Data underpinning CV22 US VitaCurves

Our "Zooming in on ZIP codes" paper introduced our VitaCurves; a series of mortality tables derived from pooling pension plan data which enable plans to use a baseline longevity assumption tailored to the true diversity of their participants.

The data underpinning the CV22 edition of Club Vita's US VitaCurves ("CV22") is a combination of a subset of the Mercer Longevity Database ("MILES") dataset, and, for the first time, directly sourced data from Club Vita member plans (together the 'Club Vita dataset').

In this paper, we provide an overview of the Club Vita dataset (**Sections 1 to 3**), describe the processing and quality controls applied to the data (**Sections 4 and 5**) and summarize the data volumes underpinning the VitaCurves (**Section 6**).

1 The heritage of the data

1.1 MILES data

The MILES data has been collected from a range of qualified defined benefit ("DB") pension plans. These private sector plans are drawn from Mercer's client base (and several other plan sponsors) and each plan has consented to the onward sharing of their longevity data with third parties.

The data relates purely to in payment annuities, and includes annuitants, disabled retirees and surviving beneficiaries of deceased retirees. Note that only plans who have committed to providing regular data updates in the future have been included in the calibration data set, to avoid future volatility in data provision.

For each plan that participated, valuation census data was collected for each plan year end covering the experience period (including the plan year ends at the start and end of the experience period). For most plans, their year-end coincided with calendar year ends so were perfectly aligned with the experience period.

1.2 Directly sourced data

Directly sourced data from Club Vita's first subscriber plans has been included for the first time in the CV22 calibration. The data provided annually by each plan undergoes thorough data validation on receipt, including analysis to identify potential unreported deaths.

The data relates to both future and in payment annuities, and includes annuitants, disabled retirees and surviving beneficiaries of deceased retirees.

As we continue to grow the collection of plans in the Club Vita dataset, we envisage that we will in the future be in a position to calibrate VitaCurves entirely using our directly sourced data.

In the rest of this document we refer to the combination of MILES and directly sourced data, used in calibrating CV22 VitaCurves, as the' Club Vita dataset'.

2 A rich and diverse dataset

The sections below summarize the data available in the Club Vita data set. Sections 2.3 onwards are restricted to individuals exposed to risk during the period to which we have calibrated our CV22 US VitaCurves (i.e. 2018 through 2020) and to the age ranges included in the calibration (see the accompanying “Calibrating CV22 VitaCurves” document for more details). Annuitants with unknown retirement health are also excluded from the data.

2.1 Range of different plans

211 different private sector defined benefit pension plans contribute data covering the 2018-2020 period. These plans cover a range of different sizes, from smaller plans through to very large plans.

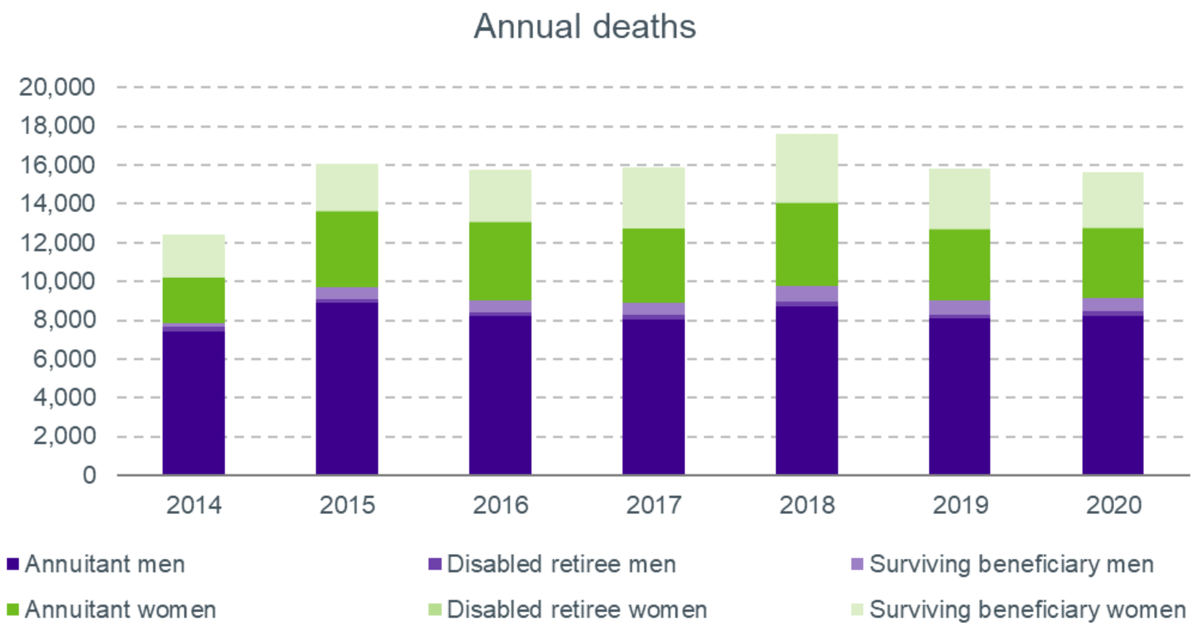
Note that the directly sourced data includes plan data provided by a number of insurers, each of which has identified the relevant ‘sub-plans’ that they took onto their books. For the purposes of this section we have treated each such sub plan as a distinct plan.

The table (right) shows the distribution of plans by size.

Number of annuitants and beneficiaries (ages 50-100)	Number of plans
<=5,000	179
5,001 – 10,000	10
10,001 – 15,000	11
15,001 – 20,000	1
20,000+	8

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 exposures and deaths over the years 2014 through 2020.



We can see how:

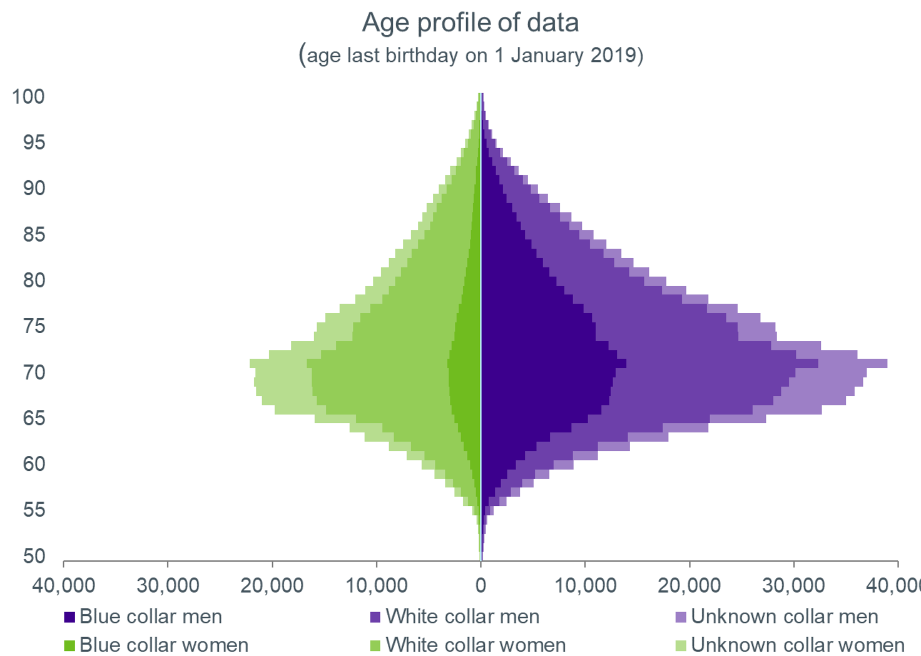
- Data volume has increased from 2014 through 2018, owing to the addition of extra plans plus the continued maturation of pension plans.
- Data volume has decreased year on year between 2018 to 2020, owing to a combination of some plans undertaking PRT activity and some exiting the dataset.
- Deaths are very similar over 2019 and 2020, despite the drop in exposure in 2020 relative to 2019. This is consistent with the COVID-19 pandemic resulting in higher mortality rates in 2020.
- There is limited data on disabled retirees reflecting the eligibility conditions for disability provisions[1].

[1] For some MILES plans none of the annuitants are marked as disabled. In these circumstances a small number of the annuitants may in fact have been disabled retirees. We have retained this data to provide maximum insight into mortality rates and differentials, accepting that if there are disabled retirees not identified as such, this may very slightly elevate mortality rates particularly at younger ages. However for directly sourced plans we have excluded plans where the retirement health is not provided (or where there is a material difference between the levels of coverage between living and dead annuitants).

2.3 Age profile of data

The dataset spans a wide range of ages from young beneficiaries through to annuitants aged over 110.

The chart to the right illustrates the age profile of the annuitant data for each collar group. Data excludes disabled retirees and surviving beneficiaries, and is based upon age last birthday on January 1, 2019[2].



The annuitant population rises rapidly from age 60 upwards. The step-up at age 64 is consistent with 65 being a popular retirement age, and so the retirees covered by this dataset (i.e. those that happened in 2020 or earlier) have reached 64 or older as at January 1, 2019.

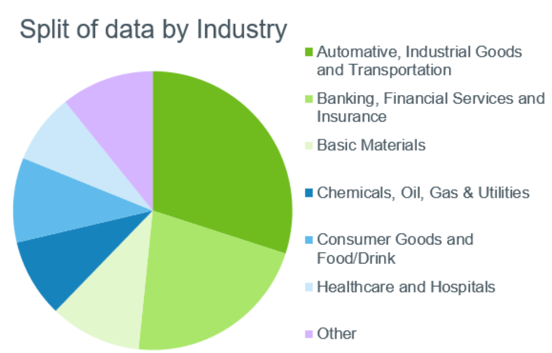
The reduction in lives between age 72 and 73 is consistent with those aged 72 being the youngest generation from the baby boom that followed the return home of G.I.s serving in WWII, to a booming US economy and the support of the G.I. bill.

The chart also highlights how:

- consistent with pension plan participation rates, there are more annuitant men than women;
- among men there is a broadly even split between blue collar (38%) and white collar (46%) annuitants; and
- the data for annuitant women is primarily white collar (62%) with a modest proportion of blue-collar annuitants (15%)

2.4 Industry mix

The data includes a mix of plans from different industries. The industry classification is based upon the 6-digit business unit code of the plan sponsor recorded on the Form 5500 used in ERISA disclosures. These codes have then been pre-grouped into six broad industries, along with an "other" group to capture plans from industries with insufficient data volumes to analyze separately.

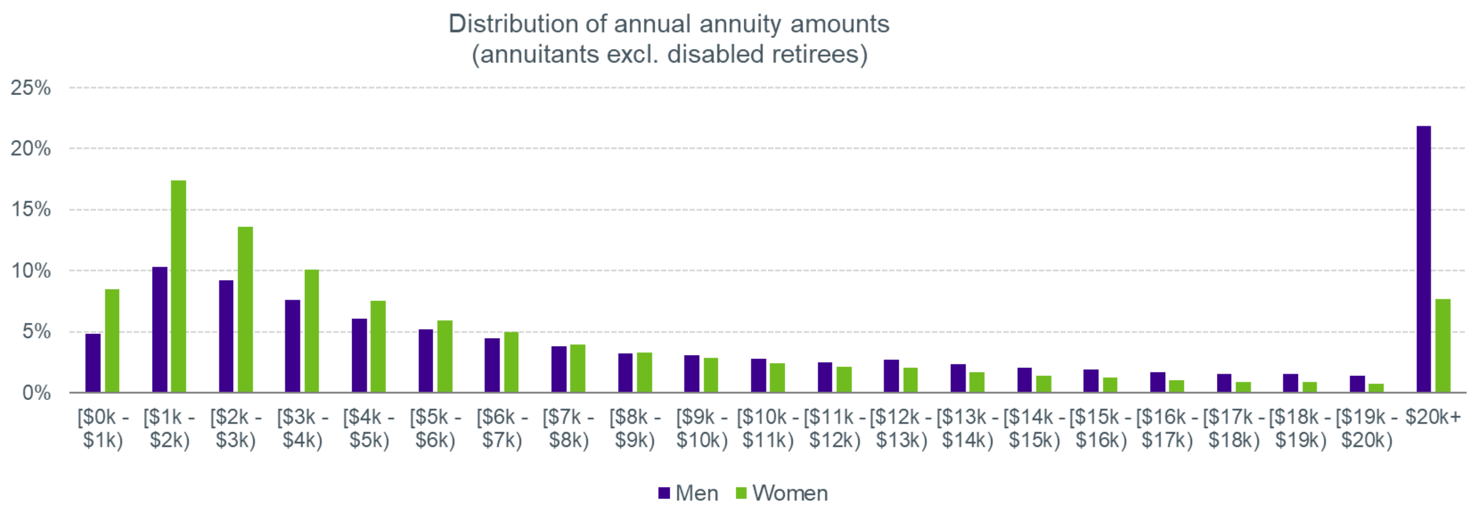


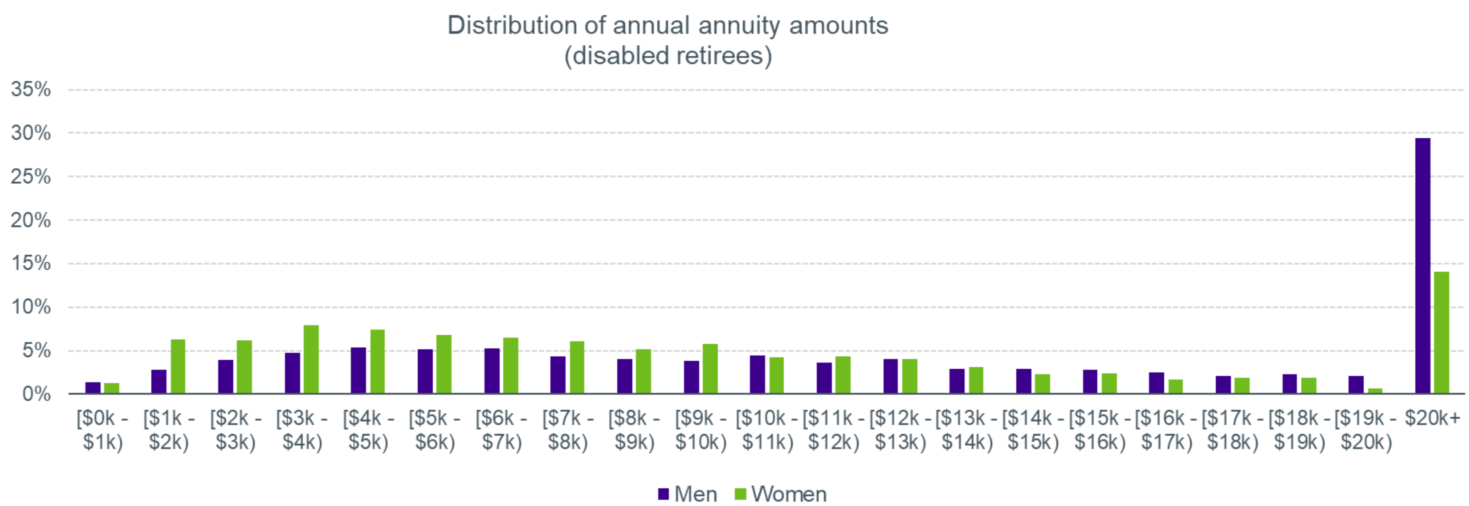
[2] We use 2019 as this is the mid-year of the period covered by the CV22 VitaCurves calibration. Note that, in the case of annuitants and surviving beneficiaries who died during 2018 but contribute to the exposed to risk in that year, we have shown them at the age they would have been on 1 January 2019, had they survived.

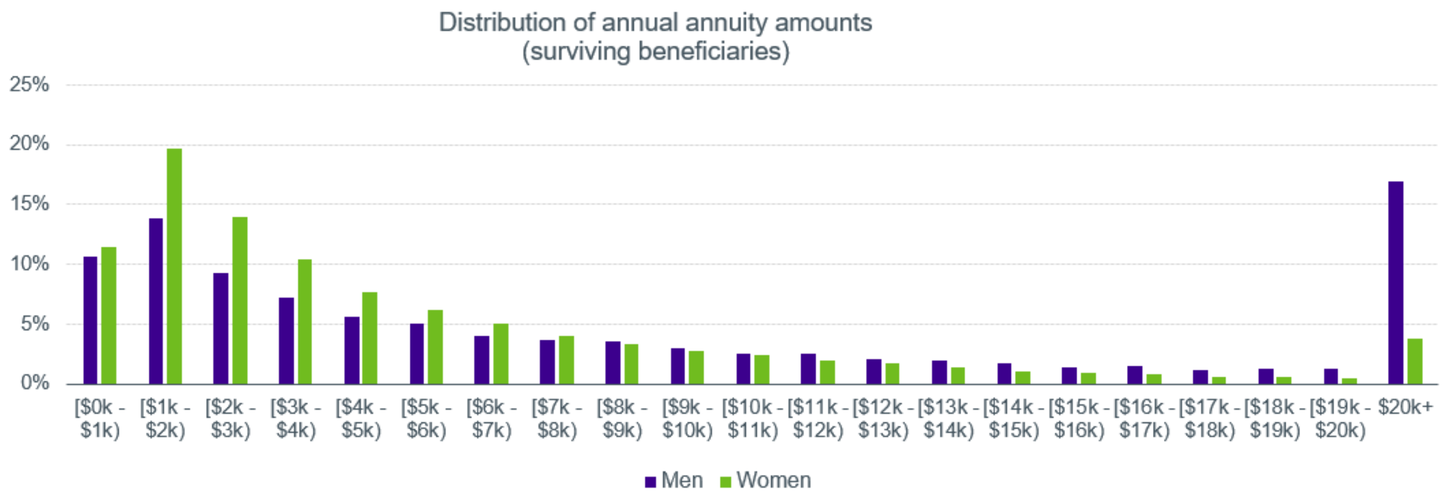
2.5 Annuity amounts

The charts below illustrate the distribution of (annual equivalent) annuity amounts in the data for annuitants, disabled retirees and surviving beneficiaries. In each case this has been split by gender. We see how:

- annuity amounts are skewed to lower amounts reflecting that modest pension incomes can be achieved either by short service or low pay
- there is a significant tail to annuity amounts
- annuity amounts are more evenly distributed for disabled retirees, reflecting the enhanced benefits sometimes payable to disabled retirees
- annuity amounts payable to surviving beneficiaries tend to be lower, reflecting that most plan participants elect less than a 100% continuation of benefit to their surviving beneficiary
- annuity amounts are slightly lower for women – reflecting historical labor force participation and potentially historical pay differentials where the benefit is linked to salary



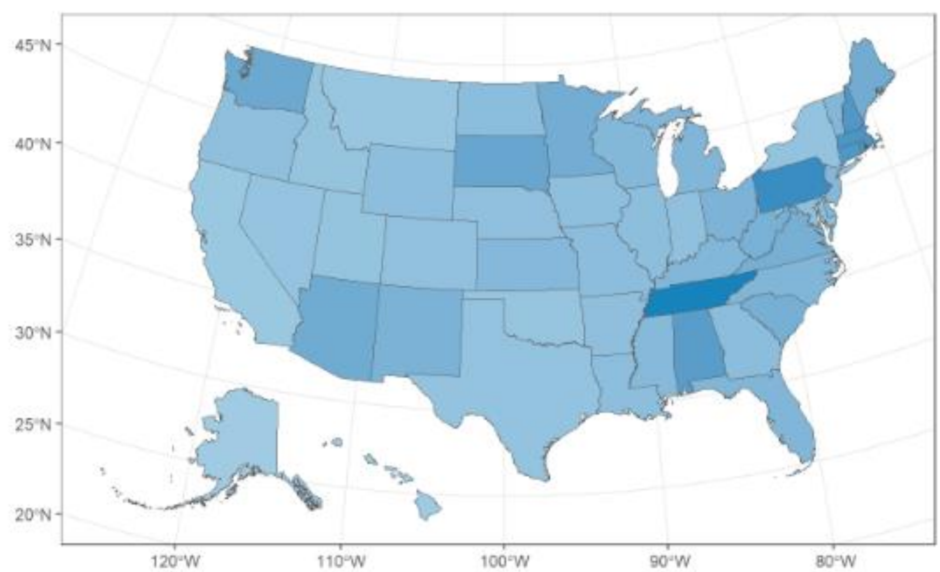




2.6 Geographical diversity

The data has strong geographical diversity, with data from all the states, and is strongly representative of the spread of the US population between states.

This can be seen from the graphic to the right which colors each state in line with the proportion of that state's *total* population which is represented by individuals in the dataset. The darker the color, the greater the proportion of that state's population which is represented in the data. (The state of residence of each individual is based upon the first 3-digits of a participant's ZIP code, which is available for 77% of the records used in the calibration.)



Note: Darker shaded areas have greater data concentrations as a proportion of the state population.

The majority of the states have between 0.05% and 0.15% of the state population represented in the dataset, indicating a broadly-even geographical mix. Concentrations are highest in Tennessee, Pennsylvania, Rhode Island and Massachusetts. The three lowest concentrations are in District of Columbia (liable to be dominated by public rather than private plan participants), Alaska and Hawaii with the fourth lowest being California (the most populous state with almost 40 million residents[3]).

[3] State populations sourced from the 2010 United States Census: "Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2018"

3 Availability and spread of key longevity predictors

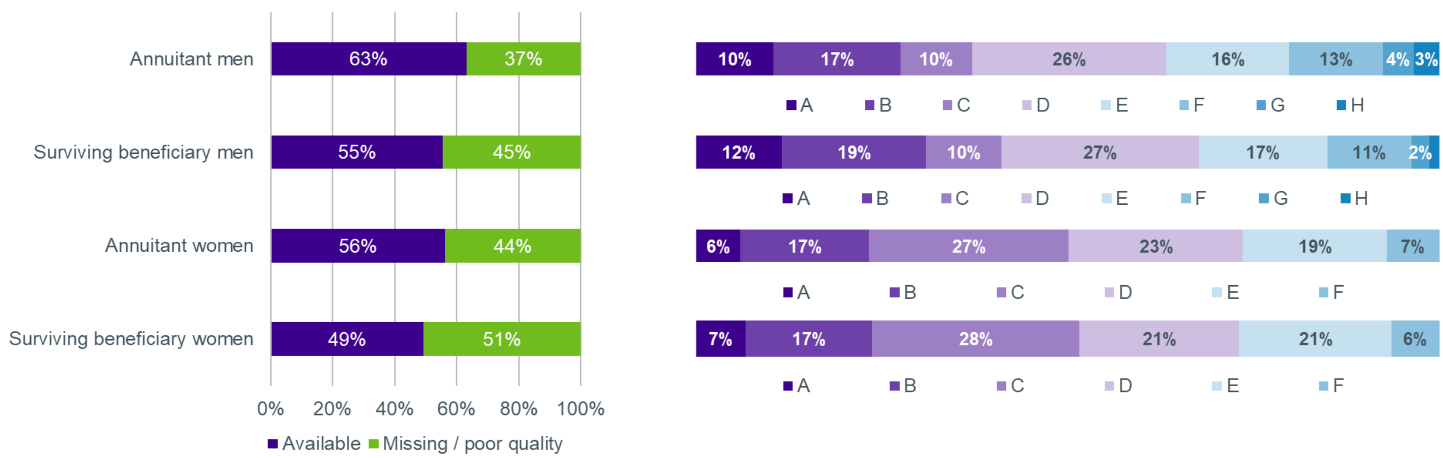
Our analysis identifies the impact on mortality of four key longevity predictors, separately for annuitants and surviving beneficiaries and for men/women[4]. It is therefore important to have good availability of data, and a spread between the values taken for each of these predictors (ZIP+4 based longevity group[5], annuity amount and, for annuitants, collar type, and form of pension). We can see from the charts below that this is the case.

3.1 Annuity

Annuity is available for almost all participants in the dataset, and is well distributed between the specific bands used for our VitaCurves as illustrated by the graphics below. (Note that annuity amounts are expressed as annual income.)

3.2 ZIP+4 longevity group

Our most detailed models rely on availability of ZIP+4 in order to identify a longevity group based upon lifestyle proxies. We can see from the charts below that ZIP+4 is generally available for 60% of the data (including surviving beneficiaries).



---

[4] We have not sought to differentiate mortality among disabled retirees at this stage, owing to the low volumes of data for disabled retirees.

[5] Where ZIP+4 is not available, we have also calibrated ZIP based longevity groups

3.3 Collar type

The collar type of plan participants is determined either at the participant level or the plan level. Where it is determined at the participant level, this is determined by the convention that a participant is:

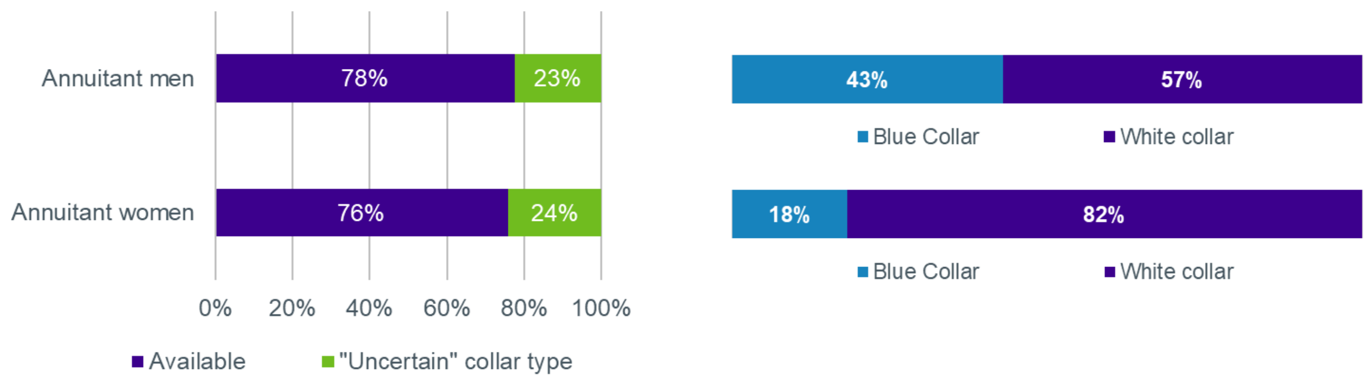
- blue collar if they are *either* hourly-paid or union
- white collar if they are *both* salaried and non-union
- unknown if neither of the above apply

For several plans in the MILES data collar type is not available at the participant level. In these cases, an indicator is provided as to the broad percentage of the participants in that plan/section that are believed to be white collar e.g. 30%. The reliability of this information is partially dependent on a degree of consistent interpretation and judgement across plan providers. Following extensive analysis on the implementation of collar as a rating factor, we have concluded to treat any values other than 0 or 1 as also being of “unknown” collar in the wider context.

For the purposes of fitting the curves we therefore use three collar type groups:

- blue collar – for those plan participants specifically identified as blue collar
- white collar– for those plan participants specifically identified as white collar
- “uncertain” collar – for all other plan participants

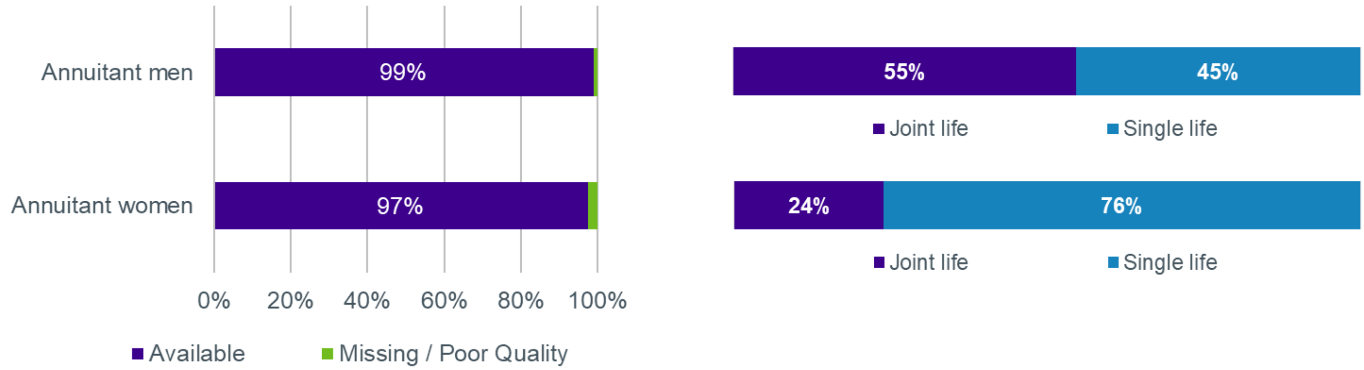
The chart below shows the volume of data where blue or white collar is specifically identified and the split between these for annuitant men and women.



To maximize data volumes, all three types are used when we fit our models which include collar type as a longevity predictor (or *rating factor* in the language of our accompanying modeling paper “Calibrating CV22 VitaCurves”).

3.4 Form of pension

Almost every record has a form of pension recorded as either a single life lifetime annuity or a joint and survivor benefit (joint life).



Annuitant men are split broadly equally between single and joint life members, whereas annuitant women are more strongly dominated by single life records.

It should be noted that our data does not record the “as at” date for form of pension. We therefore do not know whether the form of pension is “at retirement” or is more up-to-date – it appears that the data includes a mix of plans which employ both recording practices.

It is expected that an annuitant recorded as joint life would be married at retirement, and that the majority (but not all) recorded as single life would be unmarried at retirement. If form of pension data reflected the most recent information, then joint life annuitants whose spouse had pre-deceased them would become recorded as single life annuitants. The joint life annuitants at older ages are thus likely to be a mix of “true” joint life annuitants whose spouses are still alive, and “false” joint life annuitants whose spouses are deceased. Similarly, the older single life annuitants are likely to be a mix of members who were unmarried at retirement, married at retirement but elected not to take a joint and survivor benefit, and married at retirement but with a pre-deceased spouse.

We have used the form of pension field for VitaCurve calibration as it is presented in the data, implicitly assuming that it means form of pension “at retirement”, but we draw users’ attention to this ambiguity.

4 MILES data - Processing and quality control applied by Mercer

The MILES pension plan data used in our analysis as part of the Club Vita data set has been collected and processed by Mercer. The data we have received is depersonalized data. To ensure this data is suitable for the purposes of analyzing mortality rates, Mercer has carried out a number of initial quality controls and processing, as set out below.

4.1 Preliminary editing and exclusions

In processing the data Mercer has performed a number of initial edits and exclusions to ensure the suitability of the data for mortality studies. We understand that these include:

- **Death audit:** To ensure a complete record of deaths, and accurate dates of death, Social Security Numbers (“SSN”) were collected for the vast majority of plans. This enabled comparison to the Social Security Death Master File to establish dates of death[6].
- **Age ranges included:** For data prior to 2017/18 records were only included in the data where the beneficiary had an age in the range 50-120 in the year of exposure. For the 2017-18 update, records born before 1900 or after 1966 were excluded.
- **Excluded participants:** The data excludes records in relation to the following participants:
  - **Non-in payment participants** as we are interested in mortality post retirement. *(Note that this exclusion is based upon status at the start of the year and so retirements during the calendar year are excluded until the following year.)*
  - **Certain only beneficiaries** as these participants have usually died
- **Annuity amount:** To ensure comparability of benefit amounts between participants, where the benefit included a Social Security Level Income Option (“SSLIO”) the ultimate benefit level was used. Similarly, participants where the benefit amount included other short-term supplements had these supplements excluded.
- **Excluded data:** Records with missing or invalid data have been screened out according to the following:
  - Invalid or missing Social Security Number (“SSN”) for plans which participated in the SSN-based death audit (and so could not be audited as alive or dead) and for which valuation statuses indicating deaths were not available *(this impacted the 2008-2011 data only and so not the period we have used to calibrate VitaCurves)*;
  - Missing or invalid dates of birth (as they cannot be assigned an age);
  - Missing gender
  - Zero benefit amounts

  It is our understanding this impacted a very modest number of records for the period over which we have calibrated CV22 VitaCurves, and we have no reason to suspect any bias between lives and deaths within these records.

[6] A process which was performed by The Berwyn Group, Inc on behalf of Mercer.

- **Anonymization:** Dates of birth and death were adjusted to the 15th of the month (and for some deaths identified by the plan valuation census but not part of the death audit these were set to mid-year i.e. June 30). Given the broadly uniform distribution of deaths and births over any given month this will not have impacted the modeling.

4.2 Initial exposed to risk and deaths

To calculate mortality rates two key pieces of information are required: How long an individual plan participant was exposed to the "risk" of dying (known as exposed to risk); and whether a participant has died or not.

Exposed to risk

The data received by Club Vita included a computation of the exposed to risk[7] and the "death count" (i.e. an indicator whether the member died) for each individual record in the dataset for each calendar year. While we use the provided status dates to directly calculate exposures in our model, we have relied on the pre-calculated exposure/death counts in determining whether to suppress individual years (see section 5.5).

We understand that the pre-calculated exposures have been calculated by Mercer according to the following approach:

If a record was reported as being in payment as of the beginning of the plan year under consideration, then the record was flagged as being exposed to risk over that plan year i.e. given an exposed to risk value for the year of "1". This would include records who retired in the prior plan year but whose first date of payment was the first day of the plan year in consideration.

The calculation of exposed to risk also takes care with the treatment of the following cases:

- **Retirees during plan year:** For participants who retire during the calendar year and survive to the end of the year the exposure is set to 0 rather than a part year. This is to ensure consistency with the reporting of deaths as any new retirees during the year who die prior to the calendar year end are excluded from the data.
- **Temporary retirements / cessation of payments:** In some circumstances, a retiree may temporarily retire or have payments cease temporarily for some other reason (such as having payments limited to correct past overpayments.) Such records are uncommon, but when they occur, the record is flagged as exposed to risk if a payment was being made as of the beginning of the plan year, and not exposed to risk if a payment was not being made as of the beginning of the plan year. Where a record is not deemed to be exposed to risk in a plan year as a result of this, but is known to have died during the plan year (either via death audit, or via actuarial valuation census at the beginning of the following plan year), this is not recorded as a death but is treated as censored data (so as to avoid introducing a bias and overstating mortality) i.e. the record counts as "0" in both the exposed to risk and death fields contained in the MILES dataset.
- **Deceased just after plan year end:** In rare circumstances, a record may have deceased in the first few weeks after a plan year end and is recorded as deceased rather than alive at the plan year end in the actuarial valuation census file. In these circumstances the exposures and deaths reflect the actual timing of the death i.e. they would be recorded as exposed to risk in the year they died and as a death in that year.

Deaths

In some cases, an exact date of death was not included in the valuation census data files. That a death had occurred was derived from a change in status between valuation dates. In other words, where an individual is alive in one

[7] Technically "initial" exposed to risk, which is designed for use when calculating the probability of a plan participant dying over the next year.

valuation census file, and is deceased in the following valuation census file, then this can be identified as a death during the year.

Non calendar year census files

The majority of contributing plans have a "plan year" (i.e. the 1-year period between any two valuation census files) that is equivalent to calendar year, running from January, 1 to December, 31. However, a small number of contributing plans have a "plan year" that does not align with calendar years. In these cases, care is needed in computing the exposures and deaths for each calendar year.

- **Survivors:** For participants surviving the experience period i.e. to the end of the plan year ending on or after December 31, 2020 they can be assigned an exposed to risk for each calendar year based on that survivorship.
- **Deaths:** The handling of deaths depends on whether a date of death is known (either via the valuation census file or the death audit) or not. Where:
  - A date of death is known the participant has been assigned exposure to the calendar years they were alive in, and the death to the calendar year in which it occurred
  - A date of death is not known – and so the death has been imputed by virtue of the member being alive in one valuation census file and deceased in the next census file – the death has been allocated based upon the plan year in which they died.

    If, for example, a plan year end is June 30, and the June 30, 2013 census file showed the participant alive, but the June 30, 2014 census file showed them as deceased then the death would be assigned to the 2013 *calendar year*. In this example the participant would have exposure of 1 for 2013, and be shown as a 2013 death, and no exposure would be shown for 2014.

    This is a pragmatic approach adopted by Mercer in the MILES dataset. It will mean that a very small proportion of deaths in any calendar year are likely to have actually happened in the following calendar year, and that, in aggregate, the exposure for those deaths will be understated (i.e. exposures are very slightly understated as the exposure for the calendar year of actual death is omitted)

    To gain comfort that this approach is immaterial we analyzed the potential impact as part of publishing our first edition of VitaCurves. Information from Mercer enabled us to identify which plans this may be an issue for, along with an indicator in the dataset which identifies which deaths have been assigned in this way. The small proportion of plans for whom deaths are inputted in this way, coupled with the small proportion of plans with plan year ends differing from the calendar year led to an estimate of only a very small proportion (around 3.5%) of total deaths in 2014-16 were susceptible to having been reported in a different calendar year to that in which they occurred. Our sensitivity testing verified that the issues around timing of death, and potential understatement of exposures had no material impact on the resulting mortality rates from our modeling.

Excluding years with potential recording issues

The exposed to risk and death calculations also controlled for periods where the pension plans may have not consistently recorded lives and deaths – for example years during which the plan carried out partial or full buyout transactions (as deaths will not have been tracked after the transfer to the insurance company), or where there were changes of plan administration. To control for this, the plan experience is excluded for these specific calendar years in Mercer's exposure and death calculations to ensure no bias is created (i.e. the exposures and deaths are set to 0 for these years).

4.3 Data for surviving beneficiaries

There are some inherent challenges in the collection of data relating to the surviving beneficiaries of retirees in pension plans. The process for tracking beneficiaries in pension plan data varies significantly; in particular, where beneficiary data is missing, an assumption may be made about the existence of a surviving beneficiary (or otherwise) until the administrator can make contact with a surviving beneficiary (or otherwise) directly. Consequently, it is an accepted limitation that certain data fields in the MILES data represent estimates for some beneficiaries.

There is also a chance that, in some cases, notional records for beneficiaries may be recorded until it is determined that the beneficiary does not exist. Depending on precise recording practice this has the potential to overstate exposure (false recording of a beneficiary) and to overstate deaths (if cessation of a “false” beneficiary is marked as a death). On balance, we suspect that this may have led to some overstatement of mortality historically within the dataset.

Further, it is more challenging to death audit the data for beneficiaries as they are often tracked under the SSN of the original (deceased) participant, which may lead to some under-reporting of mortality.

However, it appears that the data provided for the most recent years in the MILES 2021 update may be subject to any such limitations to a lesser extent than previous years. We have additionally applied our own data quality controls to screen out records which may be susceptible to false deaths, for example, annuities with guarantee periods, as described in Section 5. Overall, we have confidence in the surviving beneficiary data and subsequent calibrations (when combined with directly sourced data) for CV22 VitaCurves.

However, users should still be aware of the previously described limitations, and so the potential for remaining frailties in the data when relying upon the beneficiary curves.

5 Club Vita quality controls

The directly sourced plan data has been subject to extensive quality controls on receipt from the respective pension plans. In addition, we have sought to apply additional quality controls to the MILES data, provided by Mercer to Club Vita, to replicate, as closely as possible, the additional data checks that have been applied to the directly sourced pension plan data.

5.1 Earliest useable date

We recognize that some plans may not have a complete record of deceased pensioners prior to some point in time. For example, when pensions administration was first computerized it was common practice to periodically 'purge' (i.e. delete) the records of deceased members in order to save on (expensive) disk space. Similarly, where plan administration is moved between platforms, historical deaths may be left behind. If we were to include these years in our analysis, we would not be observing all the deaths. We therefore set for all plans an **earliest useable date** which represents the first point in time from which we are confident we have complete recording of lives and deaths.

Setting the earliest useable data is done for each data set provided annually by each of the directly sourced pension plans who are Club Vita subscribers.

For the MILES dataset this means that we identify for each plan the first calendar year where there are no clear concerns over the completeness of the data. To do this we check for each plan whether there are any years where either the exposure or the deaths "jump" up in a manner that indicates under-reported data in prior years. (We do this excluding the "risk transfer" years described in section 4.2.).

5.2 Latest useable date

With mortality data there is always a risk that some deaths have been incurred but not reported ("IBNR") at the point of reporting. To ensure that mortality rates are not underestimated we also carry out analysis to verify the point up to which we believe we have full and complete death data. This leads to a **latest useable date** (LUD) for each scheme.

Again, setting the latest useable data is done for each data set provided annually by each of the directly sourced pension plans who are Club Vita subscribers.

In the context of the MILES dataset the risk that the valuation census data may be exposed to issues with incurred but not reported deaths ("IBNR") is likely to be higher towards the end of the study period by virtue of the "time lag" that can exist in reporting deaths. As a result, we have performed some high-level checks on each plan to establish whether this is a potential concern and concluded that only a very small number of plans saw a sharp "drop-off" in death counts relative to exposed-to-risk in the final calibration years.

5.3 Quality flags

Where receiving data direct from pension plans, we screen the data 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 either as "good", "suspicious" or "bad". Where the volumes of "suspicious records" are high, these are converted to "bad", otherwise "good". Missing data is marked as "bad".

The data screening for this calibration of VitaCurves has relied on the data processing and cleaning performed by Mercer on the MILES data as described in section 4.1, as well as that we have applied to the directly sourced data. This essentially provides the value of a data field to Club Vita where those checks suggest it is "good", and otherwise the data is returned as missing (which we mark for annuity amount and ZIPZIP+4 as having a “bad” quality flag[8]). We then additionally mark as “bad”:

- **For annuity amount**: Any records with a benefit amount that is:
  - Zero or negative.
  - Suspiciously low or high.
  - Suspiciously round to $100 or $1000.
- **For ZIP+4 code:** Any records which have not been able to be mapped to a ZIP+4 longevity group either because no ZIP+4 was provided, or it is not recognized as a valid ZIP+4 code (for example as overseas or due to a transcription error).
- **For ZIP code:** Any records which have not been able to be mapped to a ZIP longevity group either because no ZIP was provided, or it is not recognized as a valid ZIP code (for example as overseas or due to a transcription error).
- **For collar type:** Any records *not* specifically identified as blue or white collar (i.e. those records for which a broad “propensity” to collar type was instead provided).
- **For pension form:** Any records with missing or invalid i.e. not equal to “joint” or “single”.

We have then performed two additional levels of quality flagging at the plan level:

- If a plan has a large proportion of excluded records (more than 60%) for a specific longevity predictor (e.g. ZIP code) then the whole plan’s data is excluded from the analysis of the impact of this predictor[9].

  The rationale for this is that where the data is held so sparsely it is more liable to be incorrect / not up to date.
- If a plan has a material bias (greater than 20%) between the proportion of records marked “bad” among the living and the deceased records, then the plan is excluded from the analysis of the impact of that predictor on mortality rates.

  This is to avoid distortions in estimated mortality rates owing to either too many deaths are missing data on that longevity predictor (understating mortality rates) or too much exposure is missing for that predictor (overstating mortality rates).

These checks are performed separately in relation to the quality of data for each longevity predictor (annuity amount, ZIP+4 longevity group, ZIP longevity group, collar type and pension form), and separately for annuitants, disabled retirees and (where appropriate) surviving beneficiaries (in each case separately by gender). In order to ensure that biases are not introduced at specific points along the mortality curve, the bias check is not just performed on the entire age range, but also the age ranges 70+ and 75+, where the plan has more than 300 lives in this age range.

[8] For collar group there are no “missing” values as either “blue”, “white” or “unknown” is returned and we model “unknown” as a group in its own right.

[9] This does not apply to collar type as we model the “unknown” group so are not concerned if a relatively small proportion of the plan’s membership is known to be “blue” or “white” collar

As a result of these processes, both the MILES data and the direct sourced data will be assigned quality flags which reflect the quality of each record, on an individual and plan level, for each of the key characteristics.

5.4 Identifying plans with M&A activity

Within the MILES dataset, participants newly entering for the latest updates are identified with an "N" prefix in their unique member identification key, and otherwise are prefixed with "M". For most plans these new "N" entrants would be new retirees, or new surviving beneficiaries.

Some plans may have also grown owing to the sponsor undertaking M&A. In these cases, the new entrants may be from a different plan which may continue to be administered separately and so have different potential data issues. By looking at the volume and age profile of these new participants we have identified one plan which looked likely to have undergone an M&A in the most recent data update and treated them as two separate plans for the purposes of setting the earliest useable date, latest useable data and quality flags described in sections 5.1-5.3.

Specifically, a plan is identified as likely to be an M&A if its proportion of "N" records to "M" pensioner records exceeds 20%, and both:

- The proportion of "N" pensioner records aged 70+ (age nearest) at January 1, 2019 exceeds 25%; and
- The proportion of "N" pensioner records whose benefits commenced before January 1, 2018 exceeds 50%.

Note that there are no such concerns around M&A activity in the directly sourced data. However there is data that has been provided from a number of insurers. In these cases we have been able to identify the underlying pension plans within the insurer's portfolio, and have treated each of these as an individual plan, for the purposes of setting the earliest useable date, latest useable data and quality flags described in sections 5.1-5.3

5.5 Plan suppressions

Individual members within the MILES dataset will have a flag indicating any years of their experience was suppressed. For example, a suppression event may be a risk transfer exercise, and the remaining unsuppressed records are the members which did not participate in that exercise.

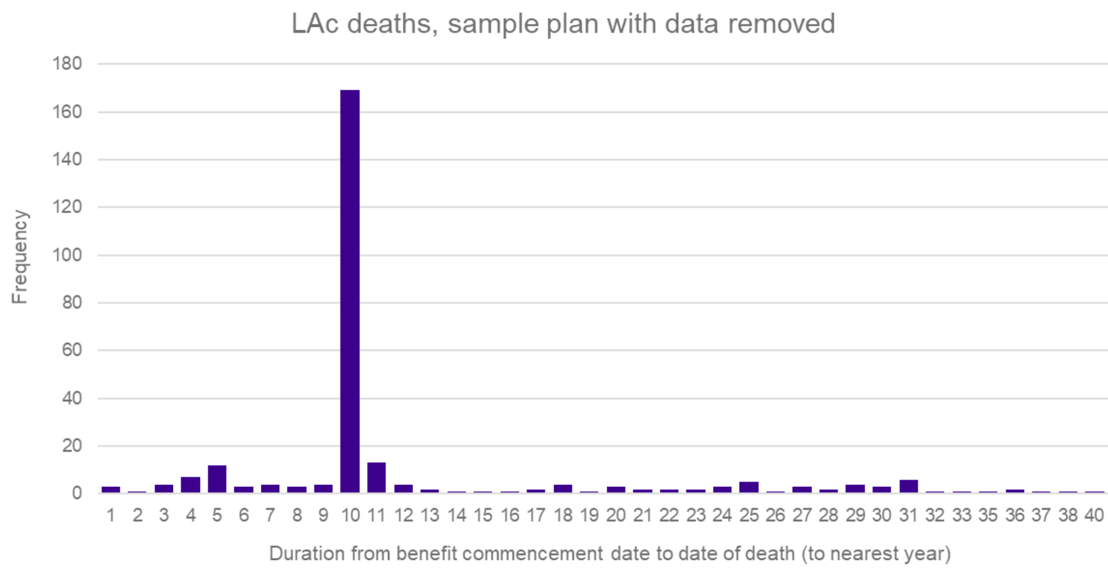
We identified the years in which any suppressed experience existed for each plan and excluded the whole plan-year from the calibration dataset. This control is a precaution in case of any residual under-reporting of deaths and/or unknown selective effects between the members in a plan which were or were not suppressed. Due to the difficulty in assessing the extent of such effects on a plan-by-plan basis, a pragmatic decision was made to employ this approach.

Note that there are no such concerns around plan suppressions for the directly sourced data.

5.6 Life annuities with a guarantee period

We have seen suggestive evidence that records indicated to have a life annuity with a guarantee period ("LAc" members) within the MILES dataset may have been mis-recorded as deaths at the end of their guarantee period. This is particularly the case for beneficiaries. In order to avoid over-stating mortality, we have removed all LAc beneficiaries from the MILES dataset.

We also analyzed the distribution of the time difference between records' date of death and date of benefit commencement for LAc pensioners. If cessations at the end of the guarantee period are mis-recorded on death, we would expect to see an anomalous spike in deaths at common guarantee periods, e.g. 10 years. Combined with a materiality criterion based on the number of underlying records, we used this to identify two plans who are likely subject to this issue and removed all their LAc records from the data. The chart below illustrates this analysis for one of these plans:



For 2018-2020, this exclusion results in the removal of 4,256 (or 0.4%) of the unique records, of which 1,847 are surviving beneficiaries and 1,722 are annuitants.

Note that there are no such concerns around LAc members for the directly sourced data.

6 Data contributing to CV22 VitaCurves

The data volumes contributing to the calibration of CV22 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. Annuitants with unknown retirement health are also excluded from the data used.

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

	Exposed to risk (life years)	Deaths
Annuitant men	598,977	25,081
Annuitant women	354,566	11,536
Disabled retiree men	14,935	633
Disabled retiree women	4,200	122
Surviving beneficiary men	33,351	2,197
Surviving beneficiary women	156,984	9,534
Total	**1,163,012**	**49,103**

7 Want to know more?

If you have any questions on this data document or would like to know additional details regarding our methods for fitting our US VitaCurves, please get in contact with any of the team. We would be delighted to hear from you.



Conor O’Reilly FFA
Head of Analytics
conor.oreilly@clubvita.net

Erik Pickett PhD FIA CERA
Head of Content
erik.pickett@clubvita.net

Dan Chan
Longevity Modeler
dan.chan@clubvita.net

February 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 methodology used for the calibration of the CV22 edition of US VitaCurves. The Research is based upon Club Vita LLP's understanding of legislation and events as of February 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 "Zooming in on ZIP codes", "Calibrating CV22 VitaCurves" and "Adjusting CV22 for COVID-19", this paper complies with the relevant Actuarial Standards Board's Actuarial Standards of Practice (ASOP) and 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 calibration of 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 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.

Annuitant men

	Exposed to risk	Deaths
Initial Data covering calibration age range	598,977	25,081
Annuity amount	591,402 (99%)	24,731 (99%)
ZIP+4	379,299 (63%)	16,360 (65%)
ZIP	451,274 (75%)	19,530 (78%)
Collar Type	464,227 (78%)	19,479 (78%)
Pension Form	592,265 (99%)	24,676 (98%)
Four variables (annuity, ZIP+4, collar type, pension form)	258,770 (43%)	11,245 (45%)

Annuitant women

	Exposed to risk	Deaths
Initial Data covering calibration age range	354,566	11,536
Annuity amount	342,941 (97%)	11,206 (97%)
ZIP+4	199,526 (56%)	6,315 (55%)
ZIP	223,208 (63%)	7,030 (61%)
Collar Type	269,072 (76%)	8,864 (77%)
Pension Form	346,663 (98%)	11,178 (97%)
Four variables (annuity, ZIP+4, collar type, pension form)	120,954 (34%)	3,902 (34%)

Disabled retiree men

	Exposed to risk	Deaths
Initial Data covering calibration age range	14,935	633

Disabled retiree women

	Exposed to risk	Deaths
Initial Data covering calibration age range	4,200	122

Surviving beneficiaries, men

	Exposed to risk	Deaths
Initial Data covering calibration age range	33,351	2,197
Annuity amount	32,306 (97%)	2,142 (98%)
ZIP+4	18,506 (55%)	1,157 (53%)
ZIP	26,262 (79%)	1,726 (79%)
Two variables (annuity, ZIP+4)	17,725 (18%)	1,126 (51%)

Surviving beneficiaries, women

	Exposed to risk	Deaths
Initial Data covering calibration age range	156,984	9,534
Annuity amount	156,082 (99%)	9,486 (99%)
ZIP+4	77,187 (49%)	4,672 (49%)
ZIP	116,895 (74%)	6,948 (73%)
Two variables (annuity, ZIP+4)	76,890 (49%)	4,650 (49%)