Data underpinning CV22 Canadian VitaCurves

In this paper we provide an overview of the data used to calibrate the 'CV22' edition of our Canadian VitaCurves, which were released in July 2023. The data used to calibrate our VitaCurves is referred to as the **VitaBank**. Our Canadian VitaBank is a pool of the data contributed by all of our Canadian club member plans.

In this paper, we provide an overview of the Canadian VitaBank (**Sections 1 to 3**), discuss our data quality control process (**Section 4**) and summarize the data volumes underpinning this edition of VitaCurves (**Section 5**).

Details of the methods we use in calibrating VitaCurves are set out in our 'Calibrating CV22 Canadian VitaCurves' paper.

1 The heritage of the data

The data collected by Club Vita in Canada covers more than a million participant records, including current and former active participants (i.e. employees accruing pension benefits), deferred pensioners (i.e. former employees with an entitlement to receive a pension when they ultimately retire), pensioners, survivors and deceased members.

The VitaCurves we calibrate relate to mortality conditional on being in retirement (i.e. in receipt of a pension). Consequently the analysis for the production of the VitaCurves focuses on the records for

- pensioners and survivors currently in receipt of benefits; and
- deceased members (who at the time of death were either a pensioner or survivor).

2 A rich and diverse dataset

This section summarizes the plans and data contributing to our CV22 Canadian VitaCurves. Sections 2.3 onwards are restricted to the individuals exposed to risk during the period we have used to calibrate our CV22 Canadian VitaCurves (i.e. 2018 through 2020) and the age ranges included in the calibration (see the accompanying 'Calibrating CV22 Canadian VitaCurves' document for more details).

2.1 Range of different plans

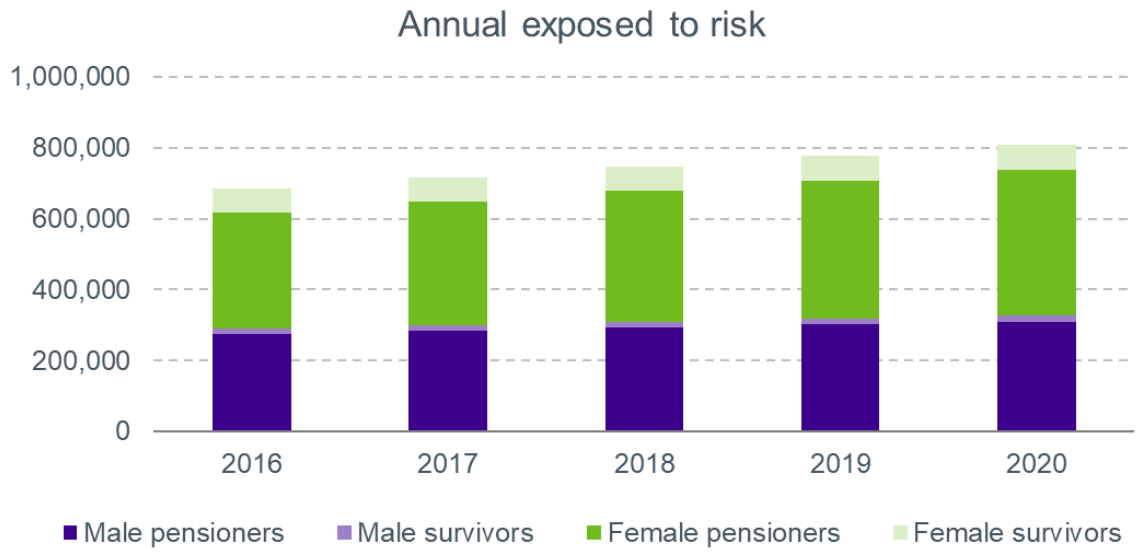
92 different defined benefit pension plans contribute data covering the 2018-2020 period. These plans cover a range of different sizes, from smaller plans with less than 2,500 pensioners and beneficiaries (58 plans) through to very large plans with more than 30,000 pensioners and beneficiaries (9 plans).

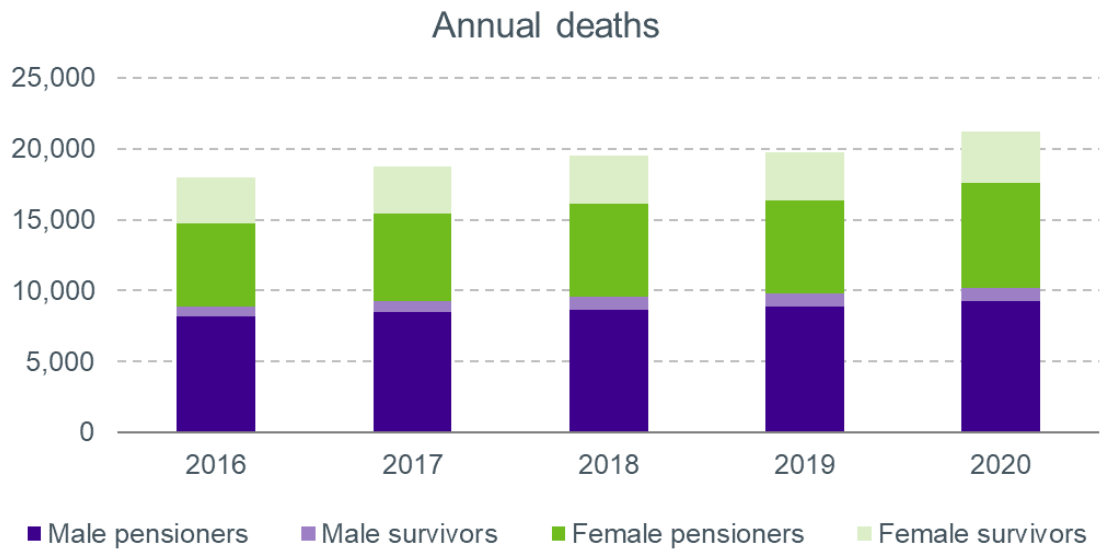
The table to the right shows the distribution of the plans by size.

Number of pensioners and beneficiaries	Number of plans
<2,500	58
2,500 – 9,999	19
10,000 – 29,999	6
30,000 or above	9

2.2 Profile of lives and deaths over time

A key factor in calibrating life tables is to have enough lives and deaths to enable robust calibration of mortality rates. The charts below highlight the split of lives-based exposures and deaths in our Canadian VitaBank over the years 2016 to 2020.

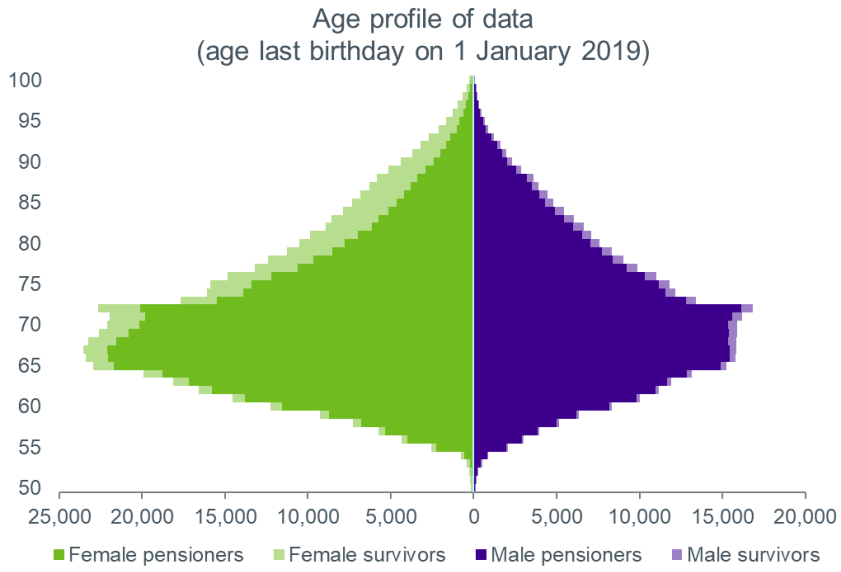




We can see a gradual growth in both exposures and deaths for both males and females over 2016 to 2020. In calibrating CV22 we include data over 2018 to 2020, and it is therefore encouraging to see the steady increases in exposure over that period.

2.3 Age profile of data

The dataset spans a wide range of ages, from young pensioners in their 50s through to pensioners and survivors aged over 100. The chart below illustrates the age profile of the pensioner and survivor data, based upon age last birthday on January 1, 2019[1].



We can see that the pensioner population rises rapidly from age 60 upwards. The step-up at age 65 is consistent with it being a popular retirement age.

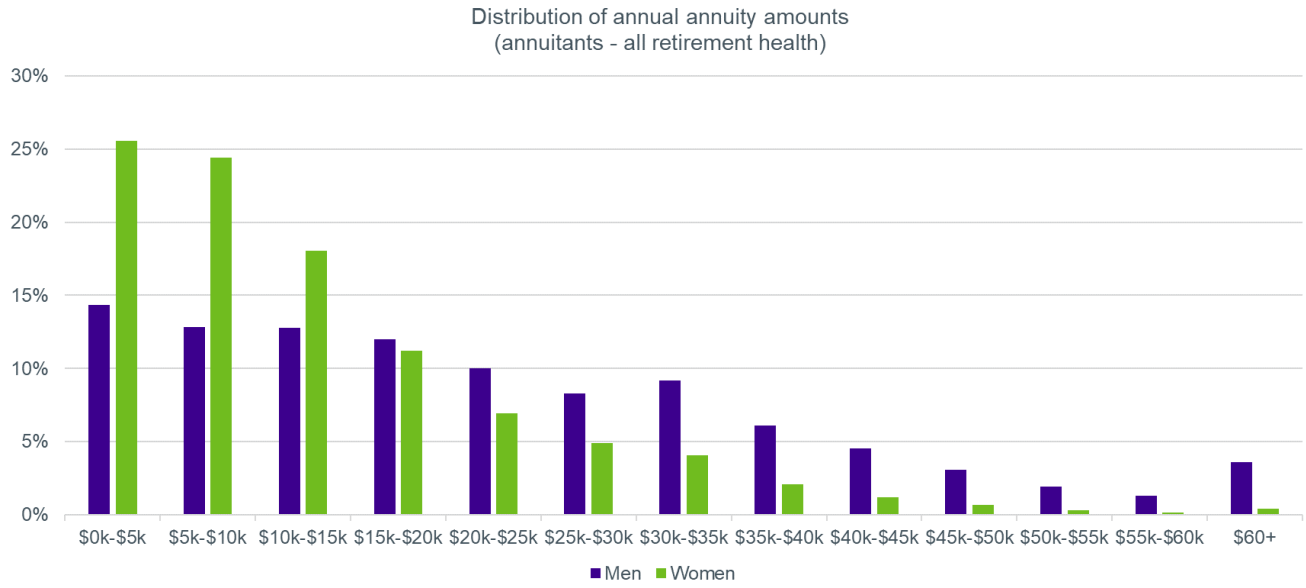
The chart also highlights how:

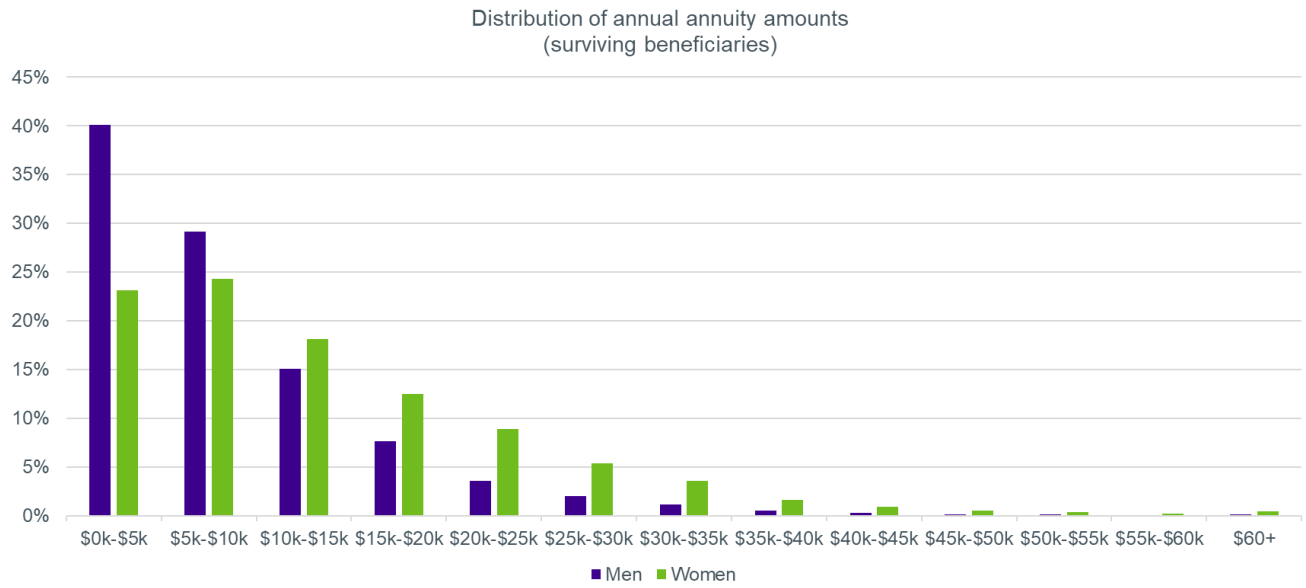
- there are significantly more pensioners than survivors;
- the pensioner data has slightly more women than men;
- the number of male survivors is much smaller than the number of female survivors.
- female survivors are heavily skewed toward older ages (as expected, as the pensioner has to die before the associated survivor benefits come into payment).

[1] The chart shows lives-based exposures over 2019. We use 2019 as this is the mid-year of the period covered by the VitaCurves calibration.

2.4 Pension amounts

The charts below illustrate the distribution of annual pension amounts in the CV22 Canadian VitaCurves data for pensioners and surviving beneficiaries[2]. In each case this has been split by gender.





We see that:

- pension amounts are skewed to lower amounts;
- there is a significant 'tail' to pension amounts, particularly for male pensioners;

[2] Based on lives-based exposures over the calibration period.

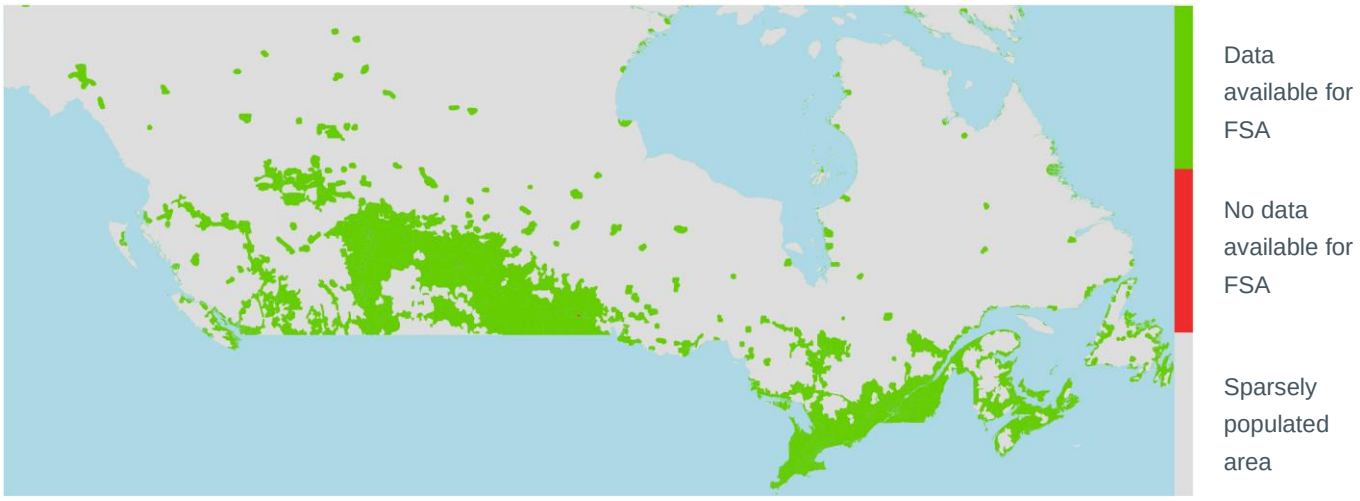
- pension amounts payable to survivors tend to be lower, reflecting that most plan participants elect less than a 100% continuation of their pension to their surviving beneficiary; and
- pension amounts are typically lower for women.

2.5 Geographical diversity

The data underpinning our Canadian VitaCurves also provides a broad geographical coverage, as highlighted by the map below.

Each Forward Sortation Area ('FSA', the first three characters of a postal code) has been coloured in green if pensioners and/or survivors living in that FSA were included within the data underpinning our CV22 Canadian VitaCurves.

Data Coverage by Forward Sortation Area (FSA)



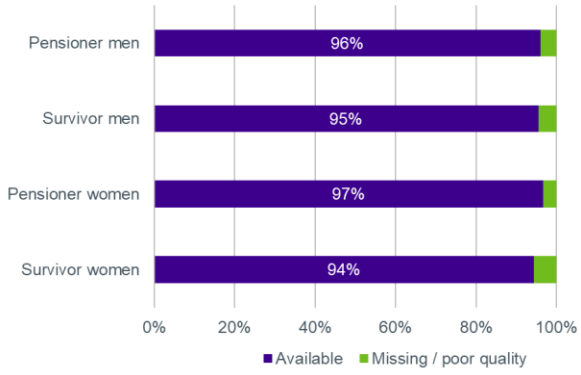
There are approximately 1,600 Forward Sortation Areas in Canada, of which only a handful are not represented in our data (i.e. those marked red on the map above). Sparsely populated areas are masked in light grey colour.

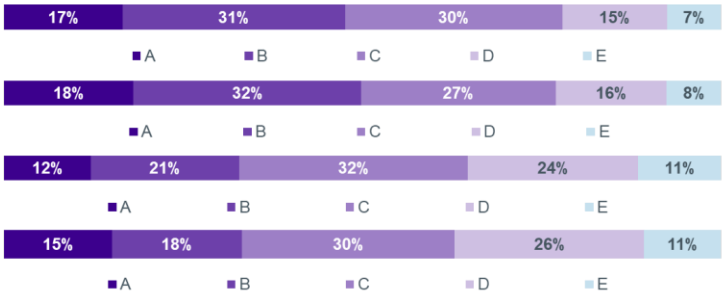
3 Availability and spread of key longevity predictors

Our analysis identifies the impact on mortality of five key longevity predictors, separately for pensioners and surviving beneficiaries and for males and females. It is therefore important to have good availability of data, and a spread between the values taken for each of these predictors (postal code based longevity groups, pension and salary amounts, occupation and pension form). We can see from the charts below that this is the case[3].

3.1 Longevity group

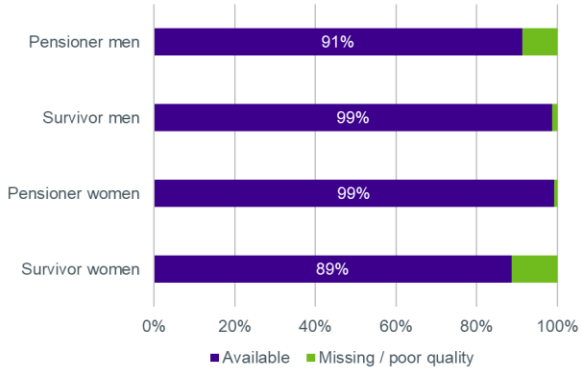
Our most detailed models rely on postal code information in order to identify a longevity group based upon lifestyle proxies. We can see from the charts below that postal code data is generally available for over 95% of the data. We also see how the most 'extreme' groups (those with the longest / shortest life expectancies) represent a smaller proportion of the overall population distribution. This is consistent with our experience in the UK and US.

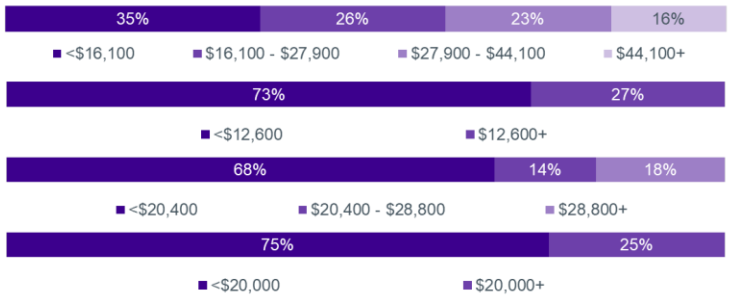




3.2 Pension amount

Pension amount is available for almost all participants in the dataset. The chart below illustrates on the left-hand side the proportion of pensioners and survivors for whom we have reliable pension amount, and on the right, the distribution of exposures by each pension band for male and female pensioners and survivors.

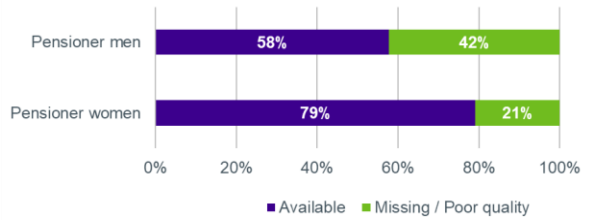


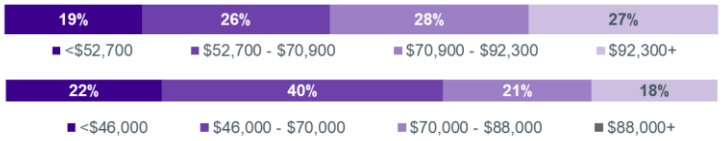


[3] Note that throughout this section the percentages reflect the share of the lives-based exposure over the calibration period.

3.3 Salary amount

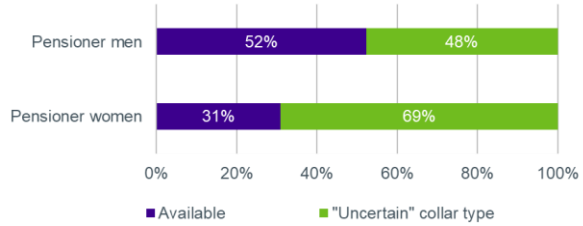
Last known salary is available for around 58% of pensioner men and 79% of pensioner women. Exposures are more evenly distributed across different salary bands for male than female pensioners.

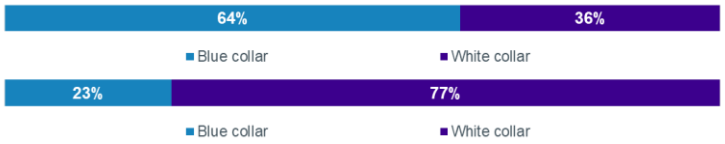




3.4 Occupation

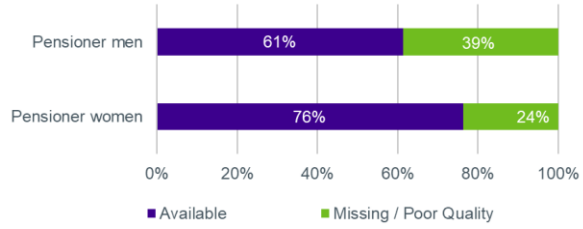
Some pension schemes in our data set identify whether their members worked in blue collar or white collar roles, where available. However the coverage of this variable is lower than for the other predictors used, with around half of pensioner men and 1/3 of pensioner women having a specified collar type. For those who are provided with a collar type, 64% of pensioner men are classed as blue collar, while 77% of pensioner women are classed as white collar.





3.5 Form of pension

The form of pension has been provided for a higher proportion of pensioner women than pensioner men. In both cases, where information is provided, a higher proportion of pensioners have joint life pension amounts, although for women the split is fairly even.



4 Data pre-processing

Only data which has been through our initial quality control process enters our statistical analysis. The data quality control process is designed to ensure the data for each pension plan is as reliable as possible. The individual participant data is checked for each plan. Basic checks are carried out relating to the plausibility of consecutive dates, missing data, etc. Where data problems are found and the data can be cleaned, or the information is non-essential, the record will be retained; otherwise the record will be excluded from VitaBank.

Data is also screened against a range of quality criteria, to ensure that any obvious errors, inconsistencies, or artificial biases which may arise as a facet of administrative processes do not distort our analysis. Individual records are flagged as a result of this screening, and may be excluded from the calibration. Details of our data processing and additional quality controls can be found in the Canadian VitaCurves technical documentation.

5 Data contributing to CV22 Canadian VitaCurves

The data volumes contributing to the CV22 calibration of Canadian VitaCurves are set out in the table below. Note that the data used covers the 2018 to 2020 period and is restricted to 'good quality' data for any specific longevity predictor where it is used in the model.

Further breakdowns of the data by participant categories and longevity predictors are provided in Appendix A.

	Exposed to risk (life years)	Deaths
Pensioner men	906,967	26,801
Pensioner women	1,170,351	20,524
Survivor men	46,752	2,764
Survivor women	212,541	10,434
Total	**2,336,611**	**60,523**

6 Want to know more?

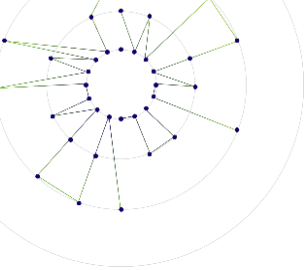
If you have any questions regarding this data document or would like additional details regarding our methods for fitting our CV22 Canadian VitaCurves, please contact our team. We would be delighted to hear from you.

Saeed Ahmadi PhD
Saeed.Ahmadi@clubvita.net

Conor O'Reilly
Conor.OReilly@clubvita.net

July 2023

For and on behalf of Club Vita LLP



Reliances and Limitations

In this paper (the 'Research'), Club Vita LLP has provided an overview of the data used for the calibration of the seventh generation of Club Vita's Canadian VitaCurves. The Research is based upon Club Vita LLP's understanding of legislation and events as of July 2023 and therefore may be subject to change. Future actuarial measurements may differ significantly from the estimates presented in the Research due to experience differing from that anticipated by the demographic, economic or other assumptions. The Research should not be construed as advice and therefore not be considered a substitute for specific advice in relation to individual circumstances and should not be relied upon. Where the subject of the Research refers to legal matters, please note that Club Vita LLP is not qualified to give legal advice, therefore we recommend that you seek legal advice if you are wishing to address any legal matters discussed in this Research. Please be advised that Club Vita LLP (not its respective licensors) does not accept any duty, liability or responsibility regarding the use of the Research, except where we have agreed to do so in writing.

When read along with the 'Calibrating CV22 Canadian VitaCurves' paper, this paper complies with the Financial Reporting Council's Technical Actuarial Standard (TAS) 100: Principles for Technical Actuarial Work.

Appendix A: Volumes of data used in our models

The sections below describe the volumes of data contributing to the CV22 calibration of Canadian VitaCurves for each participant category and longevity predictor. In interpreting these tables please note that:

- The exposed to risk is measured as life years;
- The data relates to the 2018 to 2020 calendar years used for the calibration;
- The data is restricted to ages included in the calibration; and
- The data volumes for each longevity predictor is restricted to participants with 'good' participant and plan level quality flags.

Pensioner men

	Exposed to risk	Deaths
Initial Data covering calibration age range	906,967	26,801
Pension amount	828,021 (91%)	24,163 (90%)
Postal code	870,799 (96%)	25,516 (95%)
Collar Type	473,944 (52%)	13,866 (52%)
Pension Form	556,194 (61%)	16,048 (60%)
Three variables (pension, postal code, pension form)	530,951 (59%)	15,162 (57%)

Pensioner women

	Exposed to risk	Deaths
Initial Data covering calibration age range	1,170,351	20,524
Pension amount	1,160,036 (99%)	20,265 (99%)
Postal code	1,131,483 (97%)	19,481 (95%)
Collar Type	362,396 (31%)	5,803 (28%)
Pension Form	892,746 (76%)	14,943 (73%)
Three variables (pension, postal code, pension form)	863,166 (74%)	14,204 (69%)

Survivors, men

	Exposed to risk	Deaths
Initial Data covering calibration age range	46,752	2,764

Survivors, women

	Exposed to risk	Deaths
Initial Data covering calibration age range	212,541	10,434
Pension amount	188,324 (89%)	9,095 (87%)
Postal code	200,546 (94%)	9,702 (93%)
Two variables (pension, postal code)	177,272 (83%)	8,426 (81%)