covariates.
3. Stratum-level quality flag: The data for each plan is divided into four strata, which are based on the combination of gender and pensioner type. Stratum-level quality flags are set to Bad if the underlying individual-level quality flags exceed certain thresholds. The stratum-level quality flags involve looking at both live and deceased records separately, as well as different age ranges to assess whether Bad data introduces any statistical biases.

Figure **1** shows the results of individual tests performed on pension and salary amounts. In this analysis, we have excluded those cases where data is missing, and instead are focused on identifying suspicious data. Some degree of suspicious data (e.g., very small pension amounts) is expected for each plan; however, we are trying to identify any systematic data issues that could bias our analysis. Overall, we can see that the majority of the data looks reasonable (gray solid circles), while there are some records with suspicious characteristics.

Figure 1



2.4 Data classification

To facilitate the application of the models we develop to any plan member, we fit our models only using categorical covariates. Therefore, a set of mortality rates can be relatively easily assigned to any plan member based on their longevity profile. Alternatively, continuous variables could be used—particularly for pension and salary amounts, as these are inherently continuous random variables. While it's possible to fit models using pension and salary in the continuous form, we do not believe this is practical when implementing the models for actuarial work (i.e., it is preferable to have a finite number of mortality curves). Therefore, we classify pension and salary amounts into bands. Before dividing pensions and salaries into bands, we first need to ensure that they are consistently measured.

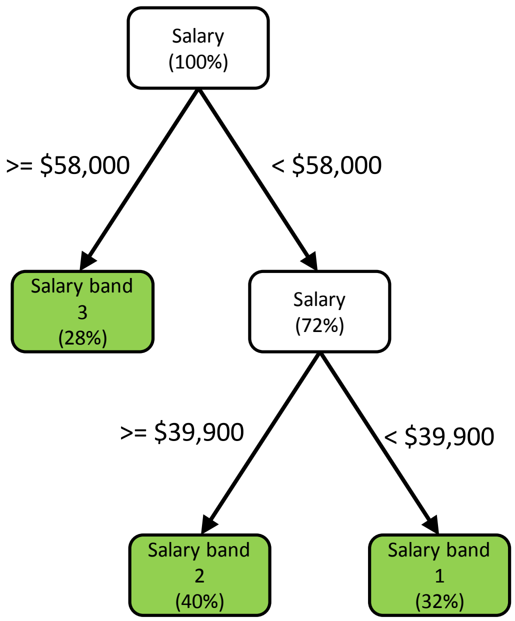
The approach we have used is to revalue pensions and salaries to a common revaluation date (in our case, December 2015). This is done by adjusting actual pension amounts by the difference in the value of the Consumer Price Index (CPI) at the revaluation date and the effective date of the pension amount. For salary amounts, the data provided is the salary the pensioner earned prior to leaving employment. Like pensions, we revalue salaries with CPI from the date the member left employment to the revaluation date to normalize the purchasing power of all salary amounts.

In the CPM study, pension bands were set based on $500 increments of monthly pension amounts up to $6,000. While evenly distributed bands ease implementation, such a selection of bands is subjective. To minimize subjective judgment, we employed statistical clustering analysis to develop pension and salary bands. To do this, we first divided records into many initial bands with a minimum level of exposures in each band. The standardized mortality ratio (SMR)[1] was then calculated for each initial band to assess the relative level of mortality while standardizing for age differences between bands (e.g., due to smaller pension amounts being more prevalent at older ages). We then follow Breiman, et al., (1984) and use the recursive partitioning and regression trees approach to cluster the initial pension and salary bands into an optimal number of discrete bands. To evaluate the size of the tree (i.e., the optimal number of bands), we analyzed the reduction in the relative error obtained by applying a 10-fold cross-validation as the tree grows. This means that the data is first divided into 10 equal bands and the tree is fitted using nine bands. The error is then obtained based on the performance of the model on the 10th segment. This procedure is repeated on all bands and the results are averaged and scaled to provide a cross-validation relative error. Figure 2 illustrates the constructed regression tree using salary amount information for female pensioners. Two breakpoints ($39,900 and $58,000) are identified to create three salary bands. The percentage of exposures within each salary band is given in parentheses.

[1] Ratio of observed number of deaths for the band, to the expected number of deaths based on all bands combined. The deaths are weighted based on the distribution of exposures for all bands combined to minimize bias introduced by variations in the age distribution for each band.

Figure 2

Regression tree on salary amount for female pensioners



Our clustering process to create pension bands (based on annual pension amount) is performed separately for each of our four strata. For salary bands, the clustering process is performed only for each of male and female pensioners, as salaries are not available for survivors.

Another mortality rating factor that requires clustering is postal code. Since there are approximately 900,000 Canadian postal codes, using them directly for mortality analysis is impractical. Therefore, instead of using postal codes directly, they can be linked to geodemographic segments which represent Canadians with different lifestyles (e.g., eating and exercise habits), socio-economic characteristics (e.g., affluence and level of education), and geographic differences (e.g., urban or rural). For our modelling, we have utilized a third-party geodemographic segmentation system that divides Canadians into about 70 different segments. With this as a starting point, we perform a similar clustering analysis as that described above for pension and salary bands. However, instead of clustering based on the SMR, we have used the crude life expectancy (calculated using method (II) in Chiang, (1984)) for each geodemographic segment. By using different geodemographic segments, our pensioner population can be successively divided into homogeneous and relatively stable bands within different age groups. This enables us to estimate life expectancy at each geodemographic segment, which will be used in our clustering analysis. We performed our clustering separately for male and female pensioners. The result is what we refer to as postal code-based lifestyle groups or postal code-based longevity groups. We refer interested readers to Breiman, et al., (1984) and Ripley, (1996) for further discussion on this classification approach.

Care is needed when carrying out the pension and salary bands and longevity group clustering outlined above. These bands and longevity groupings should be homogenous and contain reasonable exposure sizes, and should also be distinct enough to explain different variations in baseline mortality.

We tested our pension and salary bands and longevity groups to make sure that they are statistically different from each other. In particular, we apply the following criteria:

- Each salary/pension band and longevity group should contain at least 5 percent of the exposures.
- Crude life expectancy at age 65 for each consecutive band or longevity group should be different by at least half a year.
- Crude life expectancy at age 65 for each consecutive band or longevity group should be statistically different at a 95 percent confidence level.

Using the above criteria, we created five longevity groups for male and female pensioners. Table 2 provides a summary of the pension and salary bands.

Table 2

Summary of pension and salary bands by stratum						
Stratum	**Pension bands**	**Median annual pension**	**% of total stratum exposures**	**Salary bands**	**Median salary**	**% of total stratum exposures**
Male pensioners	<$16,400 $16,400–$34,000 >$34,000	$8,573 $24,065 $44,320	37% 36% 27%	<$51,000 $51,000–$63,800 $63,800–$88,000 >$88,000	$44,586 $56,420 $75,570 $100,775	30% 19% 29% 22%
Female pensioners	<$13,100 $13,100–$34,400 >$34,400	$6,211 $19,389 $41,915	59% 34% 7%	<$39,900 $39,900–$58,000 >$58,000	$34,782 $46,179 $74,783	32% 40% 28%
Male survivors	<$11,100 >=$11,100	$4,586 $15,474	79% 21%	n/a	n/a	n/a
Female survivors	<$15,100 >=$15,100	$7,173 $21,310	70% 30%	n/a	n/a	n/a

Of the clustering performed to determine the pension bands, salary bands, and longevity groups, pension bands proved to be the most challenging due to a lack of clear differentiation of SMR for small pension amount clusters. This was not unexpected, since pension is being used to proxy the influence of affluence on mortality, but those with small pensions will include a mix of individuals with low affluence and long service, and high affluence and short service. It is not until pension amounts reach high levels that lower levels of mortality become apparent.

We observed an increasing trend in SMR for female pensioners with low pension amounts followed by a gradual decline as pension increases. As a result, the first female pensioner pension band includes almost 60 percent of exposures. Salary bands provided more differentiation compared to pension bands, particularly for male pensioners.

2.5 Data summary

The data used in the modelling covered by this paper comes from a diverse range of plans of different sizes. Table 3 summarizes the number of plans by the size of their pensioner and survivor populations.

Table 3

Number of plans by size of pensioner and survivor population	
Number of pensioners and survivors combined as at December 31, 2014	**Number of plans**
Less than 2,500	15
2,500–9,999	8
10,000–29,999	4
30,000 or more	7
Total	**34**

Naturally, those plans with larger pensioner and survivor populations will have a correspondingly larger influence on the data used in our modelling. Therefore, it was critical to ensure the largest plans consisted of a diverse range of members in respect to affluence, occupation, and geography in order to minimize the risk of introducing bias into our modelling. In contrast, more homogeneous populations are acceptable for smaller plans. Given that it is currently common within the Canadian pension industry to differentiate mortality based on whether a plan covers private or public sector employees, it is worth noting the composition of our data in terms of private versus public sector. While the proportion of public sector plans is quite low compared to the total number of plans, about 70 percent of all pensioners and survivors come from public sector plans, given that public sector plans tend to be very large.

Figure 3 shows the total exposures and deaths by gender, pensioner type (pensioners versus survivors), and year of exposure. The total exposures and deaths over 2012 to 2014 are about 1.4 million and 38 thousand, respectively. We can see that the number of exposures and observed deaths for male survivors is limited.

Figure 3

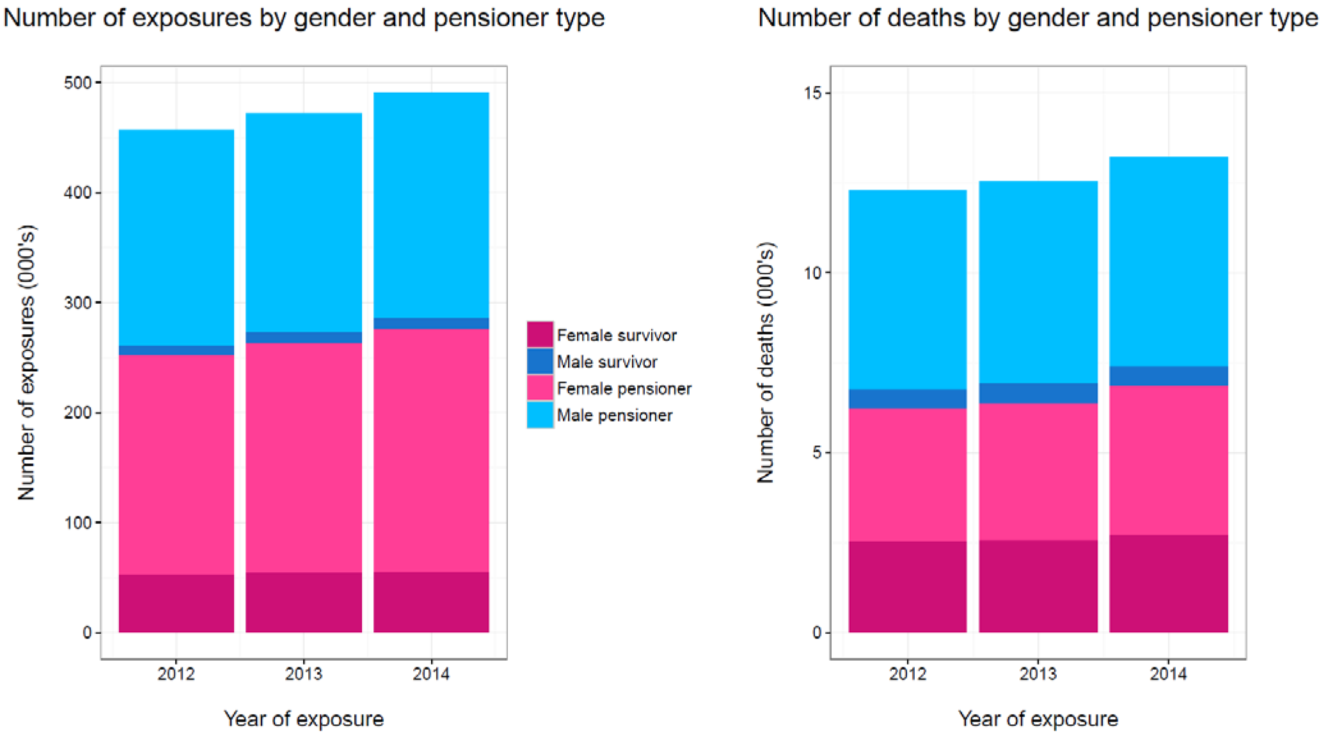
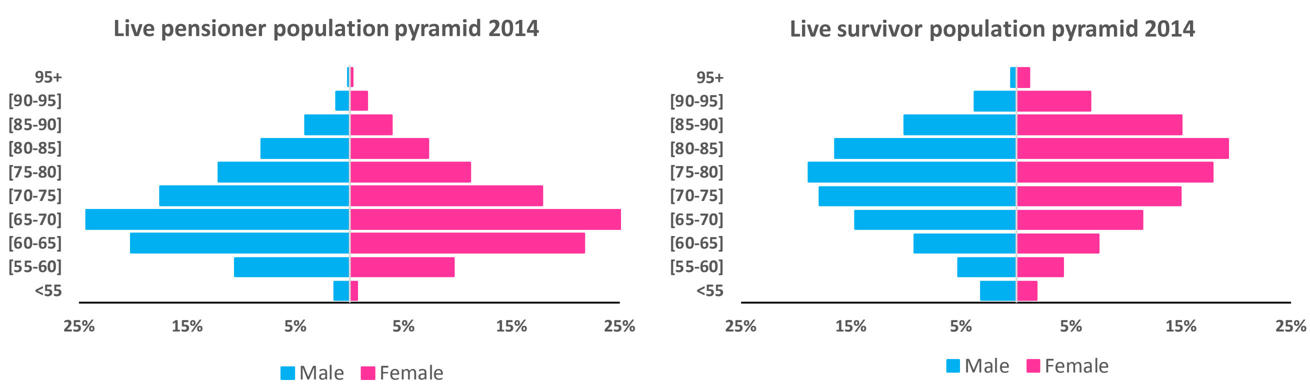


Figure 4 shows the age pyramid for live pensioners and survivors by gender in 2014. It is clear that the pensioner populations are concentrated at younger retirement ages, while the survivor population is skewed toward older ages. This is expected, since survivors begin receiving benefits, and hence become exposed to the risk of death, only following the death of the original member pensioner.

Figure 4



2.6 Data coverage

Ideally, good quality data would be available for all covariates, for all plans; however, the reality is that some plans simply do not have data available for some covariates. This is particularly true for occupation and retirement health type, as plans tend to either have this information for all members or not have it at all. Salary prior to retirement is similar, with some plans having very robust data and others having no data, but some plans also have partial data (e.g., only for recent retirements) which results in the data having to be excluded. We aim to maximize the insight of the data available while minimizing the risk of introducing statistical biases. As we will show later, different combinations of covariates are modelled under the same statistical framework and models are discarded if they do not pass our modelling criteria.

The following points describe the availability of good quality data for the mortality rating factors investigated:

- Age, gender, and pensioner type were available for virtually all data (i.e., above 99.9 percent). Records with missing dates of birth, gender, or pensioner type were omitted from all analysis.
- Public versus private sector was available for all data.
- Industry was assigned for all plans based on the characteristics of the plan sponsor.
- Postal code was available for over 99 percent of the data set and included representation of over 99 percent of all forward sortation areas (i.e., the first three characters of a postal code) in Canada.
- Pension amount was available for almost 91 percent of the data set. While some plans had very small levels of missing pension amounts, the data where pension was unavailable were largely due to mortality data being sourced from non-pension post-retirement benefits being provided to defined benefit pensioners.
- Salary prior to retirement was available for 43 percent of male pensioners and 44 percent of female pensioners.
- Occupation was available for 39 percent of male pensioners but only 14 percent of female pensioners. Occupation information was particularly absent for public sector plans and data was limited for blue collar female pensioners.
- Retirement health data was available for 55 percent of male pensioners and 49 percent of female pensioners. Of those with good quality retirement health data, only 2.6 percent retired with a disability pension.

3. Model specification

To determine the covariates that best explain differences in baseline pensioner mortality and to develop a set of statistical models as a function of age and covariates, we follow Madrigal, et al., (2011) and use a GLM framework. This section gives a brief overview of GLMs.

Let $\vec{x} = (x_1, x_2, \dots, x_p)$ be a p dimensional vector of covariates (e.g., age, affluence group, etc.). Each of these $x_t, t = 1,2, \dots, p$ represents one factor that may have a significant effect on baseline mortality. Given a set of covariates (i.e., $\vec{x}$) for a pension plan member, let $Y_i|\vec{x}\ , i = 1, 2, \dots, n$ be a Bernoulli random variable that takes only two possible values in any period of time as follows:

$$Y_i|\vec{x} = \begin{cases} 0, if \text{ member } is\ alive \\ 1, if\ member\ is \text{ deceased.} \end{cases}$$

In the GLM, we assume that expected value of $Y_i|\vec{x}$ (i.e., the probability of death) can be written as

$$E(Y_i|\vec{x}) = q_x = \frac{e^{\alpha+\vec{\beta}\vec{x}}}{1+e^{\alpha+\vec{\beta}\vec{x}}} \;, \tag{1}$$

where α and β are the unknown parameters to be estimated. Here α can be thought of as a general feature that is common among all covariates and $\vec{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ is a p-dimensional vector that represents the individual effect of each covariate. From equation (1), we can derive the following which represents the form of our GLM:

$$\log(\text{Odds}) = \log\left(\frac{q_{\vec{x}}}{1-q_{\vec{x}}}\right) = logit(\,q_{\vec{x}}) \;=\; \alpha + \vec{\beta}\vec{x}. \tag{2}$$

In other words, $\log(\text{Odds})$ or $logit(q_{\vec{x}})$ is a linear combination of the covariates. In that sense, equation (2) can be regarded as a generalization of the Perk's law of mortality, as described in Perks, (1932). The above GLM is referred to as a logistic GLM and offers many advantages, including the following:

- It can easily incorporate a multivariate framework. This is an appealing feature that enables us to capture the interaction effects of each covariate with age.
- Each term can be statistically tested to see if its inclusion is significant or not in explaining variations in mortality.
- The key metric of interest (i.e., rates of mortality) is modelled directly.
- Equation (1) always preserves the range of probability to be within the interval [0,1].
- It allows non-linear effects with age to be captured.
- There is no need for any additional distributional assumptions (e.g., normality assumption on the dependent variable or homogeneity of variance) as can be seen in many linear models.
- Probabilities extremely close to 0 or 1 can be estimated.

As with any mortality modelling approach, the GLM has its disadvantages. In particular, a large amount of data is required to reliably estimate $q_{\vec{x}}$. Therefore, it's generally not feasible to use a GLM approach for any one individual pension plan. Also, although most Canadian actuaries are likely familiar with linear regression, the logit transformation and application for mortality may be new to many. Developing a GLM model involves complex procedures that often require extensive statistical expertise. While objective statistical criteria are defined to guide the model selection process, judgment may be needed when selecting the preferred model where there is no clearly superior one.

Traditionally in Canada, baseline mortality assumptions have been developed for both pension and insurance applications using graduation techniques. For example, the CPM study graduated mortality rates using a Whittaker-Henderson method as explained in Lowrie, (1982). In this non-parametric approach, crude rates are smoothed by considering both lack of fit and lack of smoothness. An expert opinion is needed to apply the Whittaker-Henderson graduation method to find appropriate smoothing parameters. Similar to GLM, judgment is required; however, here it relates to the parameters.

In addition to the logistic GLM, there are other alternative statistical modelling frameworks commonly used for mortality. For example, rather than working with individual-level data, one may group data and use a binomial model or even a Poisson regression model for the grouped data. Complex survival models are another alternate approach. We refer interested readers to Richards, (2008) and Cox, (1972) for more information on applying survival models to pensioner data.

To prepare the data for analysis, we determine the number of deaths and exposures for the three-year period as follows:

A_{xi} : A 0/1 indicator that represents if the member i at age x died during the year of exposure;

ETR_{xi} : The proportion of a year that individual i, age x is exposed to the risk (ETR) of death;

$\hat{q}_{xi}$: Estimated mortality rate at age x for individual i obtained from fitting a GLM;

$A_x = \sum_i A_{xi}$: Total number of deaths at age x;

$ETR_x = \sum_i ETR_{xi}$: Total number of exposures at age x; and

$E_x = \sum_i \hat{q}_{xi} \times ETR_{xi}$: Total number of expected deaths at age x.

4. Model development

In this section, we provide details on how we developed our models. This includes the analysis of crude rates, determining the functional form of age (i.e., an age-only model), and fitting and selecting preferred models. Performance of the models are investigated using different statistical and actuarial tests.

4.1 Crude rate analysis

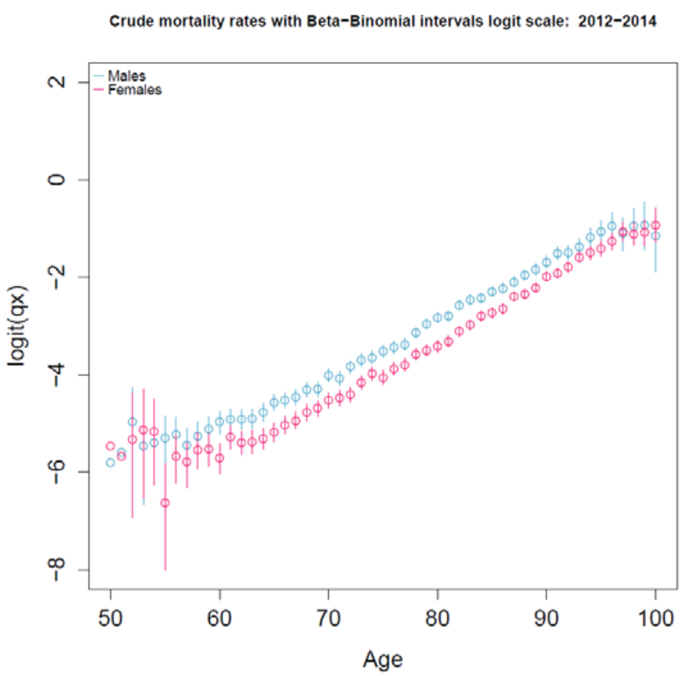
Once exposures are calculated, it is possible to determine the crude mortality rates (i.e., total number of observed deaths within a period divided by the total number of person-years exposed in that period). These unadjusted mortality rates are extremely important and can provide valuable information that can be used for model development.

In Figure 5, crude mortality rates are shown on the logit scale for males versus females, for pensioners and survivors combined. Crude rates at each age are indicated by the empty circles. To assess the credibility of each estimated mortality rate, we performed a Monte Carlo simulation (as explained in theAppendix) and applied a Bayesian approach to find the 95 percent beta-binomial confidence intervals as shown by the vertical lines. This means that the uncertainty around the q_x estimates and observed number of deaths are captured by assuming a beta distribution (as a prior information) and the binomial distribution, respectively. Graphically, longer vertical lines mean less certainty around the q_x estimates, and the lower the empty circle, the lower the mortality at that age.

As expected, we observe that women have generally experienced lower mortality compared to men. And within the 60 to 95 age range, we see that on the logit scale the crude rates resemble a straight line by age—supporting the use of the logistic GLM. We see more uncertainty around the q_x estimates below age 60 and above age 95. In addition, the q_x estimates themselves outside the 60 to 95 age range exhibit suspicious characteristics. For instance, the crude rates for females counter-intuitively exceed those for males below age 55, and both male and female crude rates appear to level off above age 95.

Furthermore, one can see that the gap between male and female mortality rates is larger at younger ages than older ages. This convergence feature is called the "compensation law of mortality", and refers to when the significance of a mortality factor (in this case gender) diminishes as age increases.

Figure 5



Based on the observations noted above, we have selected a fitting age range from 60 to 95 for our analysis. We reassessed this age range for each different stratification of males and females—namely pensioners and survivors, and ill-health pensioners and non-ill-health pensioners. Based on this analysis, we have identified the following fitting age ranges for different strata:

Table 4

Fitting age range by stratum		
Stratum code	**Stratum description**	**Fitting age range**
FPA	**F**emale **P**ensioners, **A**ll retirement health	60–95
MPA	**M**ale **P**ensioners, **A**ll retirement health	60–95
FWA	**F**emale survivors (**W**idows), **A**ll retirement health	60–95
MWA	**M**ale survivors (**W**idowers), **A**ll retirement health	65–90
FPI	**F**emale **P**ensioners**, I**ll-health retirement	60–90
MPI	**M**ale **P**ensioners**, I**ll-health retirement	60–90
FPN	**F**emale **P**ensioners**, N**on-ill-health retirement	60–95
MPN	**M**ale **P**ensioners**, N**on-ill-health retirement	60–95

Table 4 highlights that we've reduced our fitting age range for male survivors (i.e., widower), and male and female ill-health pensioners. This is due to the small sample sizes of these strata. In fact, due to the limitations of the ill-health data, we have not directly fitted models for ill-health and non-ill-health pensioners to develop baseline mortality rates, but instead adjusted the all retirement health models for ill-health and non-ill-health relative mortality. We address how we incorporate the effect of retirement health in section 6.

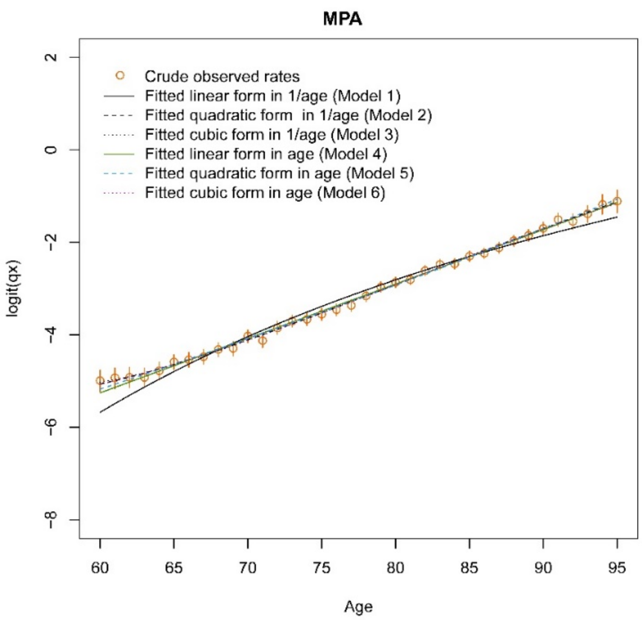
4.2 Age-only model

After stratifying the data by gender and pensioner type, we investigated the appropriate age functional form of the model.

4.2.1 Fitted values

For illustration, we consider the male pensioner data set with all retirement health types (i.e., MPA). To develop our age-only model, we fitted six different logistic regression models including the following: linear, quadratic (second-degree), and cubic (third-degree) models, both directly with age and the reciprocal of age (i.e., 1/age). While we do not dismiss other age-functional forms, our rationale to consider linear/reciprocal forms in age is mainly due to the linear pattern that was exhibited when inspecting the crude mortality rates as shown in Figure 5. As mentioned in Madrigal, et al., (2011), by considering the reciprocal of age as a possible functional form, we allow our models to reflect the compensation law of mortality when incorporating additional covariates. Figure 6 shows estimated crude rates (i.e., empty circles) with 95 percent beta-binomial confidence intervals over the fitting age range of 60–95, and the fitted curves.

Figure 6



Visual inspection reveals that model 1 (the linear model in the reciprocal of age) provides a poor fit to the data by underestimating the observed crude rates at younger and older ages. Models 4 and 5 (linear and quadratic forms directly in age) also slightly underestimate crude rates over the age of 60–63. Models 2, 3, and 6 seem to provide a good fit to the data.

4.2.2 Goodness-of-fit tests

To assess the goodness of fit of alternative models, we have investigated different statistical criteria to measure quality of the fit including the following:

- **Akaike information criterion (AIC):**

  The AIC can be used to compare goodness of fit for different models when fitted to the same data set. It is defined as

$$AIC = 2k - 2\log(\hat{L}),$$

where k is the estimated number of parameters and $\hat{L}$ is the maximum value of the likelihood function evaluated at estimated parameters, as defined in the Appendix. The AIC therefore evaluates the complexity of the model (by taking into account the number of estimated parameters), while also assessing goodness of fit (by considering the likelihood function). Thus, more complex models are penalized by a constant rate.

- **Bayesian information criterion (BIC):**

  The BIC is similar in concept to the AIC and is defined as

$$BIC = \log(n)k - 2\log(\hat{L}),$$

  where n is the number of observations. The BIC penalizes more complex models to a greater degree than the AIC. The preferred models are those that provide low AIC or BIC when comparing different models.

- **Hosmer-Lemeshow test statistic (HL):**

  The HL test statistic was proposed by Hosmer & Lemeshow, (1980) and is a modified version of the Pearson goodness of fit statistic. In the logistic regression modelling framework, the test is based on first sorting the observations in ascending order according to their predicted probabilities. Then data is divided into g groups (e.g., g=10) such that the first group includes observations with the lowest predicted probabilities and the last group contains observations with the highest predicted probabilities. The test statistic is then defined as

$$HL = \sum_{k=1}^{g} \frac{(o_k - n_k\overline{o_k})^2}{n_k\overline{o_k}(1-\overline{o_k})},$$

  where n_k is the total number of the subjects in group k, o_k is observed number of deaths in group k, and $\overline{o_k}$ is the mean estimated probability of death in group k. A smaller HL statistic is preferred. In this paper, the AIC and BIC values have been our primary criteria to test over parameterization, while the HL test was a secondary criterion.

Table 5 provides the functional forms for the six fitted models, along with the AIC, BIC, and HL test statistics. Parameters of each functional form are denoted by a, b, c, or d and will be estimated throughout the fitting process. The table has been colour-coded where red denotes the poorest result and the darkest green the best result.

We can see that model 2 (i.e., the quadratic model in the reciprocal of age) outperforms other models based on the AIC and BIC. It also gives a relatively small HL test statistic. Generally, the model(s) with the smallest AIC or BIC is the preferred model(s). Using the definition of these criteria, this means that likelihood $\hat{L}$ should be large and the number of estimated parameters k should be small. When comparing a set of models using AIC or BIC, we can first find $\Delta_j = AIC_j - AIC_{min}$, where AIC_j is the AIC of the j^{th} model and AIC_{min} is the minimum AIC values among all models. As mentioned in Burnham & Anderson, (2002), models with $\Delta_j > 10$ have less support for further consideration. Using AIC values given in **Error! Reference source not found.**, we have $AIC_{min} = 129097$, so $\Delta_j = 309, 0, 1, 18, 11$ *and* 3 for models $j = 1, 2, \ldots, 6$, respectively. This means that models 1, 4, and 5 have less support for further consideration. Using the AIC, BIC, and HL statistical tests, models 2, 3, and 6 perform best.

Table 5

Functional forms for the six fitted models and their goodness-of-fit test results

Model	Functional Form	Model Name	AIC	BIC	HL
1	$logit(q_x) = a + bx^{-1}$	Linear form in 1/age	129406	129429	233.88
2	$logit(q_x) = a + bx^{-1} + cx^{-2}$	Quadratic form in 1/age	129097	129131	16.31
3	$logit(q_x) = a + bx^{-1} + cx^{-2} + \mathrm{d}x^{-3}$	Cubic form in 1/age	129098	129143	14.84
4	$logit(q_x) = a + bx$	Linear form in age	129115	129137	24.86
5	$logit(q_x) = a + bx + cx^2$	Quadratic form in age	129108	129142	23.60
6	$logit(q_x) = a + bx + cx^2 + dx^3$	Cubic form in age	129100	129145	15.72

4.2.3 Actuarial tests

We compared the six models using formal actuarial tests as described briefly in the Appendix. Overall, 12 different actuarial tests were carried out including the following: sign test, Kolmogorov-Smirnov (KS) test, run test, chi-squared test, standardized deviations (SD) test, cumulative deviations (CD) test, serial correlations (SC) test, actual over expected (AoE) test, monotonic test, and life expectancy (LEX) comparison test. We refer the interested reader to Forfar, et al., (1988) for details.

Table 6 shows the results of the actuarial tests for the six fitted age-only models. We can see that model 1 fails seven tests, while models 4 and 5 both fail the chi-squared test. Models 2, 3, and 6 perform equally well, passing all 12 of the considered actuarial tests.

Table 6

Results of actuarial tests for the six fitted models

Model	Sign	KS	Run	Chi-squared	SD	CD	SC	AoE	Monotonic	LEX(65)	LEX(75)	LEX(85)
1	PASS	PASS	FAIL	FAIL	FAIL	PASS	FAIL	PASS	PASS	FAIL	FAIL	FAIL
2	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
3	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
4	PASS	PASS	PASS	FAIL	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
5	PASS	PASS	PASS	FAIL	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
6	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

After going through the results of the statistical and actuarial tests, we exclude models 1, 4, and 5 from further consideration and focus only on models 2, 3, and 6 for further testing.

4.2.4 Coefficients of candidate models

Table 7 shows the estimated parameters for models 2, 3, and 6 with their corresponding lower and upper 95 percent confidence intervals, as well as the t-values and probability values (p-values) for each estimated coefficient. The estimation of the parameters is carried out by maximizing the log likelihood function (defined in the Appendix) using the iteratively reweighted least squares method as explained in Fox, (2010). Each observation is weighted according to its corresponding exposure. Because the p-values are greater than 0.05, the coefficients of the linear, quadratic, and cubic terms in model 3 are not statistically significant at a 95 percent confidence level. In contrast, all the terms appearing in models 2 and 6 are statistically significant, other than the intercept term in model 6. We therefore drop out model 3 and only further consider models 2 and 6. Alternatively, instead of examining the p-values given in Table 7, one can apply 95 percent confidence intervals to test the contribution of each parameter. If a 95 percent confidence interval does not include zero, then we have enough evidence to support statistical significance at a 95 percent confidence level. For example, none of the provided confidence intervals for model 2 contain zero; therefore, all the terms are statistically significant.

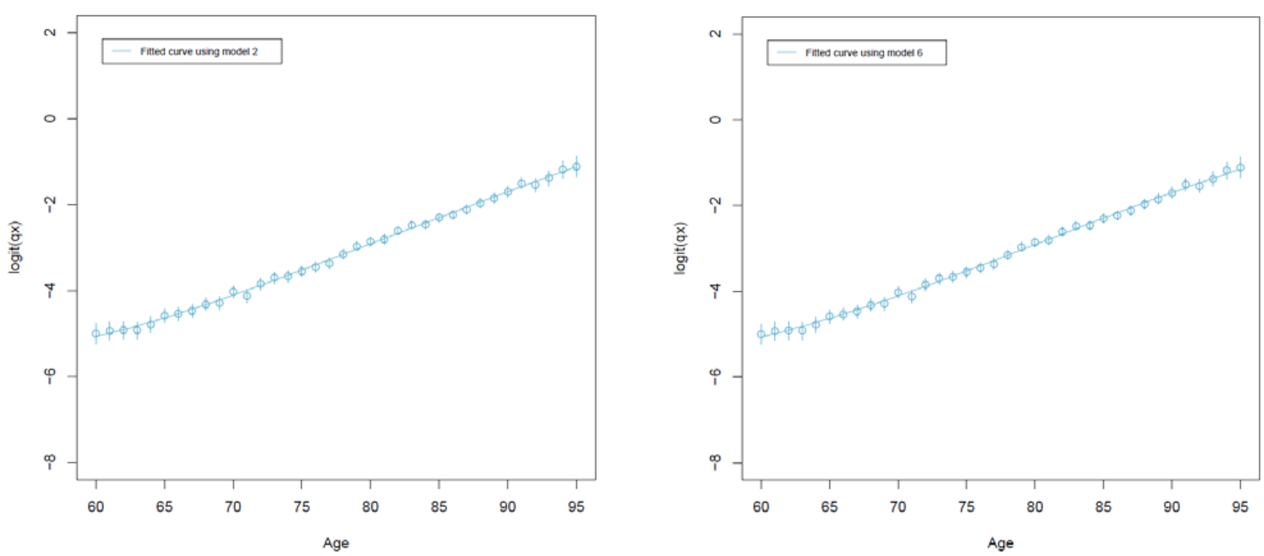
Table 7

Estimated parameters with 95% confidence intervals						
Model: functional form	**Parameter**	**Estimate**	**Lower 95% confidence interval**	**Upper 95% confidence interval**	**t value**	**p-value**
$2: logit(q_x) = a + bx^{-1} + cx^{-2}$	a	1.69×10^{1}	15.715	18.137	27.388	0
	b	-2.39×10^{3}	-2575.113	-2207.478	-25.497	0
	c	6.43×10^{4}	57420.102	71227.011	18.262	0
$3: logit(q_x) = a + bx^{-1} + cx^{-2} + \mathrm{d}x^{-3}$	a	1.14×10^{1}	1.355	21.401	2.225	0.026
	b	-1.13×10^{3}	-3400.702	1145.105	-0.973	0.331
	c	-3.07×10^{4}	-201261.49	139809.318	-0.353	0.724
	d	2.36×10^{6}	-1871203.419	6594352.48	1.094	0.274
$6: logit(q_x) = a + bx + cx^2 + dx^3$	a	6.66×10^{0}	-3.81	17.121	1.246	0.213
	b	-6.01×10^{-1}	-1.01	-0.191	-2.876	0.004
	c	8.98×10^{-3}	0.004	0.014	3.323	0.001
	d	-3.71×10^{-5}	-0.00006	-0.00001	-3.203	0.001

4.2.5 Fitted versus crude rates

Figure 7 shows the fitted curves for models 2 and 6 compared against the estimated crude rates over the fitting age range of 60 to 95. Both models fit the data very well.

Figure 7



Model 2: $logit(q_x) = a + bx^{-1} + cx^{-2}$

Model 6: $logit(q_x) = a + bx + cx^2 + \mathrm{d}x^3$

4.2.6 Fitted age-only models

After considering all models and the test results described above, we choose model 2 to represent our age-only model for male pensioners since it is a more parsimonious model, with better goodness-of-fit test results. In addition, by including terms using the reciprocal of age, model 2 offers the favourable attribute of allowing for the compensation law of mortality.

The model selection process outlined in 0 to 0 was also performed for male survivors, female pensioners, and female survivors. Table 8 shows our final age-only models for each stratum. Using these fitted models, the curtailed period life expectancy (i.e., curtailed at age 95 since this is the end of our fitting age range) at age 65 of male pensioners exceeds that of male survivors by 1.9 years, with a differential of 1.3 years for female pensioners over female survivors.

Table 8

Final age-only models for each stratum		
Stratum code	**Model name**	**Age functional form**
MPA	Quadratic form in 1/age	$logit(q_x) = a + bx^{-1} + cx^{-2}$
MWA	Quadratic form in 1/age	$logit(q_x) = a + bx^{-1} + cx^{-2}$
FPA	Cubic form in 1/age	$logit(q_x) = a + bx^{-1} + cx^{-2} + \mathrm{d}x^{-3}$
FWA	Cubic form in 1/age	$logit(q_x) = a + bx^{-1} + cx^{-2} + \mathrm{d}x^{-3}$

4.3 Univariate analysis

After completing the fitting of the age-only models, we performed a univariate analysis of different rating factors to begin the process of determining which covariates best explain differences in baseline mortality.

We first visually inspected the crude rates for each covariate outlined in subsection 2.1, and reviewed the corresponding crude life expectancies to assess the degree of differentiation between covariate categories. We found that the largest degree of differentiation occurred for retirement health type, longevity group, and salary band.

We then added each covariate to the age-only model for both male and female pensioners of all retirement health types, and applied a chi-squared test to check if the reduction in the residual deviance was statistically significant or not. Overall, we found that for both male and female pensioners, adding each covariate improved the age-only model at a 95 percent confidence level. The only covariate that was not significant over three years of mortality experience (i.e., 2012 to 2014) was the year of exposure. However, year of exposure is particularly important when developing mortality improvements. In this case, a longer history of data and a more complicated functional form of year of exposure—to allow passage of time—is needed.

While the results of our univariate analysis were informative, we next tested the importance of each factor in the presence of other factors using multivariate data analysis.

4.4 Multivariate analysis

Typically, pension plan administration systems have multiple mortality rating factors stored on their plan members. Performing a multivariate analysis allows the determination of whether using rating factors available on pension plan administration systems, other than age, enhances the mortality assumptions that can be created for plans to use and improves each plan's ability to assess its longevity risk.

Our multivariate analysis included all the covariates outlined in subsection 2.1. We applied a stepwise regression, using the BIC to select the best possible model for both male and female pensioners of all retirement health types. We refer interested readers to Venables & Ripley, (2002) for details of the stepwise regression method.

Among all the considered mortality rating factors listed in subsection 2.1, the stepwise regression found the following to be the most informative covariates:

- Age;
- Postal code, which is translated into one of five longevity groups;
- Pension amount;
- Salary at retirement (or earlier exit); and
- Occupation.

Based on the results of the stepwise regression, other rating factors we tested, including public sector versus private sector employment, were dropped from the final selected models.

To illustrate the stepwise regression procedure, we considered female pensioners and tested the inclusion of the most informative covariates (as outlined above), together with sector, postal district, urban versus rural, year of exposure, and season of birth. As indicated in the first column of Table 9, the final model includes only age, longevity groups, and salary. Other considered covariates have ultimately been dropped through the model selection procedure. The table presents chi-squared test results for the final model selected and the p-values show that the inclusion of these covariates is statistically significant.

Table 9

Stepwise regression model for female pensioners using salary and chi-squared test results					
Covariate	Degree of freedom	Deviance	Residuals	Residual deviance	p-value
x^{-1}	1	4440	247463	35948	< $2x10^{-16}$
x^{-2}	1	123	247462	35825	< 2 $x10^{-16}$
x^{-3}	1	7	247461	35818	$6.30x10^{-3}$
Longevity group	3	87	247458	35730	< 2 $x10^{-16}$
Salary	2	46	247456	35684	1.10 $x10^{-11}$

In addition, we simultaneously tested pension versus salary at retirement (or earlier exit) to find out which was the most significant affluence-based covariate. Our stepwise regression analysis shows that salary amount is a better longevity predictor compared to pension amount for both male and female pensioners. This can be explained by the fact that pension amounts do not depend solely on earnings level (e.g., they are heavily influenced by years of service). We also tested occupation simultaneously with longevity grouping and pension amount for male pensioners. Occupation was found to be an important mortality rating factor, particularly for male pensioners.

Table 10 provides ranges of crude life expectancy at age 65 based on method (II) in Chiang, (1984) for the four strata by each of longevity groupings, pension bands, salary bands, and occupation (i.e., only considering one covariate at a time). The result is that there is a 3.7-year range in male pensioner crude life expectancies when considering different possible factors individually (3.4 years for female pensioners), with the ranges for pension bands, salary bands, and occupation falling within that for longevity groupings.

Table 10

Range of crude life expectancies at age 65 over bands/groupings				
Stratum code	Pension bands only	Salary bands only	Longevity groupings only	Occupation groupings only
MPA	18.9–20.5	17.9–21.3	17.9–21.6	18.9–20.7
FPA	22.1–24.1	22.1–23.9	21.0–24.4	21.9–22.4
MWA	16.8–18.1	n/a	n/a	n/a
FWA	20.6–21.5	n/a	n/a	n/a

We are interested in how the covariates explain variations in mortality in a multivariate context. For example, Table 11 shows crude life expectancy at age 65, again based on method (II) in Chiang, (1984), for female pensioners by different combinations of salary bands and longevity groups. There is a 4.5-year difference in crude life expectancy at age 65 between female pensioners in the lowest salary band and lowest longevity group to those in the highest salary band and longevity group. For male pensioners, the corresponding range is 6.2 years. The multivariate results therefore expand the ranges presented in Table 10 that looked at individual covariates.

Table 11

FPA crude life expectancies at age 65 by salary band and longevity group					
Salary band	A	B	C	D	E
<$39,900	20.96	21.63	22.52	23.69	23.30
$39,900–$58,000	21.61	22.29	22.84	23.44	23.92
>$58,000	22.87	23.25	24.11	24.63	25.44

It is worth noting that while the majority of crude life expectancies in Table 11 increase by successive salary bands and longevity groups, the first salary band has a higher crude life expectancy for longevity group D than E, and longevity group D has a higher crude life expectancy for the second salary band than the first salary band. During the model calibration process, as outlined in section 5, we eventually determined that longevity groups D and E should be combined for this model.

5 Model calibration

Not all pension plans have good quality data for all the covariates that were identified by our multivariate analysis. Therefore, additional calibrations of the model (i.e., complete GLM models) are created to provide mortality assumptions when a plan does not have all covariates available.

To illustrate our model calibration process, we focus on the calibration of our model for female pensioners with good quality salary data. The same methodology has been applied for all other calibrations (i.e., different plausible combinations of strata and covariates). Table **12** shows all the calibrations where we have fitted a model. The green check marks indicate the inclusion of the rating factor, while red cross marks indicate that we have not considered that rating factor when fitting the applicable model. The total number of sets of mortality rates (i.e., survival curves) for each stratum is also given in the last column of Table 12. In developing these calibrations, we considered data availability and reliability as discussed earlier in section 2. Not all possible combinations of strata and covariates have been considered due to lack of data availability and/or exclusion during the model fitting process.

Table 12

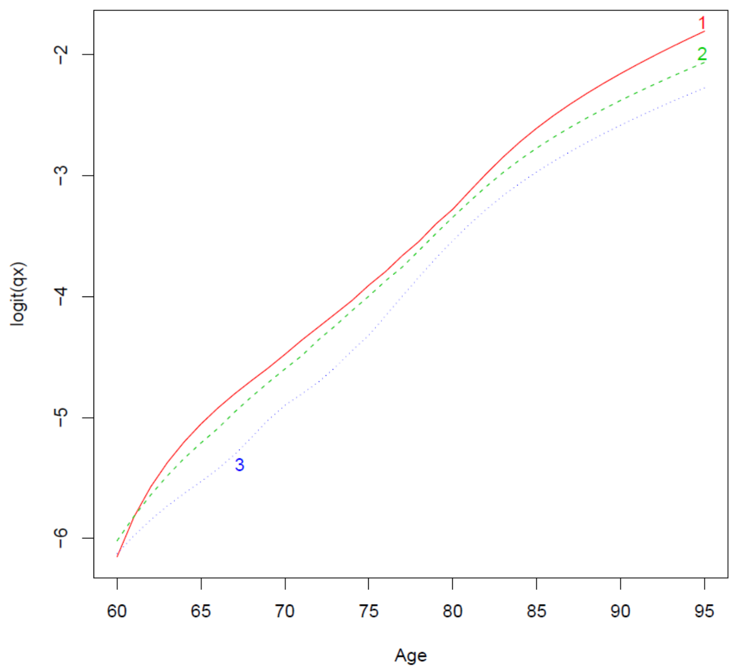
Number of fitted models by calibration and included covariates						
Stratum	**Age**	**Longevity grouping**	**Salary amount**	**Pension amount**	**Occupation**	**# of curves**
MPA/MPN	✓	✗	✗	✗	✗	1
	✓	✓	✗	✗	✗	5
	✓	✗	✗	✗	✓	2
	✓	✗	✗	✓	✗	3
	✓	✗	✓	✗	✗	4
	✓	✓	✓	✗	✗	20
	✓	✓	✗	✓	✗	15
	✓	✓	✗	✗	✓	10
	✓	✗	✗	✓	✓	6
	✓	✓	✗	✓	✓	30
FPA/FPN	✓	✗	✗	✗	✗	1
	✓	✓	✗	✗	✗	5
	✓	✗	✗	✗	✓	2
	✓	✗	✗	✓	✗	3
	✓	✗	✓	✗	✗	3
	✓	✓	✓	✗	✗	15
	✓	✓	✗	✓	✗	15
FWA	✓	✗	✗	✗	✗	1
	✓	✓	✗	✗	✗	5
	✓	✗	✗	✓	✗	2
	✓	✓	✗	✓	✗	10
MWA	✓	✗	✗	✗	✗	1
MPI	✓	✗	✗	✗	✗	1
FPI	✓	✗	✗	✗	✗	1

5.1 LOWESS plot

To determine if the number of discrete bands for each covariate continues to show a difference in the rates of mortality for a particular calibration, we apply the locally weighted scatterplot smoothing (or LOWESS) non-parametric regression method. In this technique, the data is first divided into smaller groups using a nearest neighbour's algorithm. Next, a polynomial regression in age is fitted to each group. This gives a flexible and convenient way to visualize the fitted curve when considering age as the only rating factor. We encourage the interested reader to consult Cleveland, (1981) and Cleveland, (1979) for more details on the LOWESS plot.

As an example, Figure 8 shows the LOWESS plot of the fitted mortality rates on a logit scale by age and the three salary bands. In general, salary band 1 provides higher mortality than salary bands 2 and 3 over almost the entire fitting age range. Salary band 3, on the other hand, gives the lowest mortality rate as expected. Although the salary bands become closer to each other around age 80, it is clear that there is a good distinction in mortality rates over the fitting-age range. In addition, we can observe that salary bands do not cross, other than before age 63. This suggests only a very modest interaction effect between salary band and age.

Figure 8



5.2 Contrast test

A second check applies a general contrast test of each adjacent salary band coefficient using a t-test as explained in Harrell, (2001). We performed the following three different tests:

- **Main effect model**

  We fit a logistic regression by considering only the main effects of age and salary bands and tested whether salary band 1 versus band 2, and band 2 versus band 3, are statistically different. Estimated contrasts and the p-values of the t-test are shown in columns 2 and 3 of Table 13.

- **Interaction effect model**

  We fit a logistic regression by not only considering the main effects of age and salary bands, but also included the interaction effects of age and salary bands. Column 4 shows the p-values of the t-test using the interaction effect model.

- **Individual age**

  Using the interaction effect model explained above, we also check if the adjacent salary bands are statistically significant at specific ages: 62, 72, 82, and 92. The p-values of the t-tests are shown in columns 5 to 8.

Table 13

Contrast test results for comparing adjacent salary bands (SB)							
Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7	Col. 8
Contrast	Estimated contrast main effect model	P-value main effect model	P-value interaction effect model	P-value age 62	P-value age 72	P-value age 82	P-value age 92
SB 1 vs. SB 2	0.074	0.041	0.058	0.369	0.412	0.000	0.000
SB 2 vs. SB 3	0.256	0.000	0.000	0.005	0.000	0.000	0.027

Since the estimated contrasts given in Table 13 are both positive, one can conclude that the salary bands are in the right order (i.e., on average, salary band 1 has higher mortality than salary band 2, etc.). In addition, the p-values are highly supportive of the fact that salary bands can provide a good distinction across the fitting age range for both the main and interaction effects.

5.3 Test for interaction with age

In this section, we tested the interaction effect of salary bands with age. To do so, we fitted four logistic regression models as follows:

Model 1: $logit(q_x) = a_1 + b_1 x^{-1} + c_1 x^{-2} + d_1 x^{-3} + e_{1k}\ (Salary_k),$

Model j: $logit(q_x) = a_j + b_j x^{-1} + c_j x^{-2} + d_j x^{-3} + e_{jk}\ (Salary_k) + f_{jk}(x^{-(j-1)}{:}\,Salary_k),$

$j = 2, 3, 4$ and $k = 2, 3,$

where the $Salary_k$ term stands for the main effect of salary band k, and $x^{-j}{:}\,Salary_k$ represents the interaction effect between salary band k and reciprocal of age with degree j. In the above models, terms $x^{-1},\ x^{-2}, x^{-3}$ examine the effect of reciprocal of age with degree 1, 2, and 3 on mortality while ignoring the effect of salary. Similarly, the term $Salary_k$ tests the effect of salary as a measure of affluence on mortality regardless of age. The effect of age on mortality may not be the same for all salary bands. Therefore, we added an additional term indicated by $x^{-j}{:}\,Salary_k$ to test if the effect of age on mortality depends on different levels of salary. Model 1 does not consider any interaction term, while models 2, 3, and 4 each have one term that accounts for the interaction between salary and age. The term $Salary_k$ can be regarded as an indicator variable (sometimes called a dummy variable) that captures the salary effect for different bands. We define it as follows:

$$Salary_k = \begin{cases} 1, & if\ considering\ salary\ band\ k, k = 2,3 \\ 0, & otherwise. \end{cases}$$

Based on this definition, we can capture the effect of salary band 1 by setting $Salary_2 = Salary_3 = 0$, which removes terms e_{1k} and f_{jk} in the logistic regression models above. The effects corresponding to salary band 1 are then reflected in the intercept term a_1.

Table 14 provides estimated parameters, standard errors, t values, and p-values for these four models. The first salary band is considered the reference level; the tests should be regarded with respect to the first salary band. In model 1, the p-values of the coefficients e_{12} and e_{13} are significant at 95 percent confidence level. This means that coefficients of salary band 2 and 3 are statistically different from the coefficient of salary band 1. The p-values that are highlighted in red for models 2, 3, and 4 are those that are not statistically significant. When we added a linear, quadratic, or cubic interaction term to the model, we saw that coefficients of salary band 2 and 3 are no longer statistically different from the coefficient of salary band 1. Similarly, the interaction terms of salary bands 2 and 3 with age were not found to be significant with respect to salary band 1.

Table 14

T-test results for models 1–4					
Model	**Parameter**	**Estimate**	**Std. error**	**t value**	**p-value**
1	a_1	41.3155	9.638	4.29	0.000018
	b_1	-7893.4563	2199.4383	-3.59	0.00033
	c_1	467441.65	166025.211	2.82	0.00487
	d_1	-9782738.9	4145719.962	-2.36	0.01829
	e_{12}	-0.0739	0.0362	-2.04	0.04101
	e_{13}	-0.3298	0.0466	-7.08	1.50×10^{-12}
2	a_2	42.0487	9.6739	4.35	0.000014
	b_2	-8054.1367	2207.7771	-3.65	0.00026
	c_2	478072.1	166595.5908	2.87	0.00411
	d_2	-9981985.9	4156783.771	-2.4	0.01633
	e_{22}	0.0933	0.3202	0.29	0.77075
	e_{23}	0.3451	0.4055	0.85	0.39476
	f_{22}	-13.5703	24.973	-0.54	0.58685
	f_{23}	-51.7997	31.0152	-1.67	0.09489
3	a_3	41.92852	9.66958	4.34	0.000015
	b_3	-8020.2747	2206.33775	-3.64	0.00028
	c_3	475024.62	166461.0618	2.85	0.00432
	d_3	-9893805.2	4153254.292	-2.38	0.01721
	e_{32}	0.00743	0.15962	0.05	0.96288
	e_{33}	-0.00773	0.20433	-0.04	0.96983
	f_{32}	-527.20538	945.70811	-0.56	0.57721
	f_{33}	-1873.3999	1163.61971	-1.61	0.1074
4	a_4	41.7488	9.6617	4.32	0.000016

	b_4	-7973.0634	2204.2304	-3.62	0.0003
	c_4	470999.58	166295.4731	2.83	0.0046
	d_4	-9782310.1	4149672.128	-2.36	0.0184
	e_{42}	-0.0216	0.1076	-0.2	0.8413
	e_{43}	-0.1261	0.1392	-0.91	0.3648
	f_{42}	-26670.314	47162.7411	-0.57	0.5717
	f_{43}	-88642.165	57490.679	-1.54	0.1231

We further investigate inclusion of the interaction terms in models 2, 3, and 4 by adding each term sequentially and performing a chi-squared test. We next check if the reduction in the deviance residuals is statistically significant or not. The results of the chi-squared test are given in Table 15. From this table, we see that the inclusion of the interaction terms with salary and the linear, quadratic, and cubic age form do not improve model 1. The results based on both the z test and chi-squared test are consistent with our previous visual inspection (given in subsection 5.1) that identified no evidence to support inclusion of the age and salary interaction term (i.e., the effect of salary on mortality does not vary with age).

Table 15

Chi-squared test results for models 2, 3, and 4			
Model	**Terms added sequentially**	**Deviance residuals**	**p-value**
2	$intercept$	42679	NA
	$intercept + x^{-1}$	37831	< 2x10^{-16}
	$intercept + x^{-1} + x^{-2}$	37705	< 2x10^{-16}
	$intercept + x^{-1} + x^{-2} + x^{-3}$	37699	0.014
	$intercept + x^{-1} + x^{-2} + x^{-3} + Salary$	37646	3.10x10^{-12}
	$intercept + x^{-1} + x^{-2} + x^{-3} + Salary + x^{-1}{:}Salary$	37643	0.247
3	$intercept$	42679	
	$intercept + x^{-1}$	37831	< 2x10^{-16}
	$intercept + x^{-1} + x^{-2}$	37705	< 2x10^{-16}
	$intercept + x^{-1} + x^{-2} + x^{-3}$	37699	0.014
	$intercept + x^{-1} + x^{-2} + x^{-3} + Salary$	37646	3.1x10^{-12}
	$intercept + x^{-1} + x^{-2} + x^{-3} + Salary + x^{-2}{:}Salary$	37643	0.272
4	$intercept$	42679	
	$intercept + x^{-1}$	37831	< 2x10^{-16}
	$intercept + x^{-1} + x^{-2}$	37705	< 2x10^{-16}
	$intercept + x^{-1} + x^{-2} + x^{-3}$	37699	0.014
	$intercept + x^{-1} + x^{-2} + x^{-3} + Salary$	37646	3.1x10^{-12}
	$intercept + x^{-1} + x^{-2} + x^{-3} + Salary + x^{-3}{:}Salary$	37643	0.303

5.4 Goodness-of-fit tests

We next compared goodness-of-fit tests using the four models that were considered in subsection 5.3. The AIC, BIC, and HL results are given in Table 16. Models 1 and 2 are the preferred models according to the AIC. Following the 10-unit threshold outlined in Raftery, (1995), models 2, 3, and 4 are over-parametrized based on the BIC compared to model 1. Therefore, we have very strong evidence that model 1 outperforms other models. We exclude models 3 and 4 from further consideration at this stage and focus only on models 1 and 2.

Table 16

Goodness-of-fit test results for models 1 to 4

Model	Functional form	AIC	BIC	HL
1	$logit(q_x) = a_1 + b_1x^{-1} + c_1x^{-2} + d_1x^{-3} + e_{1k}\ (Salary_k)$	37666	37729	12.4
2	$logit(q_x) = a_2 + b_2x^{-1} + c_2x^{-2} + d_2x^{-3} + e_{2k}\ (Salary_k) + f_{2k}(x^{-1}: Salary_k)$	37667	37751	12.44
3	$logit(q_x) = a_3 + b_3x^{-1} + c_3x^{-2} + d_3x^{-3} + e_{3k}\ (Salary_k) + f_{3k}(x^{-2}: Salary_k)$	37668	37751	11.09
4	$logit(q_x) = a_4 + b_4x^{-1} + c_4x^{-2} + d_4x^{-3} + e_{4k}\ (Salary_k) + f_{4k}(x^{-3}: Salary_k)$	37668	37751	11.14

5.5 Actuarial tests

A series of actuarial tests, as explained in subsection 0, have been carried out at each salary band using models 1 and 2. The results of actuarial tests are given in Table 17.

Table 17

Results of actuarial tests for models 1 and 2

Model	Salary band	Sign	KS	Run	Chi-squared	SD	CD	SC	AoE	Monotonic	LEX(65)	LEX(75)	LEX(85)
1	1	PASS	PASS	FAIL	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
1	2	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
1	3	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
2	1	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
2	2	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
2	3	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

Although model 1 fails the run test at salary band 1, both models perform well based on the actuarial tests.

5.6 Coefficient of candidate models

Table 18 provides estimated parameters, lower/upper 95 percent confidence intervals, t values and p-values for models 1 and 2. We can see that in model 2, the main effect of salary bands 2 and 3 are not statistically different from salary band 1. Similarly, interaction effects of salary bands 2 and 3 with age do not appear to be significant with respect to salary band 1. Based on these results, model 1 is the preferred model.

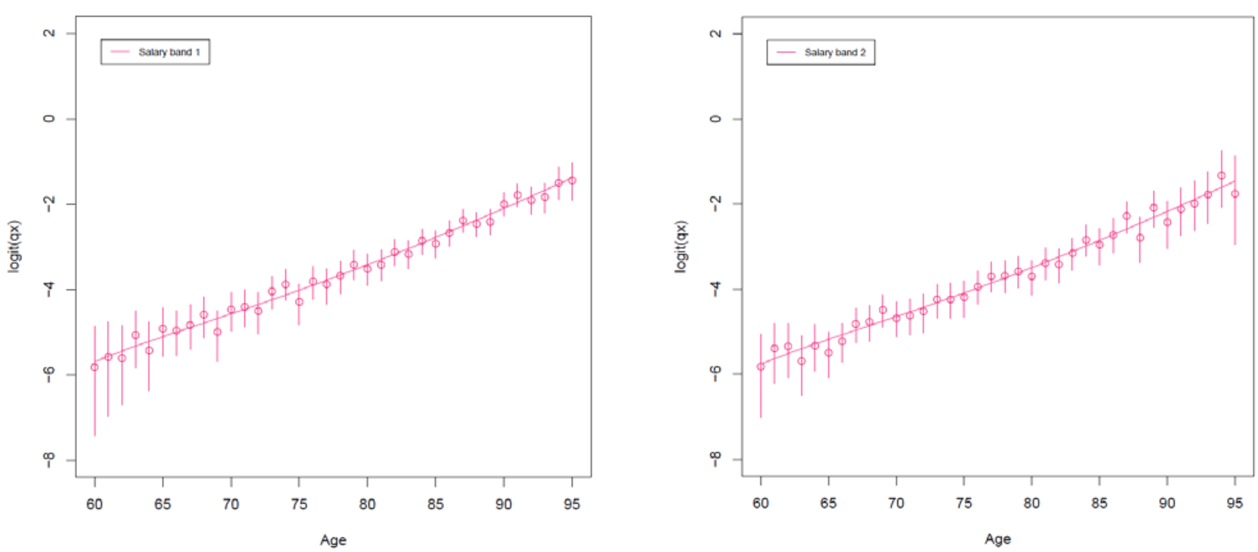
Table 18

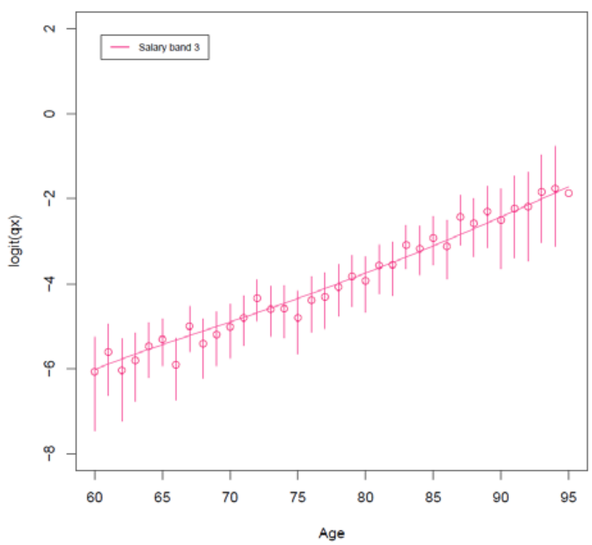
T-test results for models 1 and 2						
Model	**Parameter**	**Estimate**	**LCI**	**UCI**	**t value**	**p-value**
1	a_1	41.315	22.425	60.206	4.287	0
	b_1	-7893.456	-12204.276	-3582.637	-3.589	0
	c_1	467441.652	142038.218	792845.086	2.815	0.005
	d_1	-9782738.921	-17908200.74	-1657277.1	-2.36	0.018
	e_{12}	-0.074	-0.145	-0.003	-2.043	0.041
	e_{13}	-0.33	-0.421	-0.238	-7.079	0
2	a_2	42.049	23.088	61.009	4.347	0
	b_2	-8054.137	-12381.3	-3726.973	-3.648	0
	c_2	478072.097	151550.739	804593.455	2.87	0.004
	d_2	-9981985.931	-18129132.41	-1834839.45	-2.401	0.016
	e_{22}	0.093	-0.534	0.721	0.291	0.771
	e_{23}	0.345	-0.45	1.14	0.851	0.395
	f_{22}	-13.57	-62.516	35.376	-0.543	0.587
	f_{23}	-51.8	-112.588	8.989	-1.67	0.095

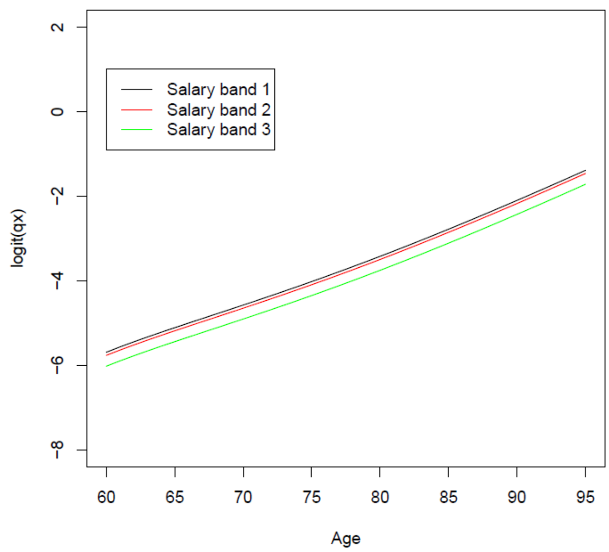
5.7 Fitted versus crude rates plots

Finally, we plotted the fitted curves over the crude rates separately by each salary band, and then the combined fitted curves, as shown in Figure 9. This has been done to make sure that the curves fit the observed data well and are not crossing. The final fitted curves using model 1 seem reasonable.

Figure 9







5.8 Calculating predicted probabilities from a logistic regression model

The main purpose of this subsection is to provide a practical example to show how the predicted probabilities of death can be obtained from the estimated coefficients of model 1 given in Table 18. For the sake of illustration, we determine q_{65} for a female pensioner in salary band 2. Using the functional form of model 1 in subsection 5.3 with $k = 2$, we have

$$
\begin{aligned}
logit(q_{65}) &= a_1 + \frac{b_1}{65} + \frac{c_1}{65^2} + \frac{d_1}{65^3} + e_{12} \\
&= 41.315 - \frac{7893.456}{65} + \frac{467441.652}{65^2} - \frac{9782738.921}{65^3} - 0.074 \\
&= -5.181879.
\end{aligned}
$$

Next, we convert $logit(q_{65})$ to q_{65} using equation (1):

$$q_{65} = \frac{e^{-5.181879}}{1 + e^{-5.181879}} = 0.005586.$$

This means that in 2013, the mid-year of the calibration period, a female pensioner aged 65 with a salary at retirement (or earlier exit) in salary band 2 will die within one year with the probability of 0.005586. The interpretation of estimated coefficients is straightforward particularly on the logit scale. For example, terms e_{12} and e_{13} in Table 18 capture the effects of salary band 2 and 3, respectively. This means that for each female pensioner aged x, the probability of death for those in salary band 2 compared to salary band 3 increases on a logit scale by $-0.074 - (-0.33) = 25.6\%$.

Using the calibration of female pensioners with good quality salary at retirement (or earlier exit) and postal code data, we developed a model equivalent of model 1 with the additional factor of longevity group. Table 19 provides the probability of death at age 65 and the change in the probability of death compared to a pensioner whose revalued salary in December 2015 is $50,000 and lives in a postal code corresponding to lifestyle group 2.

Table 19

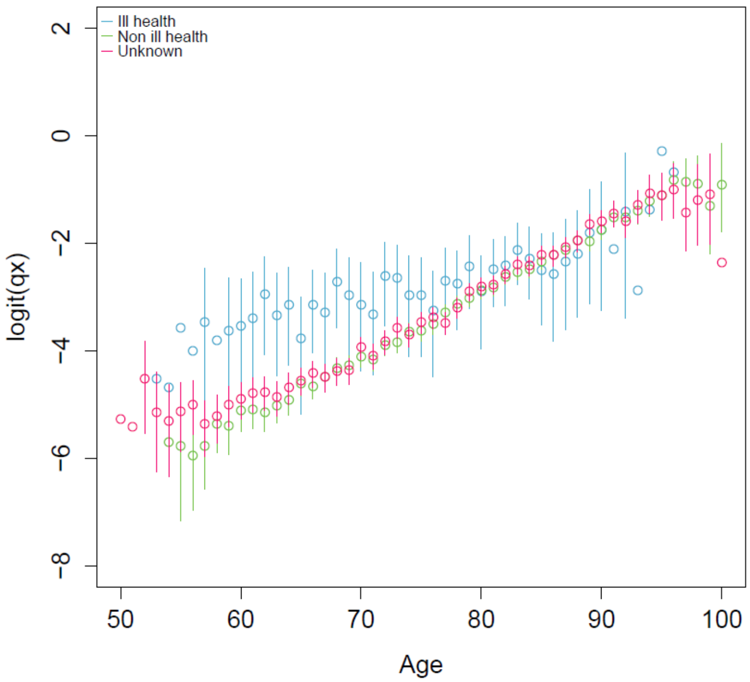
Changes in probability of death at age 65 by longevity profile		
Longevity profile	q_{65}	**% change in probability of death**
Salary band 1 & longevity group B	0.006313	8%
Salary band 2 & longevity group B	0.005827	0%
Salary band 3 & longevity group B	0.004333	-26%
Longevity group A & salary band 2	0.006782	16%
Longevity group B & salary band 2	0.005827	0%
Longevity group C & salary band 2	0.005234	-10%
Longevity group D & salary band 2	0.004402	-24%

We can see that female pensioners who are age 65 in longevity group B earning less than $39,900 (i.e., salary band 1) have an 8 percent higher probability of death than those with salary within $39,900–$58,000 (i.e., salary band 2). On the other hand, more affluent members that fall in salary band 3 have a 26 percent lower probability of death than those with earnings within salary band 2. Similarly, when compared to female pensioners at age 65 with earnings in salary band 2, being in longevity group A increases the probability of death by 16 percent compared to those in longevity group B. Compared to female pensioners in longevity group B, pensioners in longevity groups C and D have 10 percent and 24 percent lower probability of death, respectively.

6 Adjusting retirement health

By applying the same methodology explained in section 5, we calibrated a range of different models encompassing different covariates. In total, we developed 140 different curves for male and female pensioners with all retirement health types combined (i.e., ill health, non-ill health, and unknown health), as indicated in Table 12. However, our univariate analysis of the crude rates showed that health condition at retirement is a significant mortality factor. As an example, Figure 10 shows crude mortality rates for male pensioners by retirement health type. Clearly those who have retired with a disabled pension (i.e., ill health) have higher mortality rates, particularly at younger retirement ages. In this section, we briefly explain how we developed adjustment factors that vary by age, and applied them to our all-health curves to obtain corresponding non-ill-health curves, and also develop a single ill-health curve for each of male and female pensioners.

Figure 10



The following steps have been taken to find appropriate retirement health adjustment factors:

1. Filter out records with bad retirement health type quality flags using the data sets for MPA and FPA, MPN, and FPN (i.e., male/female pensioners with non-ill-health retirement type), and MPI and FPI (i.e., male/female pensioners with ill-health retirement type).
2. Fit six different logistic regression models including linear, quadratic, and cubic, both directly in age and the reciprocal of the age (as explained in subsection 0) to each data set in step 1.
3. Obtain the ratio of the predicted rates for non-ill-health to the all-health curves (i.e., FPN/FPA and MPN/MPA) and the ratio of all-health curves to the ill-health curves (i.e., FPA/FPI and MPA/MPI) for all models in step 2 for both genders.
4. Select the best set of fitted models under step 2 for males and females by considering the following criteria:
   - Compare the pensioner age-only curves discussed in section 4 based on all retirement health types to those fitted in step 2 above using only data with good-quality flags for retirement health type. We expect similar patterns between these curves.
   - Visually inspect the ill-health curves created in step 2 above. We expect these rates to increase monotonically by age.
   - Inspect that the ill-health and non-ill-health curves converge, by age, to the all-health curves.
   - Confirm the ratio of ill health and non-ill health to all health is monotonic with age.
   - Perform the AIC/BIC for goodness-of-fit tests.

Considering the above criteria, we selected the linear model with age-functional form for both male and female pensioners for our ill-health and non-ill-health curves.

We then applied the adjustment factors determined in step 3 above to obtain 142 different curves across different combinations of covariates for male and female pensioners with non-ill-health and ill-health retirement types.

7 Curve extension

While our fitted curves generally cover only ages 60 to 95, we've also extended the curves to older and younger ages to enable their use in valuing and projecting pension benefits.

We have extended our curves linearly on the logit scale to age 115 by assuming that the force of mortality[2] at age 115 is one. The maximum attainable force of mortality is chosen following Thatcher, et al., (1998) and CMI, (2009). Age 115 was selected as the maximum lifespan based on observed ages at death for the oldest-lived Canadians.

We chose a linear approach to extend our curves to older ages for the following reasons:

- **Consistency:** A linear approach preserves the relative mortality among different curves at age 95 and therefore avoids the risk of curves crossing. In other words, it does not introduce any inconsistency between different stratum/covariate profiles.
- **Monotonicity:** By adopting a linear approach, we are ensuring that the curves are increasing monotonically with age.
- **Continuity:** The extended curves do not exhibit any discontinuity at age 95.
- **Reasonability:** Rates of mortality on the logit scale are assumed to be a linear function of age. Similar to Thatcher, et al., (1998), this linear assumption holds fairly well within our data set.
- **Simplicity:** The linear approach is relatively simple to implement, which itself fits with the immateriality of the mortality rates at older ages for valuing pension benefits.
- **Consistency with experience:** While mortality experience beyond age 95 was quite limited, the linear approach produced mortality rates that generally fell within the 95 percent confidence intervals for crude mortality rates for ages above 95.

Although a linear approach fulfils all of the above properties, it may not necessarily provide a smooth extension. Note that in the application of the resulting baseline mortality curves (i.e., a pension plan valuation), it would be appropriate to set the mortality rates at age 115 to one.

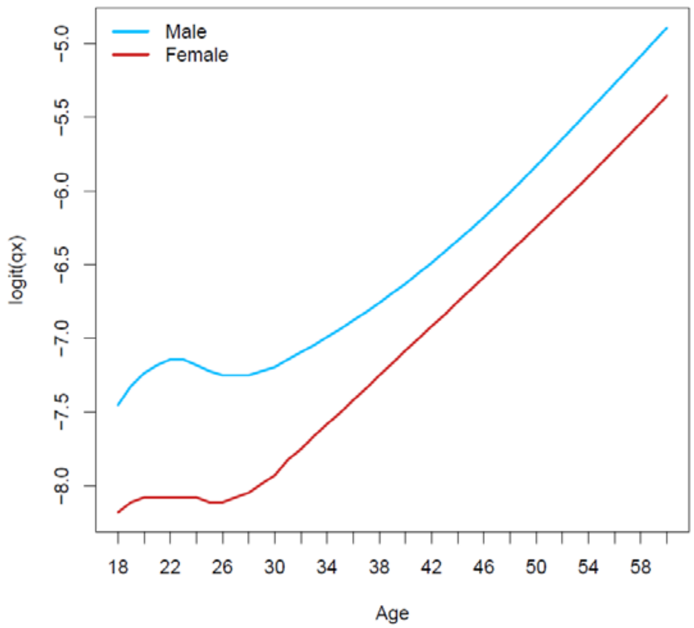
For younger ages, we again extrapolated our curves linearly, with the linear extrapolation converging to Canadian general population mortality rates. This method implicitly assumes that occupation, pension, salary, and longevity groupings for pensioner populations have an insignificant effect at young ages, owing to the uncertainty about a young person's future occupation, wealth, and socio-economic status.

Figure 11 shows mortality rates by gender on a logit scale for the Canadian general population using Statistics Canada, (2016) life tables for 2010 to 2012 from age 18 to 60. We can see that for males, extrapolating linearly from the minimum of our fitting age range (i.e., 60) to age 18 would predominantly overstate mortality rates given the observed non-linear trend at young ages. For females, linear extrapolation clearly overstates mortality rates across the whole age range of 18 to 60

[2] Let μ_x be the force of mortality for a member age x. Assuming $\mu_{x+u} = \mu_x$ for all $0 < u < 1$ and any positive integer x, we have $q_x = 1 - e^{-\mu_x}$.

due to less prevalence of so called "accident hump" effect over age 18 to 25. In addition, we observe that mortality rates for both genders seem to behave quite linearly from age 60 to about age 30. Therefore, we adopted a linear interpolation to general population rates at age 30. Mortality rates at ages below 30 were based on general population using Statistics Canada life tables for 2010 to 2012.

Figure 11

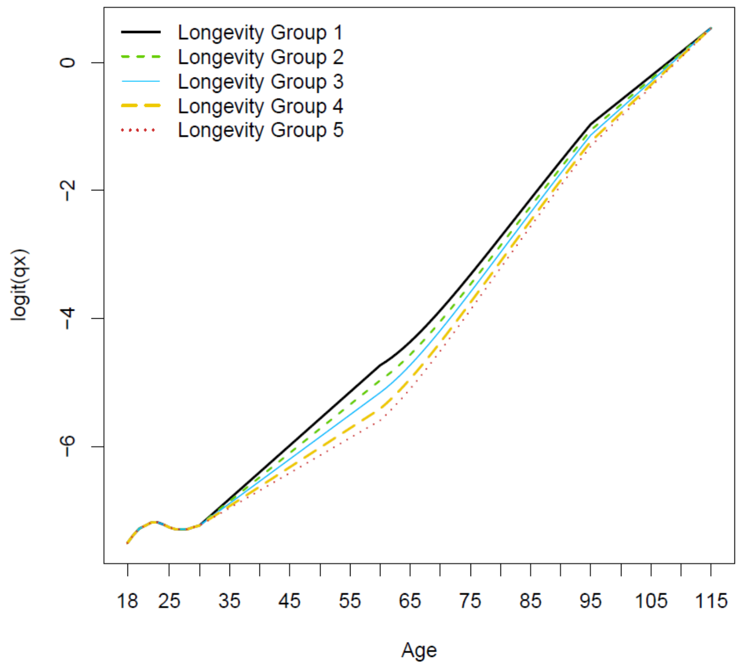


Because we calibrated our curves over the period of 2012 to 2014, the developed mortality rates can be referenced to 2013 as the mid-year of the calibration period. Therefore, to interpolate to age 30, we need to have mortality rates for general population at 2013 for both men and women.

Unfortunately, the latest available mortality rates from Statistics Canada (at the time of performing this research) covered the period of 2010 to 2012, with a mid-year of 2011 as a reference period. Consequently, we needed to project Statistics Canada rates forward two years. To do so, we forecasted mortality rates by applying the original Lee-Carter model, as described in Lee & Carter, (1992). We used data from the Human Mortality Database, (2015) by considering a fitting calendar year range of 1981 to 2011, and a fitting age range of 18 to 60 for both males and females. Mortality improvements (obtained from the fitted Lee-Carter model at each age in 2012 and 2013) are then applied to the published death rate from Statistics Canada to project the Statistics Canada death rates to 2013.

Figure 12 shows extended curves on logit scale for male pensioners with all retirement health condition for the five longevity groupings.

Figure 12



8 Parameter uncertainty

Using the design matrix of the fitted logistic regression models, one can find the standard errors of the estimated parameters. Therefore, it is possible to assess the uncertainty related to the parameter estimates and predicted mortality rates. For illustration, Figure 13 shows mean estimates of the mortality rates (solid lines) and 95 percent prediction errors (dashed lines) for female pensioners at lowest and highest salary bands. We observe that the prediction errors form a narrow range around the mean estimates over the majority of the fitted age range, with small increases at the youngest and oldest ages, highlighting the low uncertainty of our parameter estimates and therefore the high reliability of our curves.

From a practical point of view, it is also appealing to measure uncertainty around estimated life expectancies. To do so, we applied a Monte Carlo simulation by first generating a sample from predicted q_x using the fitted model and the estimated mean and variance. Curtailed period life expectancies as well as annuity factors at different ages can then be calculated using the simulated mortality rates. Note that we consider uncertainty only due to the parameter estimations and ignore variability that comes from the binomial distribution. The main reason for this is to isolate parameter uncertainty and the fact that binomial variations depend on the available exposure data within each combination of covariates.

Figure 13

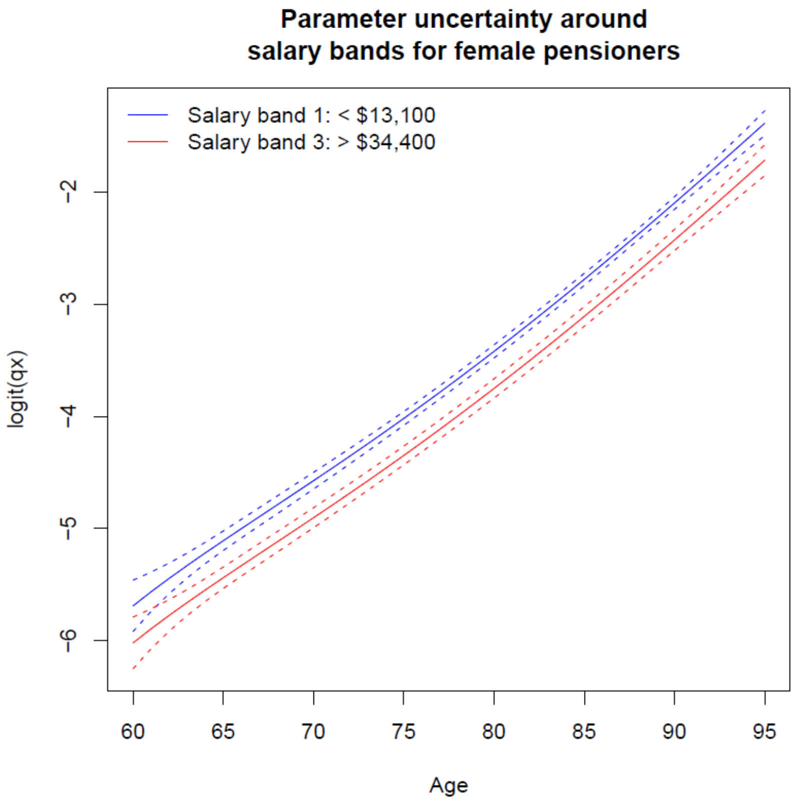


Table 20 shows estimated curtailed period life expectancy at age 65 for female pensioners with all retirement health by three salary bands. The table summarizes the fitted curtailed period life expectancy according to the model, and based on the Monte Carlo simulation together with standard deviations and 95 percent confidence intervals.

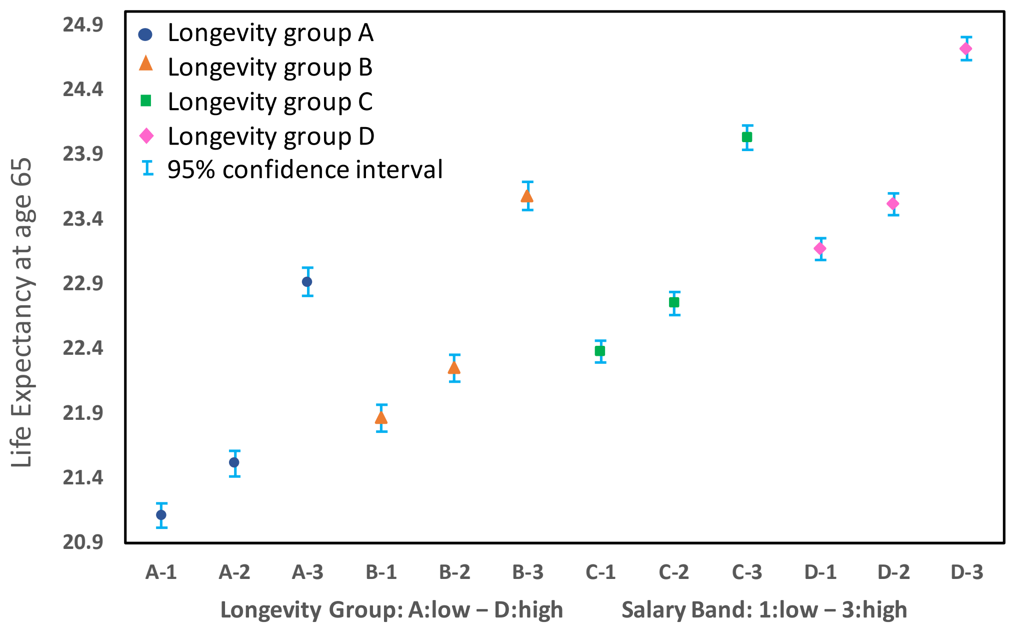
Table 20

Fitted and simulated FPA curtailed period life expectancy at age 65 with 95% confidence intervals					
Salary bands	**Fitted life expectancy**	**Mean**	**Standard deviation**	**Lower 95% confidence interval**	**Upper 95% confidence interval**
1	22.09	22.09	0.03	22.02	22.16
2	22.51	22.51	0.04	22.44	22.58
3	23.89	23.88	0.04	23.80	23.97

For illustration purposes, Figure 14 shows the estimated life expectancies at age 65 and 95 percent confidence intervals for female pensioners by different combinations of longevity groupings and salary bands (e.g., A-1 means longevity group A and salary band 1). Here, longevity groups A and D correspond to the shortest and longest living pensioners, respectively. There are only four longevity groups instead of five because we have determined that longevity groups D and E are not statistically significant within this calibration and therefore they are grouped together.

We can clearly see that mean life expectancies are increasing by each salary band within each longevity group, as well as between longevity groups overall. Although life expectancies at salary bands 1 and 2 are close, they are distinct as their 95 percent confidence intervals are not overlapping. Overall, Figure 14 shows that the uncertainty around the estimated parameters is relatively small.

Figure 14



9 Reducing baseline mortality measurement risk

In this section, we illustrate how increasing the number of mortality rating factors reduces baseline mortality measurement risk. We do this through an analysis of actual-over-expected mortality experience, first based on our highest order mortality factors (i.e., age, gender, and pensioner type), and then using all factors we found to be predictive.

Figure 15 shows the actual-over-expected mortality ratios for 2012 to 2014 experience for the plans that participated in this study. This includes a subset of plans that were not part of the data included to create the baseline mortality models. The expected mortality for this analysis accounts only for the age, gender, and pensioner type characteristics of the members underlying the different plans. The actual-over-expected ratios for each plan or group of plans from a single plan sponsor (i.e., each dot) are compared to an approximate 95 percent confidence interval developed based on the volatility of each plan's experience, which is largely a function of plan size. We can see that there is a substantial degree of dispersion in the results, with about half of the dots falling outside the confidence interval.

Figure 15

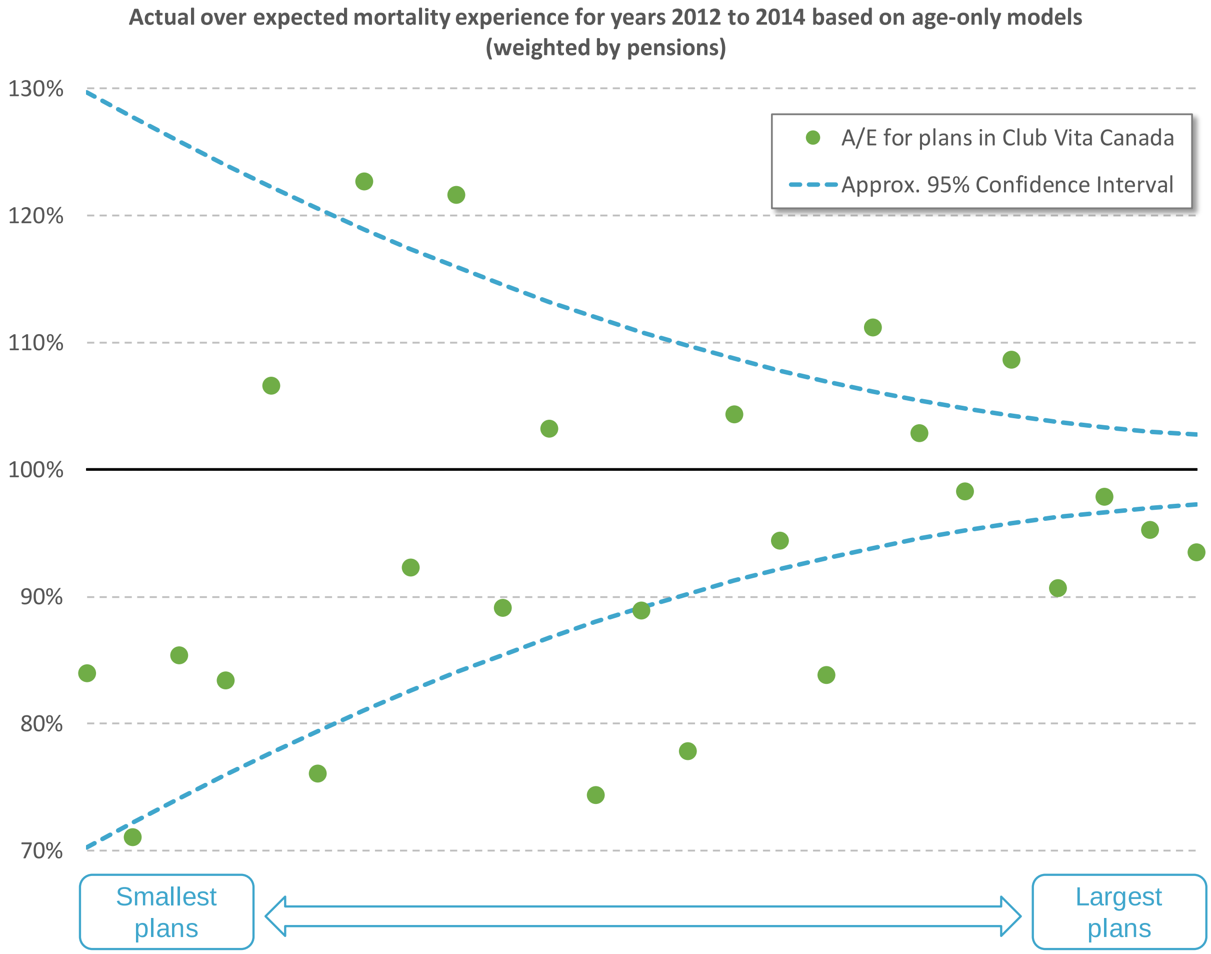
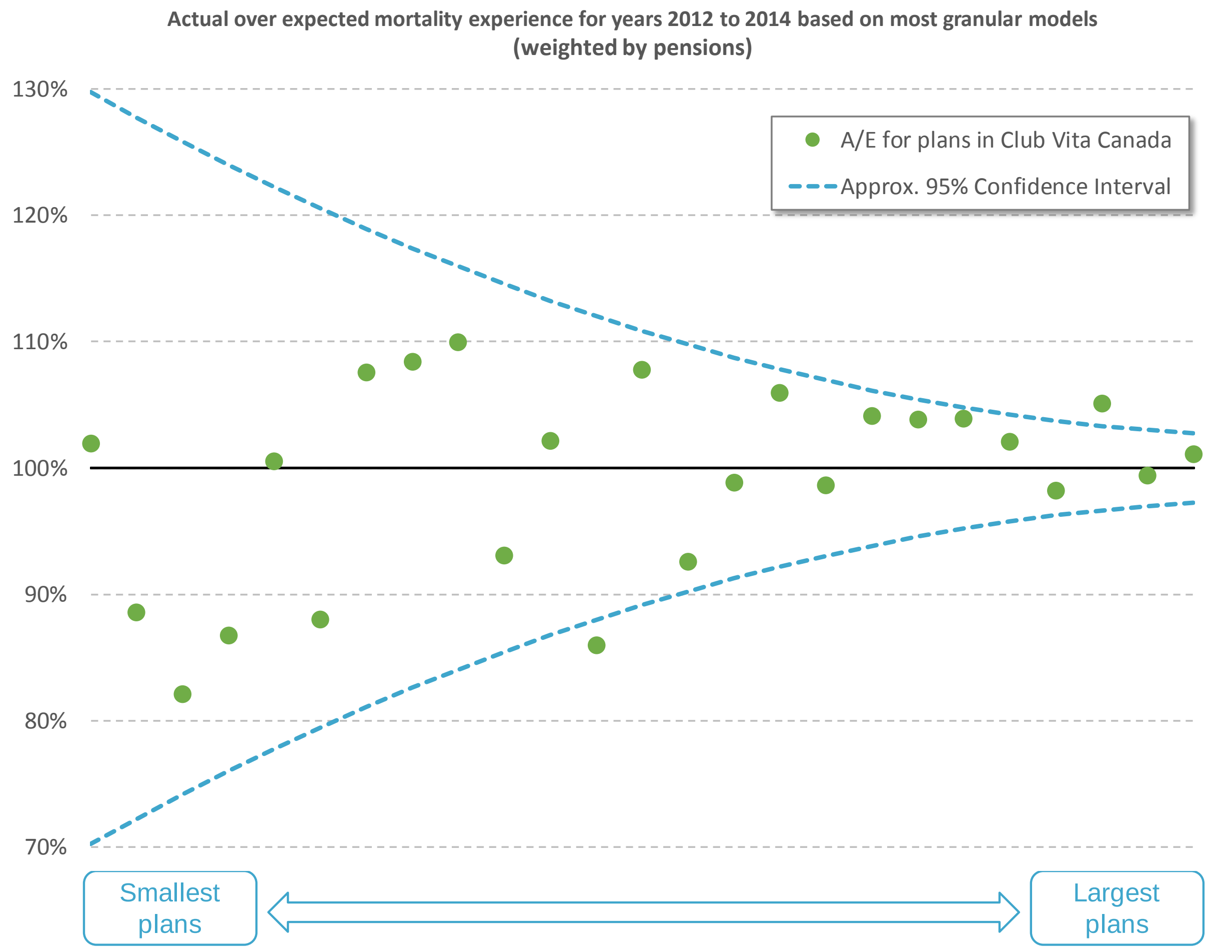

Actual over expected mortality experience for years 2012 to 2014 based on age-only models
(weighted by pensions)
130%
120%
110%
100%
90%
80%
70%
A/E for plans in Club Vita Canada
Approx. 95% Confidence Interval
Smallest plans
Largest plans

Figure 16 shows a similar analysis but now the expected mortality experience is based on all the mortality factors available for each plan. With these additional rating factors, almost all dots fall within the confidence interval. Therefore, by using a range of mortality factors, much more accurate plan-specific baseline mortality assumptions can be developed.

Figure 16



10 Comparison with the CPM study

In this section, we provide some comparisons of the results of our research to that of the CPM study.

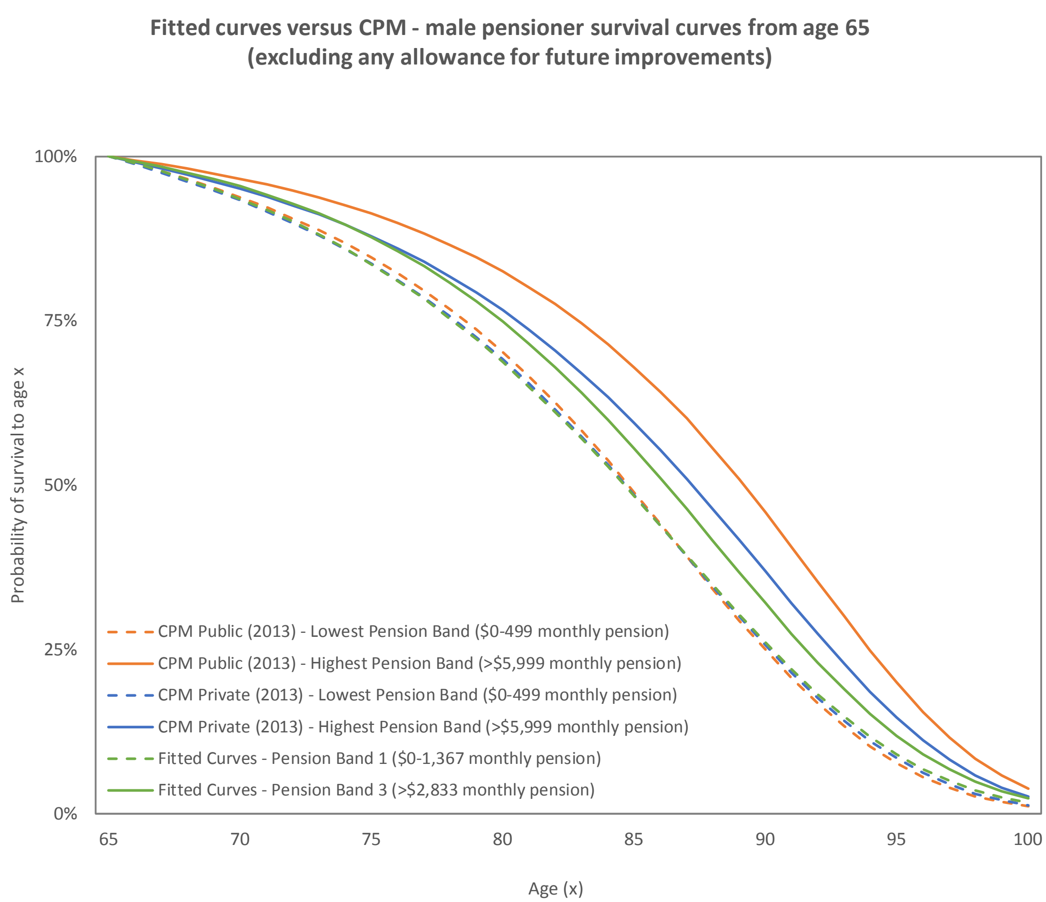
First, in Table 21 we compare the actual pensioner deaths (for males and females combined) in our data set over the period 2012 to 2014 to the deaths expected based on the CPM mortality tables (public, private, and combined projected to the applicable year). This analysis is done on pensioners only since the CPM study excluded survivors, and the actual and expected deaths take into account pension amounts (i.e., deaths are pension weighted as opposed to lives weighted). We first observe that the actual deaths were 13 percent greater than expected by the CPM public table, regardless of whether the plan members in our data set were employed by a public sector or private sector employer. The CPM combined table also expects lower mortality compared to our data set. The actual deaths are most in line with the CPM private table, irrespective of sector type.

Table 21

AoE analysis using CPM mortality tables			
AoE ratio	CPM Combined	CPM Private	CPM Public
Private sector plans	1.07	0.97	1.13
Public sector plans	1.09	0.98	1.13
All plans	1.08	0.98	1.13

Next, we review the differences in the probabilities of survival from age 65 when accounting for the effect of pension amount. Figure 17 shows survival curves from age 65 for male pensioners using the CPM public and CPM private sector tables (projected to 2013) after applying the size adjustment factors for the CPM study's lowest and highest pension size bands, together with the models developed in this paper for male pensioners in pension bands 1 and 3. In this figure, the solid lines represent high pension amounts, while dashed lines indicate low pension amounts.

Figure 17



We can see that the probabilities of survival are quite similar for low pension amounts. However, for male pensioners with high pension amounts, our fitted curves exhibit lower survival rates compared to those developed in the CPM study, particularly for public sector pensioners and private sector pensioners aged 80 and older. Therefore, our data and modelling did not support as wide a range in mortality expectations for male pensioners based on pension amount as the CPM study's pension size adjustment factors. In contrast, our data and modelling supported a somewhat wider range in mortality expectations for female pensioners than the CPM study.

Through the combination of the CPM public and private baseline mortality tables and the corresponding pension size adjustment factors, the CPM study explains a 3.9-year range in period life expectancy at age 65 for male pensioners and a 1.6-year range for female pensioners. Our modelling did not find sector type to be a predictive factor, but based on pension amount alone, our fitted models resulted in period life expectancy at age 65 ranging 1.5 years for male pensioners and 2.3 for female pensioners. As discussed in subsection 2.4, the classification of pension amounts into pension bands proved challenging due to a lack of clear differentiation in mortality for small pension amount clusters. When accounting for all of the factors included in our final fitted models, an 8.6-year range in period life expectancy is explained for males and a 7.7-year range for females. The ability to explain a wider range in mortality expectations allows pension plans to much better assess and measure the mortality of their plan members by accounting for their unique mortality characteristics.

In Table 22 and Table 23, we provide annuity factors based on our final fitted curves for the all retirement health calibration for male and female pensioners aged 65 in 2017 using a discount rate of 4 percent per annum. The tables show the annuity factors for all the different combinations of pension bands, longevity groups, and occupations for the final fitted model. Note that our calibration process explained in section 5 resulted in the combination of pension bands 1 and 2, and longevity groups D and E for male pensioners when pension band, longevity group, and occupation are all available, and occupation is not incorporated for female pensioners when pension amount and longevity group are available. The table also includes the corresponding CPM private and CPM public annuity factors, and the percentage change when comparing the annuity factors for the final fitted curves to each of the CPM factors. When determining the CPM annuity factors by pension band, the CPM size adjustment was determined based on the median monthly pension amount underlying each pension band. For the under $34,000 male pensioner band, the applicable median annual pension amount was $16,303, and therefore $1,359 per month, which equates to the CPM study's third pension band (see Table *2* for the median amounts for other pension bands). All annuity factors have been calculated using the CPM-B improvement scale projected from 2013.

Table 22

Comparison of annuity factors for male pensioners

Pension band	Longevity grouping	Occupation type	Fitted curve factor	CPM Private factor	Fitted curve/CPM Private	CPM Public factor	Fitted curve/CPM Public
<$34k	A	Blue collar	12.77	13.70	-6.8%	13.85	-7.8%
<$34k	A	White collar	13.53	13.70	-1.2%	13.85	-2.3%
>=$34k	A	Blue collar	13.14	14.17	-7.3%	14.60	-10.0%
>=$34k	A	White collar	13.87	14.17	-2.1%	14.60	-4.9%
<$34k	B	Blue collar	13.07	13.70	-4.6%	13.85	-5.6%
<$34k	B	White collar	13.81	13.70	0.8%	13.85	-0.3%
>=$34k	B	Blue collar	13.43	14.17	-5.2%	14.60	-8.0%
>=$34k	B	White collar	14.14	14.17	-0.2%	14.60	-3.1%
<$34k	C	Blue collar	13.48	13.70	-1.6%	13.85	-2.6%
<$34k	C	White collar	14.19	13.70	3.6%	13.85	2.5%
>=$34k	C	Blue collar	13.83	14.17	-2.4%	14.60	-5.2%
>=$34k	C	White collar	14.51	14.17	2.4%	14.60	-0.6%
<$34k	D	Blue collar	13.89	13.70	1.4%	13.85	0.3%
<$34k	D	White collar	14.56	13.70	6.3%	13.85	5.1%
>=$34k	D	Blue collar	14.22	14.17	0.3%	14.60	-2.6%
>=$34k	D	White collar	14.85	14.17	4.8%	14.60	1.8%

Table 23

Comparison of annuity factors for female pensioners							
Pension band	**Longevity grouping**	**Occupation type**	**Fitted curve factor**	**CPM Private factor**	**Fitted curve/CPM Private**	**CPM Public factor**	**Fitted curve/CPM Public**
<$13.1k	A	n/a	14.15	14.85	-4.7%	14.98	-5.5%
$13.1k - $34.4k	A	n/a	14.28	15.05	-5.1%	15.18	-5.9%
>$34.4k	A	n/a	14.95	15.26	-2.1%	15.38	-2.8%
<$13.100	B	n/a	14.53	14.85	-2.2%	14.98	-3.0%
$13.1k - $34.4k	B	n/a	14.66	15.05	-2.6%	15.18	-3.4%
>$34.4k	B	n/a	15.27	15.26	0.0%	15.38	-0.7%
<$13.1k	C	n/a	14.72	14.85	-0.9%	14.98	-1.7%
$13.1k - $34.4k	C	n/a	14.84	15.05	-1.4%	15.18	-2.2%
>$34.4k	C	n/a	15.42	15.26	1.1%	15.38	0.3%
<$13.1k	D	n/a	15.03	14.85	1.2%	14.98	0.4%
$13.1k - $34.4k	D	n/a	15.15	15.05	0.6%	15.18	-0.2%
>$34.4k	D	n/a	15.68	15.26	2.8%	15.38	2.0%
<$13.1k	E	n/a	15.42	14.85	3.9%	14.98	3.0%
$13.1k - $34.4k	E	n/a	15.53	15.05	3.2%	15.18	2.3%
>$34.4k	E	n/a	16.00	15.26	4.9%	15.38	4.1%

Table 22 andTable 23 show differences in annuity factors ranging from -10.0 percent to 6.3 percent for male pensioners and -5.9 percent to 4.9 percent for female pensioners. These results highlight that individual pension plan liabilities based on our fitted baseline mortality curves could differ greatly from those based on the CPM study.

11 Key findings

In this paper, we determined which rating factors available within pension plan administration systems can be used to differentiate baseline pensioner mortality and how a logistic regression model can be used in a GLM framework to develop models for Canadian pensioner baseline mortality. Our key findings are directly applicable for pension and post-retirement benefit mortality assumptions, and can be summarized as follows:

- After testing a series of mortality rating factors as outlined in subsection 2.1, we found age, gender, pensioner type, retirement health, geodemographics, salary at retirement (or earlier exit), pension amount, and occupation to be the most significant covariates for isolating variations in life expectancies among pension plan members.
- Public sector versus private sector employment was not found to be statistically significant in the presence of other considered covariates.
- A pension plan member's postal code is a vital piece of information that can be used to capture differences in geodemographics.

- Longevity/lifestyle groups can be created by applying statistical clustering methods to geodemographic segments, and these have been found to be a very important predictive factor.
- Salary at retirement (or earlier exit) can provide a better means of capturing the influence of affluence on mortality compared to pension amount for both males and females.
- Pensioners who are disabled at retirement have different mortality patterns than those that are not disabled at retirement, particularly at younger retirement ages.
- Male pensioners with the longest life expectancy are expected to outlive those with the shortest life expectancy by 8.6 years at age 65 (based on period life expectancy). For female pensioners, this differential is 7.6 years.
- By capturing the impact of a wide range of mortality factors, baseline mortality expectations can be tailored to the longevity characteristics of individual pension plans. This in turn allows plans to reduce their risk of incorrectly measuring their longevity exposure and creating inappropriate mortality assumptions.

Acknowledgements

The authors would like to thank Prof. Johnny S.H. Li from the University of Waterloo for his valuable advice and insight provided during the development of the models discussed in this paper, as well as his thoughtful review and comments on this paper. We also wish to express our sincere gratitude to the Research Executive Committee of the Canadian Institute of Actuaries and to their appointed reviewers for all their comments and suggestions.

References

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J., (1984). *Classification and Regression Trees.* s.l.:s.n.

Burnham, K. P. & Anderson, D. R., (2002). *Model Selection and Multimodel Inference.* s.l.:Springer-Verlag.

Chiang, C. L., (1984). *The Life Table and its Applications.* Florida: Robert E Krieger Publ Co.

Cleveland, W. S., (1979). Robust locally weighted regression and smoothing scatterplots. *J. American Statistical Association,* Volume 74, p. 829–836.

Cleveland, W. S., (1981). LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician,* 35(1).

CMI, (2009). *Continuous Mortality Investigation Self Administered Pension Schemes Mortality Committee (CMI). Working paper 35: The graduations of the CMI Self-Administered Pension Schemes 2000-2006 mortality experience. Final 'S1' series of mortality tables.,* s.l.: s.n.

Cox, D. R., (1972). Regression models and life tables. *Journal of the Royal Statistical Society,* 34(2), pp. 187-220.

CPM, (2014). *Final Report Canadian Pensioners' Mortality,* s.l.: Canadian Institute of Actuaries.

Forfar, D. O., McCutcheon, J. J. & Wilkie, A. D., (1988). On Graduation by Mathematical Formula. *Journal of the Institute of Actuaries,* pp. 1-149.

Fox, J., (2010). *An R and S-Plus Companion to Applied Regression.* 2 ed. s.l.:SAGE.

Harrell, F. E., (2001). *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York: Springer-Verlag.

Hosmer, D. & Lemeshow, S., (1980). Goodness-of-fit tests for the multiple logistic regression model. *Communications in Statistics,* 9(10), pp. 1043-1069.

Human Mortality Database, (2015). [Online]
Available at: http://www.mortality.org/

Lee, R. D. & Carter, L. R., (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association,* 87(419), pp. 659-671.

Lowrie, W. B., (1982). An extension of the Whittaker–Henderson method of graduation. *TRANSACTIONS OF SOCIETY OF ACTUARIES,* Volume 34.

Madrigal, A. M. et al., (2011). What longevity predictors should be allowed for when valuing pension scheme liabilities?. *British Actuarial Journal,* 16(1), pp. 1-38.

Mendenhall, W., (1982). *Statistics for Management and Economics.* Boston: Duxbury Press.

Perks, W., (1932). On Some Experiments in the Graduation of Mortality Statistics. *Journal of the Institute of Actuaries,* pp. 12-57.

Raftery, A. E., (1995). Bayesian Model Selection in Social Research. *Sociological Methodology,* Volume 25, pp. 111-163.

Report, R.-2. M. T., (2014). *RP-2014 Mortality Tables Report,* Schaumburg, IL: Society of Actuaries.

Richards, S. J., (2008). Applying survival models to pensioner mortality data. *British Actuarial Journal,* 14(2), pp. 257-303.

Ripley, B. D., (1996). *Pattern Recognition and Neural Networks.* s.l.:Cambridge University Press.

Statistics Canada, (2016). *Life Tables, Canada, Provinces and Territories, 2010 to 2012, no. 6.* [Online] Available at: http://www.statcan.gc.ca/pub/84-537-x/84-537-x2016006-eng.htm

Sullivan, A. R. & Fenelon, A., (2014). Patterns of Widowhood Mortality. *he Journal of Gerontology, Series B: Psychological Sciences and Social Sciences,* 69(B(1)), p. 53–62.

Thatcher, A. R., Kannisto, V. & Vaupel, J., (1998). The force of mortality at ages 80 to 120. *Odense Monographs on Population Ageing,* Volume 5.

Venables, W. N. & Ripley, B. D., (2002). *Modern Applied Statistics with S.* 4th ed. New York: Springer.

William, J. C., (1971). *Practical Nonparametric Statistics.* New York: John Wiley & Sons.

Appendix

A.1. Actual-over-expected test

The actual-over-expected test is developed to check if, at an overall level, the ratio of actual deaths to expected deaths is significantly different from 1 or not. The 95 percent confidence interval can be constructed using quantiles of a standard normal distribution.

A.2. Beta-binomial confidence interval

The beta-binomial confidence interval is calculated using a Bayesian approach—assuming beta distribution as a prior information—and applying a Monte Carlo simulation as follows:

1. Find total number of deaths (A_x) and total number of exposures (ETR_x) at each age (x).
2. For each age x and at each simulation:
   - Generate a random sample from a beta distribution with the shape parameter equal to $A_x + 0.5$ and scale parameter equal to $ETR_x - A_x + 0.5$.
   - Generate a binomial random sample with number of observation equal to –truncated- ETR_x and probability of success equal obtained in previous step. This can represent a hypothetical number of deaths.
   - Find the hypothetical mortality rate by using simulated number of deaths in previous step.
3. At each age, find mean, 5th , and 95th percentiles across all simulations.

A.3. Chi-squared test

The chi-squared test compares the actual versus expected deaths and tests if the distribution of the observed number of deaths is significantly different than the expected number of deaths. The test statistics are defined as

$$\chi^2 = \sum_x \frac{(A_x - E_x)^2}{E_x},$$

where A_x and E_x stand for the actual number of deaths and expected number of deaths at age x, respectively. Although the chi-squared test statistic is defined by age, it can be successively generalized to test the death distribution across other categorical variables (e.g., whether the distribution of deceased members in pension band 3 differs significantly from those in pension band 1). The test statistics approximately follow a chi-square distribution.

A.4. Cumulative deviations test

The cumulative deviations test is developed to check if, at an overall level, the observed number of deaths minus the expected number of deaths is significantly different from zero. The total number of observed deaths is compared with the total number of expected deaths using the variance of the binomial distribution. The test statistic is then compared with the appropriate quantile of a normal distribution to detect any atypical results.

A.5. Kolmogorov-Smirnov test

The KS test is designed to detect and prevent the over-smoothing of mortality rates. By comparing the cumulative number of observed deaths and expected deaths, the KS test determines if the observed and expected deaths are coming from the same distribution. The confidence level is set to be 95 percent. Details of the KS test can be found in William, (1971).

A.6. Life expectancy comparison test

We have compared fitted life expectancies (using fitted mortality rates) at ages 65, 75, and 85 with the corresponding 95 percent confidence intervals of the crude life expectancies using the observed data. The test fails if the fitted life expectancies are not within appropriate 95 percent upper and lower bounds of the crude life expectances. Details of life expectancy calculations can be found in Chiang, (1984).

A.7. Likelihood function

The likelihood function is the product of the binomial probabilities for each member weighted by exposures, i.e.,

$$L = \prod_{i=1}^{n} \left({q_x}^{ETR_{xi} \cdot Y_i} (1 - q_x)^{ETR_{xi} \cdot (1 - Y_i)} \right),$$

where q_x is defined in eq. (1). When estimating model parameters, we maximize above function using the iteratively reweighted least squares method as explained in Fox, (2010).

A.8. Monotonic test

The monotonic test ensured that the fitted mortality rates are increasing by age, as would be expected for mortality rates during retirement years.

A.9. Run test

Ideally, we would like to develop a model that captures the shape of mortality (by age) that is observed in the historical data. The run test is developed to check the randomness of residuals and to detect any deviations in the shape of mortality. The run test considers groups of residuals with the same sign (or run) and compares them with the number which would be expected if residuals are all random. Interested readers are referred to Mendenhall, (1982) for details.

A.10. Serial correlations test

The serial correlations test is similar to the sign test but the magnitude of the deviation is also taken into account. First deviations Z_x are determined at each age. Next, serial correlation r_k is calculated for a particular lag k which is then multiplied by the square root of the number of ages. The test fails if the obtained value is large enough compared to the quantile of the standard normal distribution.

A.11. Sign test

Once we fit our models, we do not expect that the number of observed deaths will deviate from the number of expected deaths (obtained from the fitted model) by age with a systematic pattern. In other words, the difference between observed and expected values (residuals) should be randomly distributed around zero over the fitting age ranges. The sign test monitors the number of positive residuals and compares it with a binomial distribution to detect any particular trend by age. The sign test will pass if the number of positive residuals are within the lower 2.5th percentile and upper 2.5th percentile of the binomial distribution.

A.12. Standardized deviations test

The chi-squared test above may fail to pick up small and consistent under or over estimations (e.g., a few large deviations can be potentially offset by a large number of small deviations). The standardized deviations test is designed to capture these and is based on Z_x scores as defined by

$$Z_x = \frac{A_x - E_x}{\sqrt{E_x}}.$$

The Z_x is calculated at each age and then is divided into intervals. Then the number of observed Z_x that falls within each interval is calculated and compared, using a chi-squared test, to the expected number of deaths that fall in each interval under the normal distribution.