Article from

In The Public Interest

January 2020

Zooming in on ZIP Codes: Using Socio-Economic Factors to Tailor U.S. Pension Plan Longevity Assumptions

By Daniel Reddy and Erik Pickett

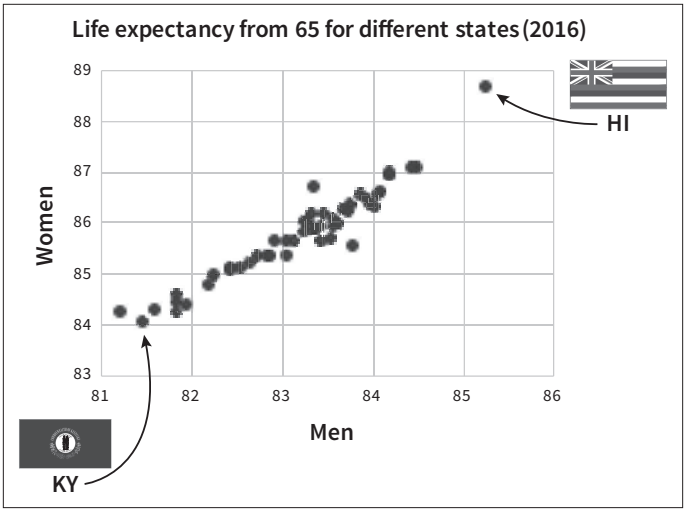
A DIVERSE NATION

The United States is a diverse nation made up of many people with distinctly different characteristics. This diversity is particularly noticeable when you analyze life expectancy.

Figure 1 shows the average life expectancy for men and women for each state in the U.S. Each state itself is made up of a diverse mix of people, but even so, the state average life expectancies are very different from state to state, with more than four years of difference in life expectancy from Kentucky to Hawaii.

What is contributing to this diversity, and how can pension plans account for it when setting their longevity assumptions?

Figure 1
State Variations in Life Expectancy



Source: Club Vita analysis of Barbieri, Magali, and John, Wilmoth. 2019. United States Mortality Table. March 4, *http://usa.mortality.org* (accessed Nov. 12, 2019).

Nurture, not Nature

Many believe that longevity is passed down through the genes we inherit from our parents, but research suggests that only about 20 percent of the differences in life expectancy comes from our genes.[1] The majority is driven by external factors such as lifestyle and environment.

Some key characteristics that indicate how long individuals will live include their level of education, whether they smoke, how much exercise they get, the type of job they have, how wealthy they are and even how much sleep they get. Many of these factors are not possible for pension plans to measure; however, Club Vita can use the data fields that pension plans do hold to create effective proxies.

How can Pension Plans Capture This Diversity?

As displayed in Table 1, the following drivers of longevity can be captured by data fields routinely held by pension plan administrators.

Table 1
Longevity Drivers and Pension Administration Proxies

	Longevity Driver	Data Item Used as a Proxy
	Lifestyle (level of education, propensity to smoke, etc.)	ZIP code
	Affluence	Ideally salary, otherwise pension amount
	Retirement health	Disabled or normal health retirement
	Occupation	Blue- or white-collar worker

Categorizing participants using these different data fields (often referred to as "rating factors") gives us a granular method for understanding a social security system's or pension plan's demographics. By comparing each participant to the experience of other participants in the Club Vita data set with similar characteristics, we can then derive a longevity assumption appropriate for that participant within the social security system or pension plan.

Zooming In

We've seen that longevity varies state to state. This is largely driven by the different lifestyles of people living in different places. But can we zoom in further and capture more diversity using details about where people live?

Marketers have long appreciated that analyzing ZIP codes helps them spend their budgets more wisely. Pension plan sponsors can repurpose these techniques to refine their understanding of the longevity of their members.

CAPTURING LIFESTYLE EFFECTS USING ZIP CODE

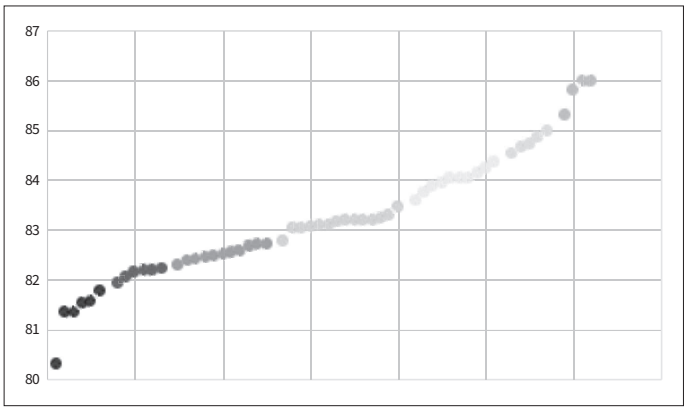
Where individuals live can tell us a lot about their lifestyle and, therefore, about how long they are expected to live. This information is encoded within the 9-digit ZIP code (commonly known as the ZIP+4 code). We prefer ZIP+4 code because there are some very large (100,000-plus resident) 5-digit ZIP codes in the U.S., limiting our ability to identify lifestyle effects. So, how do we get to capturing lifestyle effects based on ZIP+4?

First, we repurpose some key principles that marketers use (see Figure 2).

- **Marketing principle 1:** People living in the same neighborhood have similar characteristics.
- **Marketing principle 2:** Neighborhoods can be characterized by the types of people living there.
- **Marketing principle 3:** Neighborhoods with the same characteristics appear all over the country.

Figure 2
Marketing Categorizations

Figure 3
Life Expectancy by Marketing Group



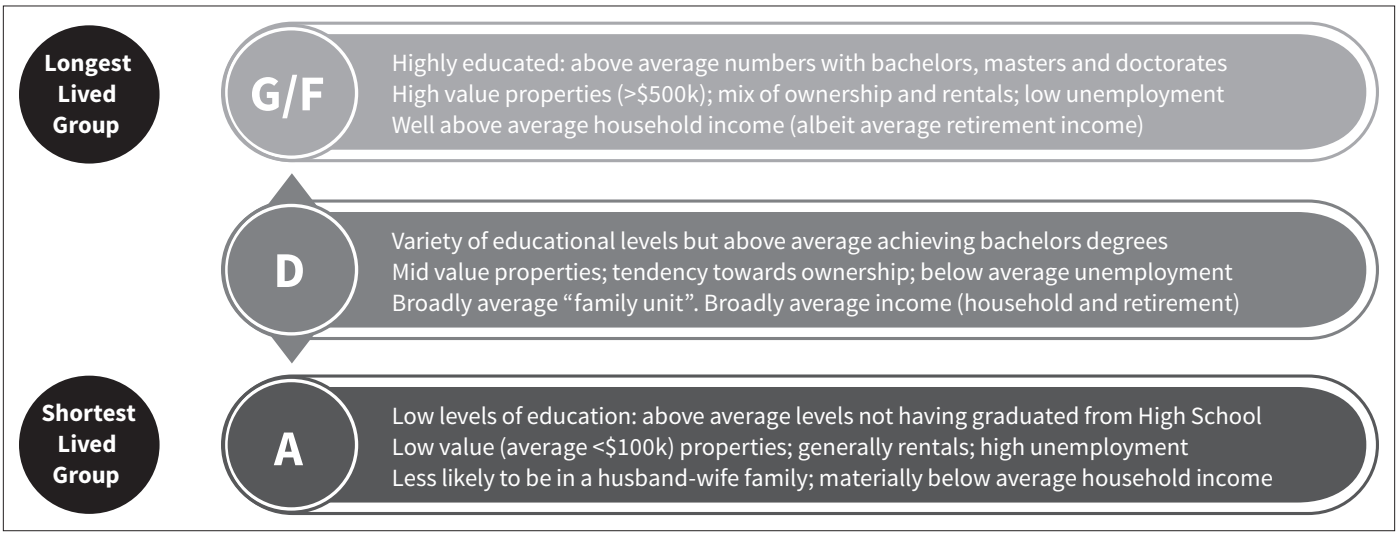
While there are more than 46 million ZIP+4 codes in the U.S., this marketing classification puts each ZIP+4 code into one of 58 different types of neighborhoods.

- **Longevity modeling principle:** Neighborhoods with similar characteristics have similar longevity.

We analyze longevity experience data for people living in each of the different marketing groups and order them from shortest life expectancy to longest life expectancy. We then use a clustering algorithm to simplify the classification of ZIP+4 codes by combining the marketing groups that have similar longevity experience.

This process gives us seven groups exhibiting distinct longevity experience for men (see Figure 3) and six for women. We color-code these "longevity groups" from light to dark, as shown in Figure 4.

Figure 4
Description of Longevity Groups



Source: Club Vita summary of longevity group characteristics

INTRODUCING VITACURVES

ZIP+4 codes enable us to capture large differences in life expectancy, but there are other factors, most notably income, that also lead to considerable variation of life expectancy. Club Vita's approach is to combine the effects of multiple factors including ZIP+4, pension amount, and blue- and white-collar worker into a highly predictive model of current (or "baseline") longevity. We call this the VitaCurves model. The techniques we are describing here have been tried and tested in the U.K.[2] and Canada.[3]

The starting point is the data set underlying our calculations (see Figure 5).

How do we Build the VitaCurves Model?

We split each data field into distinct buckets. Each individual retiree in our data set will be characterized by how the retiree's data fit into each bucket, as shown in Table 2.

We apply a statistical technique called "Generalized Linear Modeling" to our data set to build up a picture of how each data field influences an individual's longevity. We use this technique to calculate an individual longevity assumption, or VitaCurve, for each combination of our data fields. For our first-generation model, we generate 306 VitaCurves (see Figure 6, pg. 11).

The first generation of VitaCurves captures a difference in life expectancy from 65 of 8.7 years for men and 6.6 years for women. Here we show how the different ratings factors contribute to this diversity.

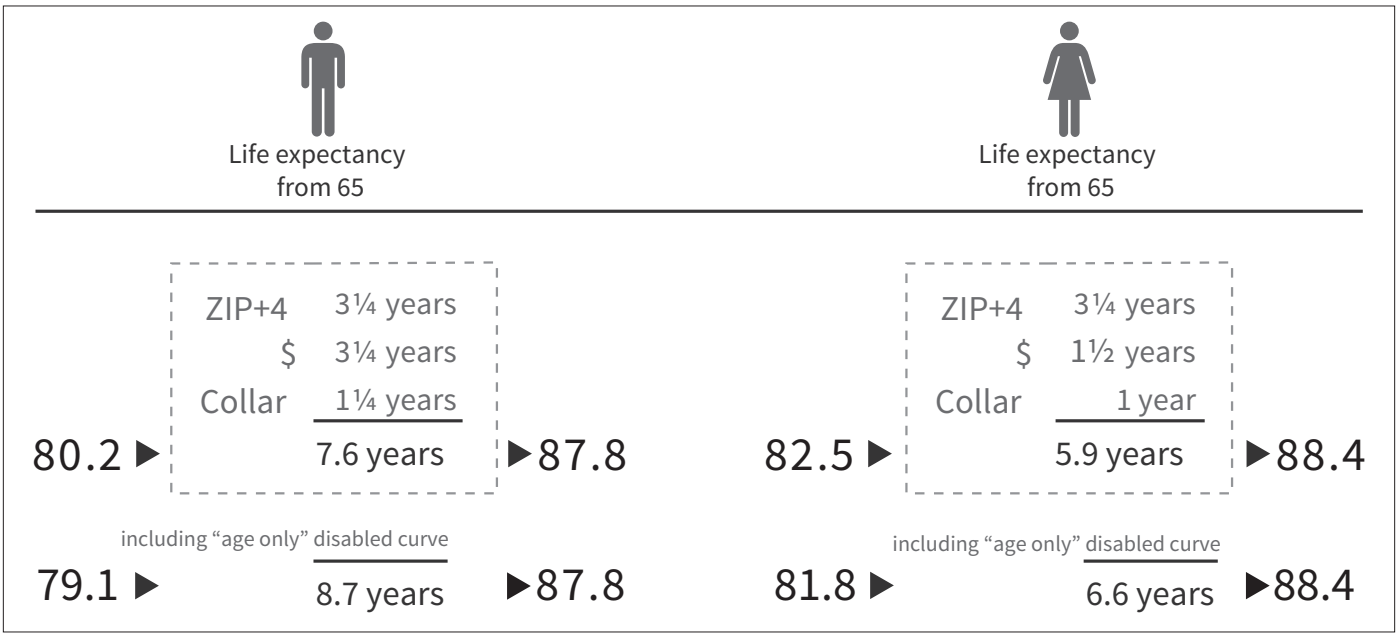
Figure 5
Data Set Characteristics

- The size of our data set is key to make our calculations statistically significant.
- The more data we have, the more we can identify the signal through the noise.
- Our first-generation U.S. VitaCurves model is built on a data set of more than 800,000 retirees from a diverse portfolio of 108 large plans.
- The richness of our data set is key to capture the full diversity between different retirees.
- The more data fields we collect, the more diversity we can capture between retirees.
- Our first-generation U.S. VitaCurves model uses the data fields: ZIP+4, pension amount, blue- and white-collar, first and second life, gender, disabled and normal health.

Table 2
Rating Factors Used in U.S. VitaCurves Model

Data Field	Retiree Men	Retiree Women
Pension amount	6 pension bands	3 pension bands
ZIP+4	7 longevity groups	6 longevity groups
Collar	Blue and white collar	Blue and white collar

Figure 6
Life Expectancy Range and Attribution to Ratings Factors of U.S. VitaCurves



Source: Club Vita US VitaCurves analytics

SOCIAL INSURANCE AND PUBLIC FINANCE

There are many implications of acknowledging and using this information in the U.S. pension system, especially when defined broadly to include all retirement plan types, Social Security and Medicare.

1. Tailored mortality assumptions lead to more accurate cost calculations and stability of liabilities when measured one year to the next.

2. Better cash flow and headcount projections lead to better understanding of both pension plans' and social security systems' future costs and sustainability. The level of plan and system maturity is better understood as well—knowing how many active employees or taxpayers and their level of income support current and future retirees is important.

3. Knowing the real difference in life expectancies for different groups helps to quantify the degree of the intra- and inter-generational equity existing in the social security system.

4. The shift in defined benefit to defined contribution type retirement benefits has been moving risk from plan sponsors to plan participants. That includes not only investment risk but longevity risk as well. Individual defined contribution savers need to understand their individual longevity risk better. Incorporating this Club Vita's analysis into tools calculating probabilities of living to certain ages (such as the Actuaries Longevity Illustrator[4]) will help individuals tailor savings targets.

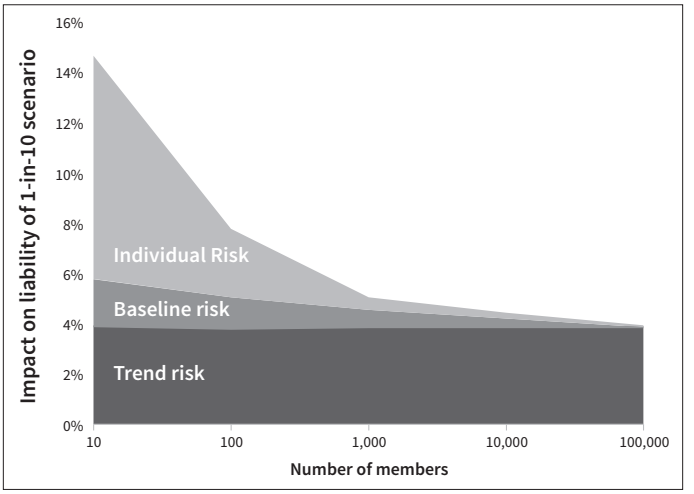
5. Longevity risk is better-understood by the actuarial community in the U.K. but is not yet part of most U.S. risk conversations. Extreme longevity has been rated among the top 15 extreme risks by the Thinking Ahead Institute in 2019.[5] Longevity risk can be broken down into subcomponents:

   a. **Individual (or idiosyncratic) risk:** This is the risk that certain members of a population live significantly longer (or shorter) lives than that predicted, driven by the natural variation in a population. The law of large numbers in a social insurance or any large pension plan means this risk is mainly mitigated by pooling the exposure of a large number of participants.

   b. **Baseline risk:** This is the risk that the exposed population differs from the population used to calculate the current longevity assumption. This could be better managed by sophisticated social insurance and pension plan sponsors using accurate assumptions that are highly tailored to their plan's population.

c. **Long-term trend risk:** This is the risk that comes from long-term mortality improvements being greater (or lower) than expected. Social insurance plans are most exposed to this risk. In 2020, Club Vita aims to understanding how the long-term trends in the U.S. vary by socio-economic factors, through the power of ZIP+4 code (see Figure 6).

Figure 6
Components of Longevity Risk



Source: Club Vita analytics of longevity risk components.

DETAILS ON THE MODELING METHODOLOGY

Further technical details on the underlying data and modeling methodology can be found in the supporting documents at *https://www.clubvita.us/zooming-in-on-zip-codes*. ■

Daniel Reddy, FSA, EA, is the CEO of Club Vita US, LLC. He can be contacted at *daniel.reddy@clubvita.net*.

Erik Pickett, CERA, FIA, Ph.D., is the chief product officer of Club Vita. He can be contacted at *erik.pickett@clubvita.net*.

ENDNOTES

1 See for example Gavrilova, Natalia S., Leonid A., Gavrilov, Galina N., Evdokushkina, Victoria G., Semyonova, Anna L., Gavrilova, et al. 1998. Evolution, Mutations, and Human Longevity: European Royal and Noble Families. *Human Biology* 70, no. 4:799–804. *https://www.ncbi.nlm.nih.gov/pubmed/9686488* (accessed Nov. 12, 2019) or Skytthe, Axel, Kirsten Ohm, Kyvik, Niels Vilstrup, Holm, and James W., Vaupel. 2002. The Danish Twin Registry: 127 Birth Cohorts of Twins. *Twin Research* 5, no. 5:352–357. *https://www.ncbi.nlm.nih.gov/pubmed/12537858* (accessed Nov. 12, 2019).

2 Madrigal, A.M., F.E., Matthews, D.D., Patel, A.T., Gaches, et al. 2011. What Longevity Predictors Should Be Allowed for When Valuing Pension Scheme Liabilities? *British Actuarial Journal* 16, no. 1:1–62. *https://www.cambridge.org/core/journals/british-actuarial-journal/article/what-longevity-predictors-should-be-allowed-for-when-valuing-pension-scheme-liabilities/3C7E032BF549D497D3ABEE506CFF67EF* (accessed Nov. 12, 2019).

3 Ahmadi, Seyed Saeed, and Richard, Brown. 2018. Canadian Institute of Actuaries, May, *https://www.cia-ica.ca/docs/default-source/2018/218068.pdf* (accessed Nov. 12, 2019).

4 Actuaries Longevity Illustrator. Welcome to the Actuaries Longevity Illustrator. *https://www.longevityillustrator.org/* (accessed Nov. 12, 2019).

5 Thinking Ahead Institute. Extreme Risks 2019. *https://www.thinkingaheadinstitute.org/-/media/TAI/Pdf/Research-Ideas/a_public/Extreme_Risk_2019.pdf* (accessed Nov. 12, 2019).