# Data underpinning ZIP+4 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 first edition of Club Vita's US VitaCurves is the Mercer Longevity Database ("MILES") dataset. In this paper, we provide an overview of our understanding of the MILES dataset (**Sections 1 to 4**), describe the additional processing we have applied to the data (**Section 5**) and summarize the data volumes underpinning the VitaCurves (**Section 6**). This paper has been shared with Mercer in advance of publication to ensure it represents a fair and accurate representation of the data and information received.

## 1 The heritage of the 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 a number of other plan sponsors) and each plan has consented to the onward sharing of their longevity data with third parties.

No personally identifiable data is included in the dataset. Club Vita worked with Mercer to supply a file containing a lookup between ZIP+4 and ZIP to our preferred geo-demographic factors. Mercer then supplied to Club Vita a copy of the MILES dataset with these factors appended to the data which we have used to calibrate our VitaCurves model. (Appendix A provides a list of the main data fields we have received and relied on in our modeling.)

The data relates purely to in payment annuities, and includes annuitants, disabled retirees and surviving beneficiaries of deceased retirees and was collected in two batches. The first batch of data was collected by Mercer during September 2014 for a study period spanning 2008-2011. An update to the data was performed during 2017 to cover the 2012-2016 period. A few sponsors who provided data during the 2008-2011 period did not provide updates in 2017. Similarly, a small number of plans (5) provided data in the update which had not previously contributed.

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 the vast majority of plans, their year end coincided with calendar year ends so were perfectly aligned with the experience period.

## 2 A rich and diverse dataset

The sections below summarize the data available in the MILES dataset. Sections 2.3 onwards are restricted to the individuals exposed to risk during the period to which we have calibrated our first generation US VitaCurves i.e. 2014 through 2016. (All charts are for the whole dataset prior to the quality controls set out in section 5).

### 2.1 Range of different plans

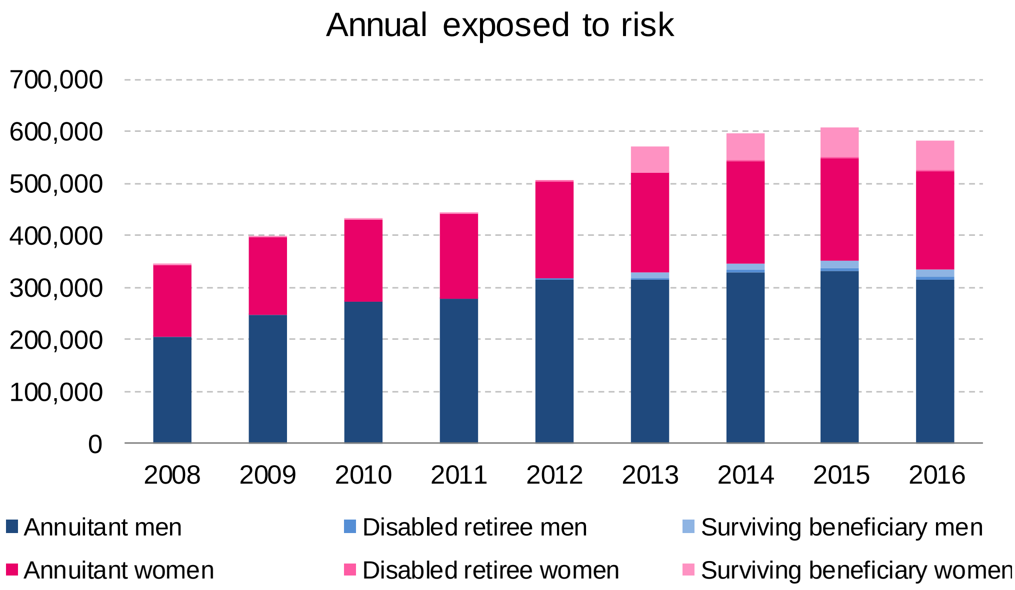
103 different private sector defined benefit pension plans contribute data covering the 2014-2016 period. These plans cover a range of different sizes, from smaller plans with less than 1,000 annuitants and beneficiaries (16 plans) through to very large plans with more than 20,000 annuitants and beneficiaries (10 plans).

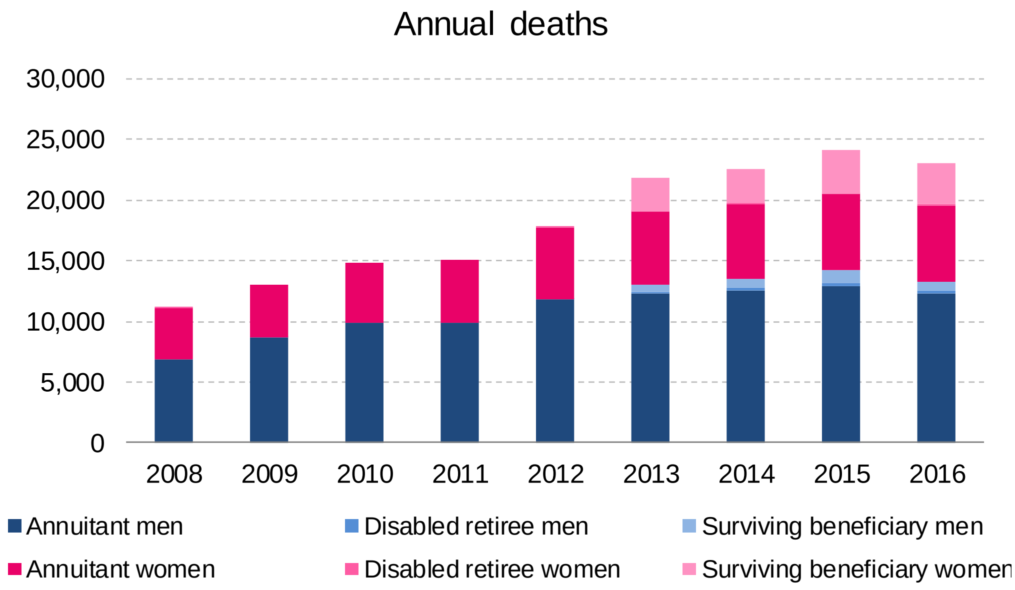
The table to the right shows the distribution of the plans by size.

| Number of annuitants and beneficiaries | Number of plans |
|---|---|
| <1,000 | 16 |
| 1,000 – 4,999 | 55 |
| 5,000 – 9,999 | 11 |
| 10,000 – 19,999 | 11 |
| 20,000+ | 10 |

## 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 calendar years covered by the MILES dataset.




We can see how:

- The data shows a marked increase in 2012 consistent with the addition of extra plans, and again in 2013, as data on surviving beneficiaries was also collected;
- Notwithstanding the step-up in 2012, data volumes generally increase over time (in particular, in terms of death counts) consistent with the maturation of US pension plans; and
- Deaths are elevated in 2015 (compared to growth in exposures relative to 2014) consistent with the relatively harsh flu season in early 2015.

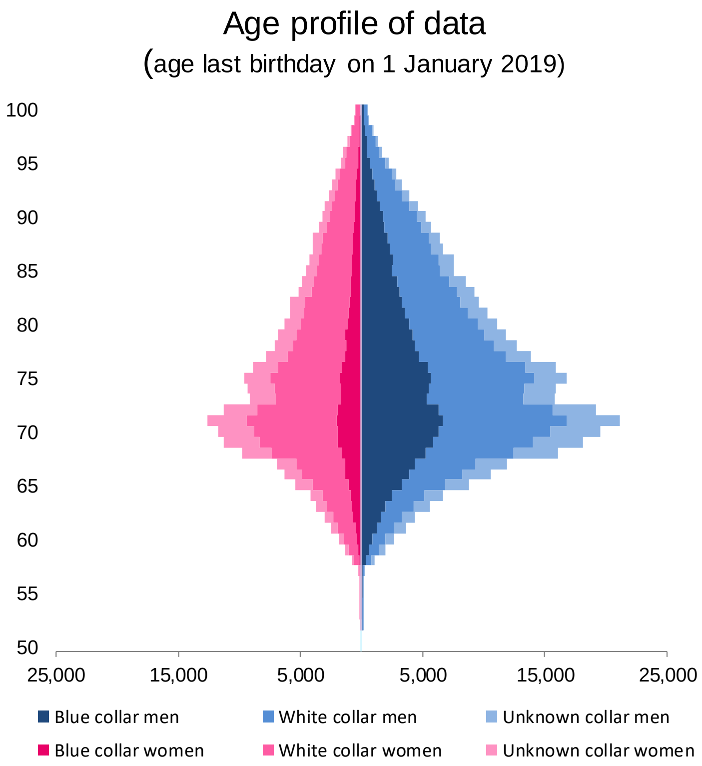
## 2.3 Age profile of data

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 (excluding disabled retirees) based upon age last birthday on 1 January 2019[1].

The annuitant population rises rapidly from age 60 upwards. The step-up at age 68 is consistent with 65 being a popular retirement age and so the retirees covered by this dataset (i.e. those that happened in 2016 or earlier) have reached 68 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.

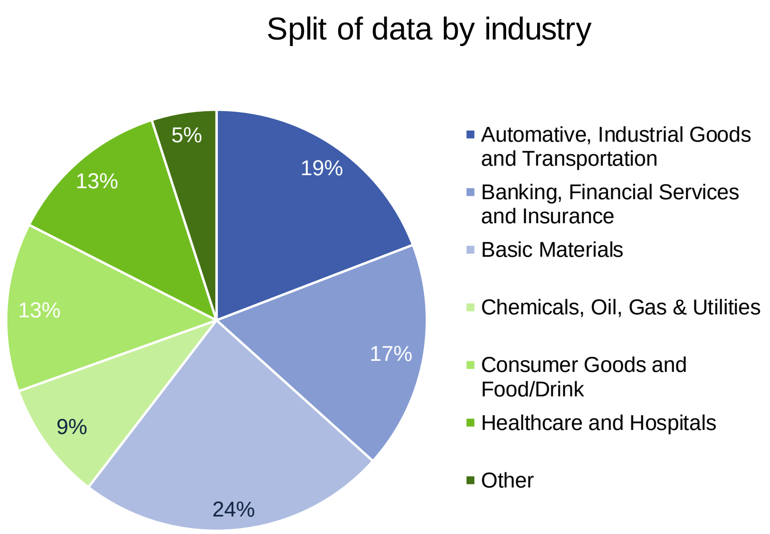



The chart also highlights how:

- consistent with pension plan participation rates, there are more annuitant men than women;
- among men there is a broadly even split between blue collar (34%) and white collar (48%) annuitants; and
- the data for annuitant women is primarily white collar (59%) with a modest proportion of blue collar annuitants (18%)

## 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.



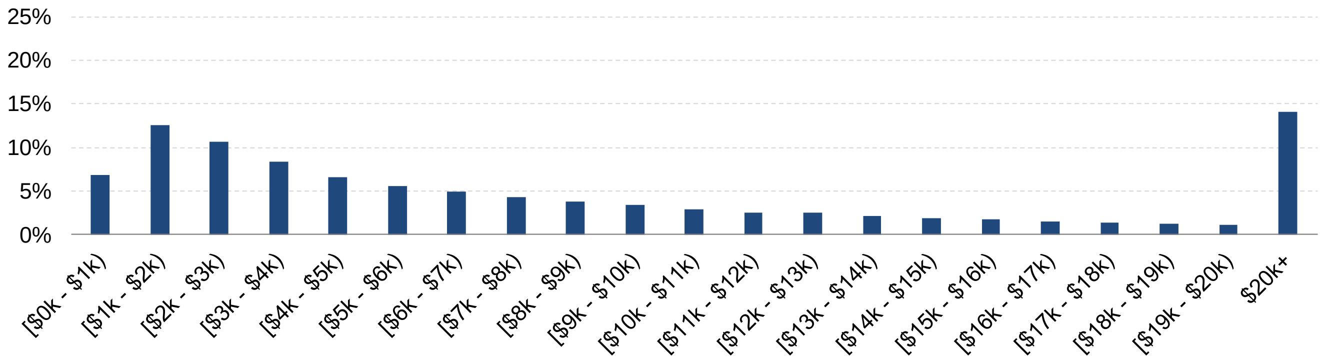

[1] In the case of deceased annuitants we have shown them at the age they would have been on 1 January 2019, had they survived.

## 2.5 Annuity amounts

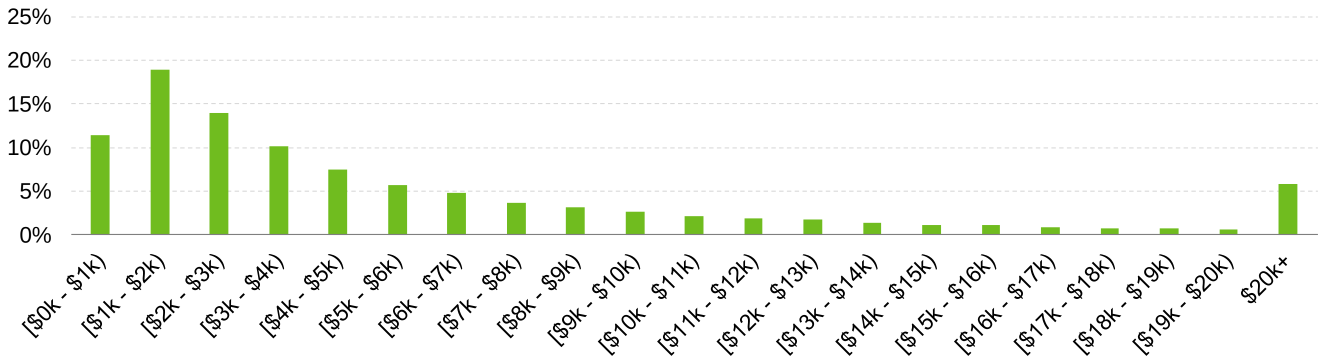
The charts below illustrate the distribution of (annual equivalent) annuity amounts in the data for annuitants, disabled retirees and surviving beneficiaries. In each case this has been split by gender. We see how:

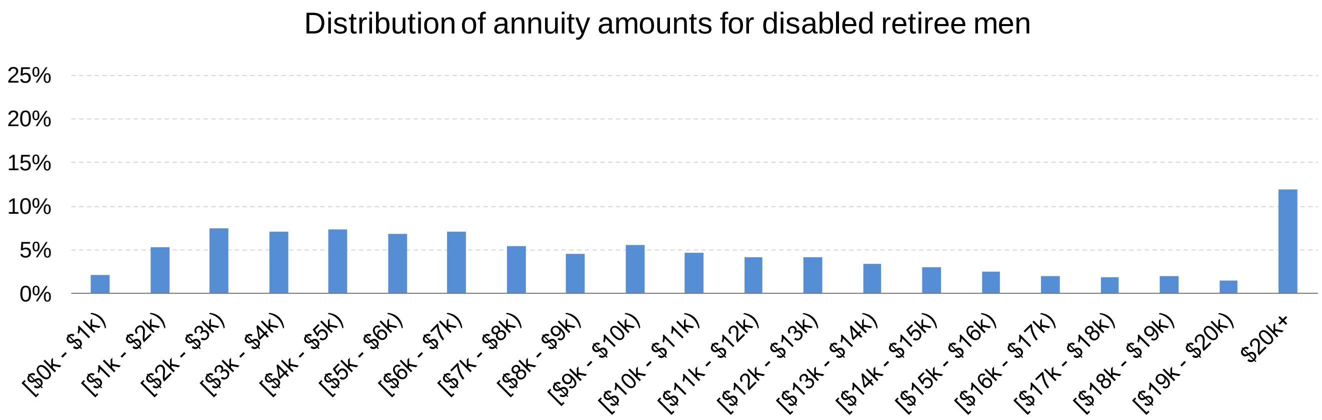
- annuity amounts are skewed to lower amounts reflecting that modest pension incomes can be achieved either by short service or low pay
- there is a significant tail to annuity amounts
- annuity amounts are more evenly distributed for disabled retirees, reflecting the enhanced benefits sometimes payable 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 labour force participation and potentially pay differentials where the benefit is linked to salary

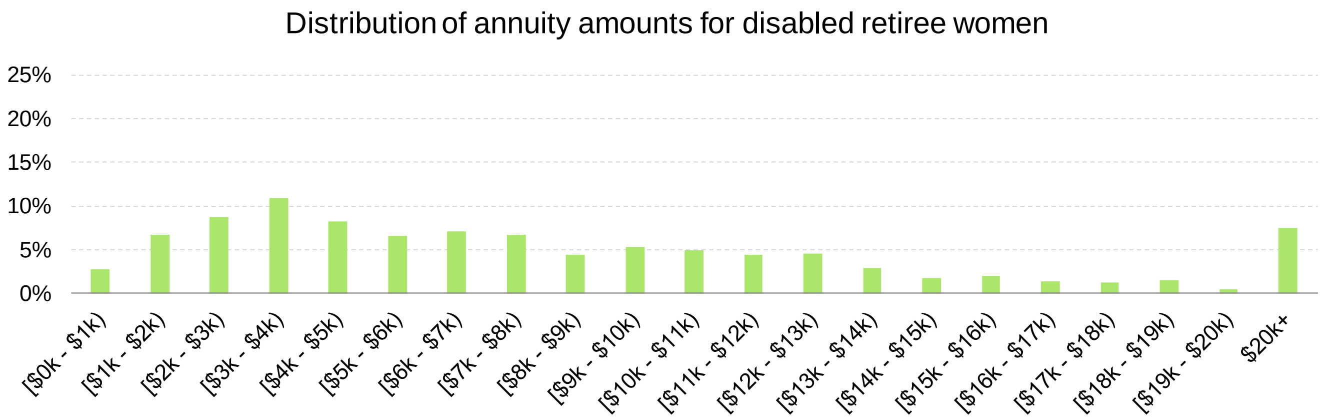
Distribution of annuity amounts for annuitant men

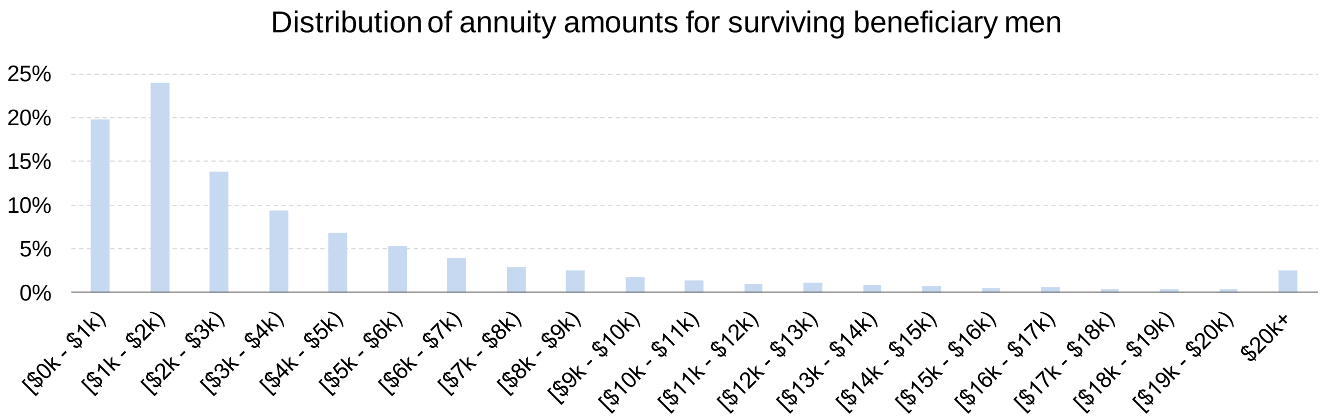


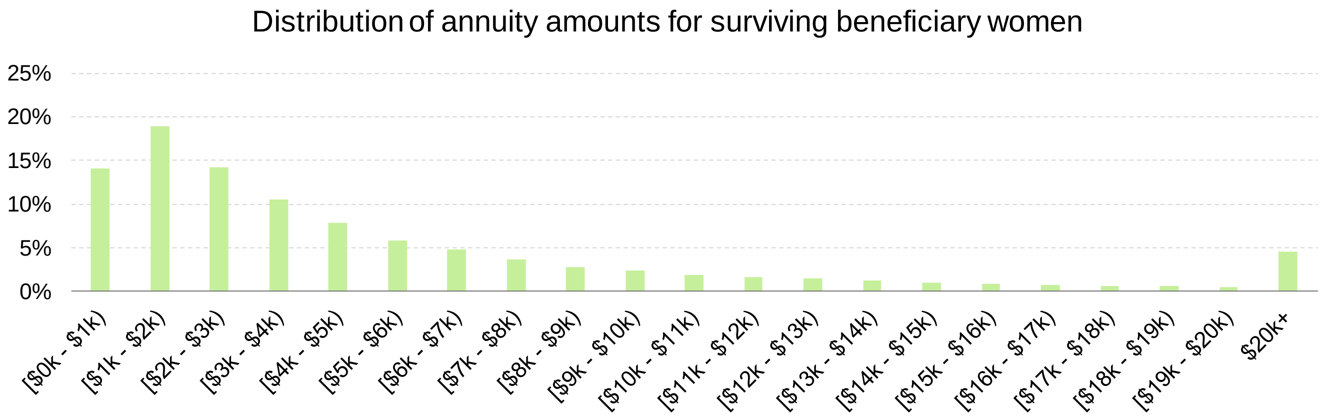

Distribution of annuity amounts for annuitant women




Distribution of annuity amounts for disabled retiree men
25%
20%
15%
10%
5%
0%
[$0k - $1k)
[$1k - $2k)
[$2k - $3k)
[$3k - $4k)
[$4k - $5k)
[$5k - $6k)
[$6k - $7k)
[$7k - $8k)
[$8k - $9k)
[$9k - $10k)
[$10k - $11k)
[$11k - $12k)
[$12k - $13k)
[$13k - $14k)
[$14k - $15k)
[$15k - $16k)
[$16k - $17k)
[$17k - $18k)
[$18k - $19k)
[$19k - $20k)
$20k+


Distribution of annuity amounts for disabled retiree women
25%
20%
15%
10%
5%
0%
[$0k - $1k)
[$1k - $2k)
[$2k - $3k)
[$3k - $4k)
[$4k - $5k)
[$5k - $6k)
[$6k - $7k)
[$7k - $8k)
[$8k - $9k)
[$9k - $10k)
[$10k - $11k)
[$11k - $12k)
[$12k - $13k)
[$13k - $14k)
[$14k - $15k)
[$15k - $16k)
[$16k - $17k)
[$17k - $18k)
[$18k - $19k)
[$19k - $20k)
$20k+

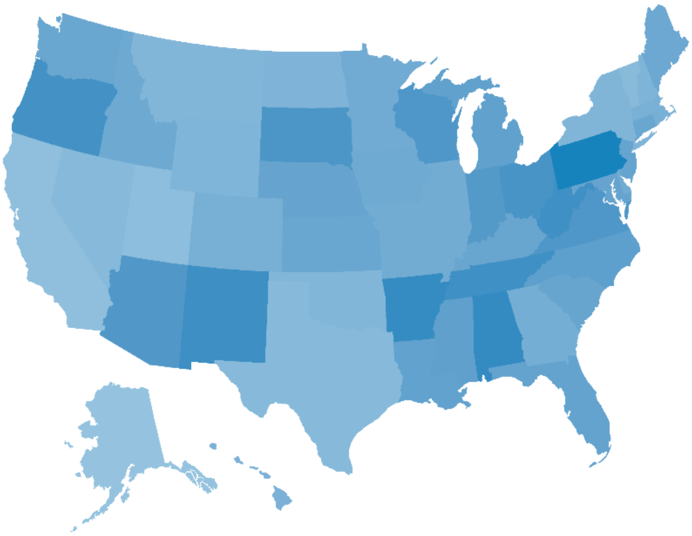

Distribution of annuity amounts for surviving beneficiary men
25%
20%
15%
10%
5%
0%
[$0k - $1k)
[$1k - $2k)
[$2k - $3k)
[$3k - $4k)
[$4k - $5k)
[$5k - $6k)
[$6k - $7k)
[$7k - $8k)
[$8k - $9k)
[$9k - $10k)
[$10k - $11k)
[$11k - $12k)
[$12k - $13k)
[$13k - $14k)
[$14k - $15k)
[$15k - $16k)
[$16k - $17k)
[$17k - $18k)
[$18k - $19k)
[$19k - $20k)
$20k+


Distribution of annuity amounts for surviving beneficiary women
25%
20%
15%
10%
5%
0%
[$0k - $1k)
[$1k - $2k)
[$2k - $3k)
[$3k - $4k)
[$4k - $5k)
[$5k - $6k)
[$6k - $7k)
[$7k - $8k)
[$8k - $9k)
[$9k - $10k)
[$10k - $11k)
[$11k - $12k)
[$12k - $13k)
[$13k - $14k)
[$14k - $15k)
[$15k - $16k)
[$16k - $17k)
[$17k - $18k)
[$18k - $19k)
[$19k - $20k)
$20k+

## 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 states *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 3 digit ZIP supplied in the data; which is available for 80% of the records.



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

The vast majority of the states have between 0.1% and 0.2% of the state population represented in the dataset, indicating a broadly-even geographical mix. Concentrations are highest in Pennsylvania, Alabama and Arkansas. The three lowest concentrations are in District of Columbia (liable to be dominated by public rather than private plan participants), Alaska and California (the most populous state with almost 40 million residents[2]).

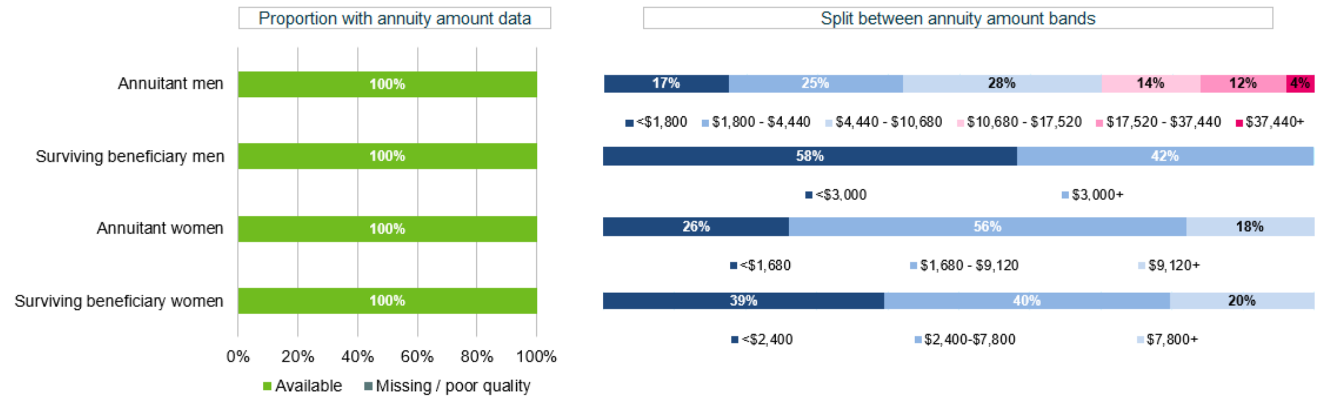
[2] State populations sourced from United States Census Bureau: "Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2018"

# 3 Availability and spread of key longevity predictors

Our analysis identifies the impact on mortality of three key longevity predictors, separately for annuitants and surviving beneficiaries and for men/women[3]. 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, annuity amount and, for annuitants, occupation). We can see from the charts below that this is the case.

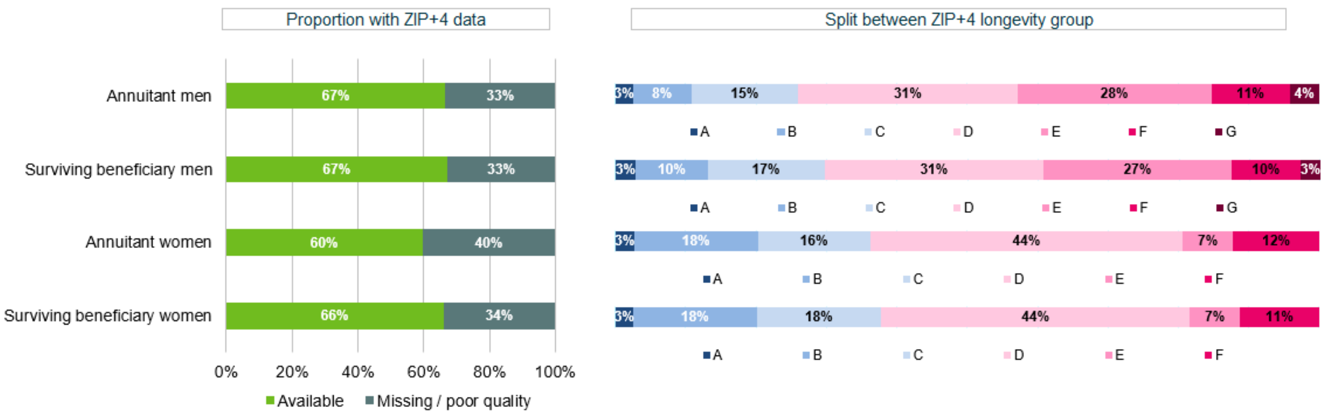
## 3.1 Annuity

Annuity is available for 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 over 70% of the data. We also see how the most "extreme" groups (those with the longest / shortest life expectancies) represent a small proportion of the overall population i.e. the outermost 3-4% of the distribution. This is consistent with our experience in the UK and Canada.




## 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:

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

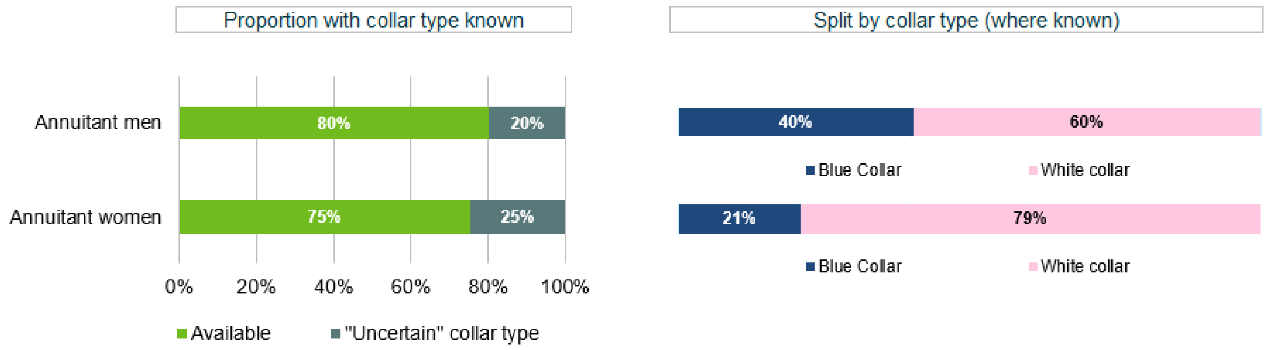
- blue collar if they are *either* hourly-paid or union
- white collar if they are *both* salaried and non-union
- unknown if neither of the above apply

For several plans 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 modeling paper).

# 4 Processing and quality control applied by Mercer

The pension plan data used in our analysis has been collected and processed by Mercer. The data we have received is depersonalise data, as set out in Appendix A. To ensure this data is suitable for the purposes of analysing 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. 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[4]. In total 85% of the dataset has been through this process.

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

- **Age ranges included:** Records were only included in the data where the beneficiary had an age in the range 50-120 in the year of exposure.
- **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 Option ("SSLO") 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 can not 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 VitaCurves, and we have no reason to suspect any bias between lives and deaths within these records.
- **Anonymization:** Dates of birth and death were adjusted to the 15th of the month. Given the broadly uniform distribution of deaths and births over any given month this will not have impacted the modeling. Information based on a participants ZIP(+4) code was appended to the data by Mercer using lookups supplied by Club Vita.

## 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[5] and the "death count" (i.e. an indicator whether the member died) for each individual record in the dataset for each calendar year. We have relied on these pre-calculated exposure/death counts in creating our VitaCurves. These have been calculated according to the following:

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.

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

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 as in one 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, 2016 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 have used information from Mercer to enable 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 mean that only a very small proportion (around 3.5%) of total deaths in 2014-16 are susceptible to having been reported in a different calendar year to that in which they occurred. Our sensitivity testing has verified that the issues around timing of death, and potential understatement of exposures has no material impact on the resulting mortality rates from our modeling.*

**Excluding risk transfer years**

These calculations have also controlled for periods where pension plans carried out partial or full buyout transactions. During the years in which these have taken place only partial reporting of deaths will have been possible (as deaths will not have been tracked after the transfer to the insurance company). To control for this, the plan experience is excluded for these specific calendar years to ensure no bias is created (i.e. the exposures and deaths are set to 0 for these years).

## 4.3 Known data limitations – 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. As a consequence, it is an accepted limitation that certain data fields in the MILES data represent estimates for some beneficiaries.

There is 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 lead to some overstatement of mortality.

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.

Given these potential distortions, we have analyzed the mortality experience among beneficiaries in the MILES data and contrasted this to the mortality experience among plan annuitants. The relative levels of mortality were broadly consistent with our a priori beliefs on the relative mortality of retirees and beneficiaries (based partially on our analysis in the UK over the past decade), and so we have no clear and obvious reason to doubt the overall credibility of the beneficiary data for the purpose of curve calibration.

As a result, we assumed the materiality of the issues described above to be low and have utilised the MILES data to calibrate curves specifically for male and female beneficiaries. However, users should be aware of these limitations when relying upon the beneficiary curves.

# 5 Club Vita additional quality controls

We have sought to add additional quality controls on to the data as provided by Mercer to Club Vita. These quality controls are designed to replicate as closely as possible the additional data checks that would be applied if we were processing the pension plan data directly.

## 5.1 Earliest useable date

We recognise 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 computerised 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.

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.).

For a small number of plans this indicated an EUD might be needed, however in all cases this was prior to 2014 leading to no restrictions on the data contributing the VitaCurves calibration.

## 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.

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 (5) saw a sharp "drop-off" in death counts relative to exposed-to-risk in the final calibration year (2016). As such, the experience in 2016 for each of these plans has been excluded from the calibration data, so as to avoid the risk of distortion from "IBNR" deaths

## 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 as described in section 4.1. 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 ZIP+4 as having a "bad" quality flag[6]). We then additionally mark as "bad":

- **For annuity amount**: Any records with zero benefit amount
- **For ZIP+4 code:** Any records which have not been able to be mapped by Mercer to a ZIP+4 longevity group (either because no ZIP+4 provided, or it is not recognised as a valid ZIP+4 code (for example as overseas or due to a transcription error)
- **For collar type:** Any records *not* specifically identified as blue or white collar in the underlying MILES data (i.e. those records for which a broad "propensity" to collar type was instead provided).

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.

  *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 and collar type), and separately for annuitants, disabled retirees and 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.

# 6 Volumes of data used in our models

The sections below describe the volumes of data contributing to the calibration of VitaCurves. In interpreting these tables please note that:

- The data relates to the 2014 to 2016 calendar years used for calibration; and
- We only restrict the data to "good quality" for any specific longevity predictor where it is used in the model

## 6.1 Annuitant men

| | Exposed to risk (life years) | Deaths |
|---|---|---|
| Initial dataset (2014-2016 exposures) | 976,320 | 37,711 |
| Apply Earliest and Latest Useable dates | 970,137 | 37,651 |
| Limit to the age range fitting over (65-95) | 810,488 | 34,620 |

[6] For collar group there are no "missing" values as either "blue", "white" or "unknown" is returned and we model "unknown" as a group in it's own right.

| | | | |
|---|---|---|---|
| Annuity amount | "Good quality" | 810,485 (100%) | 34,620 (100%) |
| ZIP+4 | Available | 600,844 (74%) | 26,177 (76%) |
| | "Good quality" | 539,927 (67%) | 23,597 (68%) |
| Collar type | "Good quality" | 810,488 (100%) | 34,620 (100%) |
| All three variables (annuity, ZIP+4,collar type) | Available | 600,841 (74%) | 26,177 (76%) |
| | "Good quality" | 539,924 (67%) | 23,597 (68%) |

Note: Annuity amount was available for all individuals, collar type was specified (blue, white, unknown) for all individuals hence we have only shown the impact of quality control for those variables. "Good quality" for collar type includes all records as where the quality flagging identified any biases in the proportions of living and deceased participants with uncertain collar type then all participants in the plan were marked as uncertain collar type. Percentages relate to the proportion of the dataset after applying earliest/latest useable dates and the age ranges restrictions that is available at that stage and shown to nearest 1%.

## 6.2 Annuitant women

| | | Exposed to risk (life years) | Deaths |
|---|---|---|---|
| Initial dataset (2014-2016 exposures) | | 586,324 | 18,649 |
| Apply Earliest and Latest Useable dates | | 580,037 | 18,606 |
| Limit to the age range fitting over (65-95) | | 478,671 | 16,326 |
| Annuity amount | "Good quality" | 478,671 (100%) | 16,326 (100%) |
| ZIP+4 | Available | 319,134 (67%) | 10,864 (58%) |
| | "Good quality" | 284,990 (60%) | 9,685 (59%) |
| Collar type | "Good quality" | 478,671 (100%) | 16,326 (100%) |
| All three variables<br>(annuity,<br>ZIP+4,collar type) | Available | 319,134 (67%) | 10,864 (58%) |
| | "Good quality" | 284,990 (60%) | 9,685 (59%) |

Note: Annuity amount was available for all individuals, collar type was specified (blue, white, unknown) for all individuals hence we have only shown the impact of quality control those variables. "Good quality" for collar type includes all records as where the quality flagging identified any biases in the proportions of living and deceased participants with uncertain collar type then all participants in the plan were marked as uncertain collar type Percentages relate to the proportion of the dataset after applying earliest/latest useable dates and the age ranges restrictions that is available at that stage and shown to nearest 1%.

We can see how while there is a bias of available ZIP+4 data for annuitant women between the in-force (67%) and deaths (58%), the quality control process has removed this bias.

## 6.3 Disabled retiree men

| | Exposed to risk (life years) | Deaths |
|---|---|---|
| Initial dataset (2014-2016 exposures) | 18,086 | 820 |
| Apply Earliest and Latest Useable dates | 18,080 | 819 |
| Limit to the age range fitting over (55-90) | 17,130 | 769 |

## 6.4 Disabled retiree women

| | Exposed to risk (life years) | Deaths |
|---|---|---|
| Initial dataset (2014-2016 exposures) | 5,708 | 170 |
| Apply Earliest and Latest Useable dates | 5,681 | 167 |
| Limit to the age range fitting over (60-80) | 4,359 | 132 |

## 6.5 Surviving beneficiaries, women

| | | Exposed to risk (life years) | Deaths |
|---|---|---|---|
| Initial dataset (2014-2016 exposures) | | 163,522 | 9,819 |
| Apply Earliest and Latest Useable dates | | 163,300 | 9,798 |
| Limit to age range fitting over (60-95) | | 151,702 | 8,903 |
| Annuity amount | "Good quality" | 151,702 (100%) | 8,903 (100%) |
| ZIP+4 | Available | 110,206 (73%) | 6,547 (74%) |
| | "Good quality" | 100,012 (66%) | 5,979 (67%) |
| Both variables (annuity, ZIP+4) | Available | 110,206 (73%) | 6,547 (74%) |
| | "Good quality" | 100,012 (66%) | 5,979 (67%) |

Note: Annuity amount was available for all individuals hence we have only shown the impact of quality control for this variable. Percentages relate to the proportion of the dataset after applying earliest/latest useable dates and the age ranges restrictions that is available at that stage and shown to nearest 1%.

## 6.6 Surviving beneficiaries, men

| | | Exposed to risk (life years) | Deaths |
|---|---|---|---|
| Initial dataset (2014-2016 exposures) | | 38,366 | 2,576 |
| Apply Earliest and Latest Useable dates | | 38,201 | 2,564 |
| Limit to the age range fitting over (65-95) | | 31,765 | 2,166 |
| Annuity amount | "Good quality" | 31,765 (100%) | 2,166 (100%) |
| ZIP+4 | Available | 23,723 (75%) | 1,576 (73%) |
| | "Good quality" | 21,326 (67%) | 1,413 (65%) |
| Both variables (annuity, ZIP+4) | Available | 23,723 (75%) | 1,576 (73%) |
| | "Good quality" | 21,326 (67%) | 1,413 (65%) |

Note: Annuity amount was available for all individuals hence we have only shown the impact of quality control for this variable. Percentages relate to the proportion of the dataset after applying earliest/latest useable dates and the age ranges restrictions that is available at that stage and shown to nearest 1%.

# 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.

**Steven Baxter**
Head of Innovation & Development
Club Vita
steven.baxter@clubvita.net

**Daniel Reddy**
CEO
Club Vita US
daniel.reddy@clubvita.net

**Erik Pickett**
Head of Content
Club Vita
erik.pickett @clubvita.net

**James Maloney**
Head of Longevity Modelling
Club Vita
james.maloney@clubvita.co.uk

October 2019

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 first generation of US VitaCurves. The Research is based upon Club Vita LLP's understanding of legislation and events as of August 2019 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" and "Calibrating VitaCurves," 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: Main data fields

| Type of field | Field | Notes |
|---|---|---|
| Identifier | RecordID | A unique identifier created by Mercer enabling queries relating to specific records to be raised. |
| Type of participant | RETorBEN | Identifies if plan participant (retiree) or a surviving beneficiary of a deceased retiree |
| | Disabled | Indicator if annuitant a disabled retiree |
| Key dates | Date of birth | Provided as 15th of the month/year of birth to protect confidentiality of participants |
| | Date benefit commenced | Date benefit commenced *(note that this can relate to commencement date of most recent type/level of benefit and so can be later than date of benefit first commencing)* |
| | Date of death | Date of death of a deceased retiree/beneficiary. Provided as 15th of the month/year of birth to protect confidentiality of participants |
| Key date information flags | BCDflag | Provides details of whether an actual or an imputed benefit commencement date has been provided. |
| | DODflag | Provides details as to how the date of death was sourced, including if it went through the specific death audit performed on the dataset i.e. consent was provided by the plan sponsor. |
| Type of benefits payable | Form of payment | If annuity is known to be a single life or a joint & survivor benefit this is recorded within this field, along with information on the broad level of continuation percentage and if a guarantee period applies. |
| | Lump sum options | Provides information on lump sum options that may have been available to the plan participant. |
| | Deferred vested cashout | Provides information on whether the plan provided any one-time deferred vested cashout offer during the study period. |
| Potential predictors of longevity | Gender | Gender of participant or surviving beneficiary |
| | Collar *(Numeric)* | Indicator if plan participant blue (1) or white collar (0), or broad percentage of blue collar in the plan. |
| | Benefit | Monthly benefit amount |
| | ZIP3 | First 3 digits of ZIP code |
| | ZIP – CV | ZIP based geo-demographic factors provided by Club Vita and appended by Mercer to data |

| Type of field | Field | Notes |
|---|---|---|
| Potential predictors of longevity | ZIP9 – CV | ZIP+4 based geo-demographic factors provided by Club Vita and appended by Mercer to data |
| | Industry code | 6 digit business unit code of the plan sponsor as per Form 5500 used in ERISA disclosures. |
| | Industry | Industry grouping of the plan. Generally based upon the 6 digit industry code. In a few cases a specific industry was provided by the plan where the 6 digit business unit code did not accurately describe the plan population. |
| Exposed to risk and deaths | Exposed to risk | Initial exposed to risk for each calendar year 2008-2016 |
| | Deaths | Binary indicator as to whether record contributes to death count for each calendar year 2008-2016 |
| | Suppression details | Provides information as to which calendar years exposures / deaths have been suppressed (set to 0) to avoid distortions in mortality reporting, for example in years of an annuity buyout transaction where the in force at the start of the year could not be tracked in full to end of year. |