MILLIMAN WHITE PAPER

# On the use of consumer data in predicting mortality risk

Daniel Glab
Michael Niemerg

## Executive summary

Understanding and predicting mortality risk is a critical task in life insurance, with potential implications in market segmentation, underwriting, pricing, and in-force management. This paper reports findings from our study on the use of consumer data in predicting mortality risk.

In this white paper, we:

- Define consumer data and provide examples of its application within the insurance industry
- Provide background on the types and usage of data in a variety of life insurance applications
- Present the results of a study we conducted to understand the impact of incorporating consumer data into models formulated to predict mortality risk
- Explain the methodology employed in this study as well as the rationale supporting key decisions that guided the analysis

Ultimately, our study concludes that the addition of consumer data is strongly predictive of mortality when used in isolation. When combined with other data types for predicting mortality risk (e.g., medical claims, credit attributes, etc.) it yields incremental predictive value.

## Consumer data defined

“Consumer data” captures various characteristics of consumers’ interests, opinions, behaviors, or culture. Consumer data is also sometimes referred to as “consumer lifestyle” or simply “lifestyle” data. Because multiple data vendors offer some version of consumer data (each distinct in its own way), there is no single, precise definition in terms of scope or content. However, such data generally includes:

- Lifestyle preferences (media consumption, habits, interests, hobbies, etc.)
- Indicators of purchase history (house, car, etc.)
- Personal financial attributes (income, mortgage size, payment history, etc.)
- Demographic details (age, gender, marital status, dependents, ZIP Code, etc.)

This data has historically been used to gain insights into an individual's future consumption patterns and social behavior.

Consumer data is obtained through a variety of means, including national and local surveys, browser cookie and social media tracking, and transaction (e.g., credit card) histories. This is often supplemented with census data. These various approaches to data acquisition can be utilized by any individual consumer data vendor, which, in turn, means that the level of granularity can vary from field to field in a given dataset; some fields may be measured at an individual level of granularity, while other fields can be either at a household or ZIP Code level.

Because of the broad scope of consumer data, some consumer data fields resemble other well-known data types, such as “social determinants of health,” “credit,” and “health” data. Despite apparent commonalities, one should not conflate consumer data with these other data types, because consumer data has a distinct focus on consumer preference and is captured at a lower level of granularity. For instance, a credit profile may have hundreds of attributes related to lines of credit, account balances, etc., while consumer data may only capture a few high-level credit-related variables such as outstanding mortgage balance.

Consumer data has been used in the marketing of many consumer goods and services, including insurance. More recently, actuaries and data scientists have explored uses of consumer data that go beyond insurance marketing. Such uses include targeting outreach for wellness and intervention programs that aim to help long-term care (LTC) policyholders,[1,2] and predicting inpatient and emergency department utilization within managed care.[3] The results have encouraged investigation into other possible uses of consumer data and prompted our research.

# Uses of data for mortality prediction

Several data types and sources have proven valuable for predicting mortality in various life insurance use cases. Some noteworthy examples include prescription histories, medical claims, credit data, MIB (formerly known as Medical Information Bureau), and motor vehicle records (MVR). These sources are quite distinct from each other in terms of elements captured, and each yields insight into an individual's mortality risk.

Prescription and medical data can be used to infer underlying medical conditions, which, in turn, generally have a clear causal linkage to mortality. The utility of this data to life insurance carriers is self-evident. In contrast, credit data was originally intended to determine an individual's creditworthiness. In general, credit data captures information such as credit utilization, timeliness of repayment, and credit-seeking behavior. None of this information has a direct connection to the individual's mortality risk. Nevertheless, credit data has been demonstrated to be predictive of mortality.

A primary means by which such data are put into practical use is through the implementation of machine learning algorithms and predictive models. Beyond underwriting, which is foundational to life insurance, there are also well-established applications of predictive models in marketing (i.e., the acquisition and retention of customers) as well as in managing in-force policyholders.[4] These models use various combinations of the data types described above as inputs to furnish a prediction of some outcome of interest, such as mortality or policy lapse.

Consumer data (as previously described) is comprised of information that cannot be uniquely linked to an individual, so it is subject to inaccuracy and precludes its use in any business activity that could result in taking an adverse action against an individual according to the Fair Credit Reporting Act (FCRA). Consequently, consumer data is not suitable for underwriting or pricing. However, given the nature of life insurance, prediction of mortality is always of interest. This is even true of marketing use cases where life insurance carriers seek to initiate outreach to prospective customers who are more likely to qualify for life insurance products.

# Study

### OVERVIEW

To explore the potential benefit that can be achieved from using consumer data to predict mortality outcomes, we examined predictive performance measures for multiple alternative predictive models. The models studied differed in terms of the composition of input data used to train each model. As we will describe, each model was trained on input data comprised of different combinations of data commonly used in the life insurance industry for predicting mortality risk. Models with and without consumer data serve as benchmarks for assessing predictive performance of consumer data.

### DATASET DESCRIPTION

The data for our analysis spans the years 2016 to 2020 and consists of approximately 21 million lives with more than 342,000 observed deaths. In addition to age and gender, we have features derived from prescription, medical, and credit data to use as input into various models.

---

[1] Gaweda, B., Krischanitz, C., Bellina, R. et al. (April 2022). Potential Data Sources for Life Insurance AI Modelling. Milliman Report. Retrieved December 11, 2023, from https://www.milliman.com/-/media/milliman/pdfs/2022-articles/4-22-22_lsc-data-science-report-ai-in-life-insurance.ashx.

[2] Milliman LARA™. Long-Term Care Wellness Initiatives: A Simulated Pilot Program. Retrieved December 11, 2023, from https://www.milliman.com/-/media/products/lara/12-10-21-milliman-lara_simulated-pilot-case-study.ashx.

[3] Chen, S. & Bergman, D. (January 15, 2020). Using Applied Machine Learning to Predict Healthcare Utilization Based on Socioeconomic Determinants of Care. American Journal of Managed Care. Retrieved December 11, 2023, from https://doi.org/10.37765/ajmc.2020.42142.

[4] Batty, M. et al. (April 2010). Predictive Modeling for Life Insurance. Deloitte Consulting LLP. Retrieved December 11, 2023, from https://www.soa.org/globalassets/assets/files/research/projects/research-pred-mod-life-batty.pdf.

In the remainder of this section, we review basic descriptive summaries that provide context as to the composition of the data used for our study, in order to provide readers with some insight into general patterns and to identify potential limitations in our study dataset.

### Baseline mortality and hit rates by data source

For each data source considered in our study, we define a "hit" as the case when an individual from our study has at least one non-null feature corresponding to that data source. Conversely, when no data can be found from a given data source for a given individual, we refer to that individual as a "no hit." The hit rate of a data source is an important consideration when evaluating predictive performance achieved from using that data source. Indeed, strong predictive performance for a data source with a limited hit rate undermines any real-world benefit that can be attained from implementing a model utilizing that data.

From Figure 1, we see that the cohort defined by consumer data has the lowest hit rate of any data type in our dataset, but it also has the lowest mortality. Our calculations of relative mortality utilize the 2015 Centers for Disease Control and Prevention (CDC) Mortality Table for expected deaths. It is interesting to note that, while some credit data attributes also appear as consumer data, the hit rate for credit data is significantly higher than any other type.

**FIGURE 1: HIT RATES AND RELATIVE MORTALITY BY DATA SOURCE**

| | % OF LIVES | RELATIVE MORTALITY |
|---|---|---|
| **Prescription hit** | 77.7% | 95.9% |
| **No hit** | 22.3% | 112.3% |
| **Medical hit** | 76.9% | 99.5% |
| **No hit** | 23.1% | 102.2% |
| **Credit hit** | 94.4% | 96.3% |
| **No hit** | 5.6% | 163.1% |
| **Consumer hit** | 72.1% | 92.8% |
| **No hit** | 27.9% | 125.6% |

### Data coverage by demographic group

Starting with gender, we see in Figure 2 that our study population is not too far off a roughly even gender distribution. The consumer data hit rate is higher for males than for females, but the disparity is not particularly extreme.

**FIGURE 2: HIT RATES BY GENDER**

| GENDER | % OF LIVES | LS HIT |
|---|---|---|
| **Female** | 51.60% | 70.80% |
| **Male** | 48.40% | 73.40% |

In reviewing age, note that consumer data is only available for individuals 18 and older. Additionally, from Figure 3, one can see that our study population is older than the overall U.S. population. Because our study population is primarily comprised of life insurance applicants, this overrepresentation of older ages is not surprising.

**FIGURE 3: HIT RATES BY AGE GROUP**

| AGE GROUP | % OF LIVES | LS HIT |
|---|---|---|
| **18 – 29** | 12.80% | 52.60% |
| **30 – 39** | 20.50% | 68.00% |
| **40 – 49** | 18.20% | 71.90% |
| **50 – 59** | 20.30% | 75.50% |
| **60 – 69** | 18.10% | 80.00% |
| **70+** | 10.10% | 84.10% |

With regard to consumer data, there is a distinct disparity in hit rates across age groups. This disparity can be explained by noting that consumer data captures attributes reflecting various life events that would seem to accumulate with age: home purchase, number of dependents, etc. This relationship between consumer data and age reveals the potential for confounding their associated mortality risks if care is not taken when using the data.

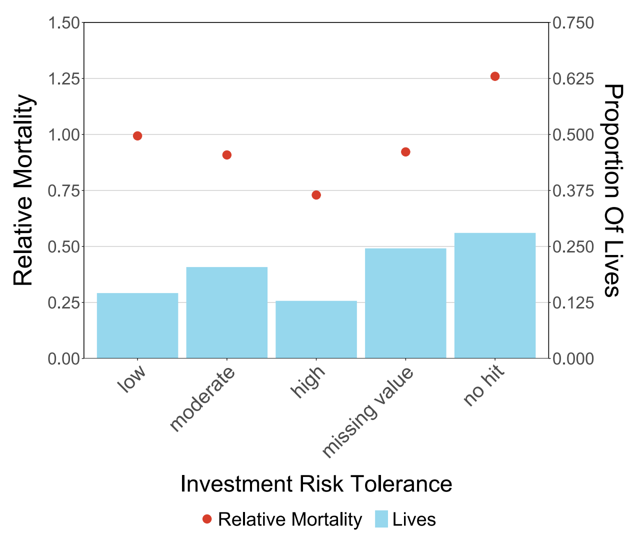
### Overview of consumer features

To get a sense of how mortality risk can be explained by consumer features, we review a representative example of the relationship between mortality and the variable "Investment Risk Tolerance" (IRT). IRT is a prediction of an individual's risk tolerance for investments ranked from "high" (aggressive), to "moderate" (moderately aggressive), to "low" (conservative). It should be noted the prediction is generated by the vendor that supplied our data; we do not have direct insight into the methodology used to produce it. Additionally, "no hit" indicates that there is not consumer data for this individual, while "missing value" indicates that the individual is a consumer hit but does not have a valid value for this variable.

The variation in mortality risk (measured on a relative mortality scale) over the different possible values of IRT is seen in the vertical position of the plotted red points in comparison to the left axis. The relative frequency of each value can be determined by looking at the height of distribution bars in comparison to the right axis.

It would appear that higher IRT tends to coincide with lower mortality risk. Moreover, individuals in the no hit category would tend to have the highest mortality risk. This general relationship in mortality risk across IRT categories roughly holds across age and gender groups; however, the distribution of lives across IRT categories does appear to vary somewhat when split by demographic groups (see Appendix). We will address this heterogeneity by controlling for demographic variables in our subsequent analysis of multivariate models.

**FIGURE 4: EMPIRICAL RELATIVE MORTALITY AND DISTRIBUTION OF INVESTMENT RISK TOLERANCE**



## MODELING APPROACH

Our analysis focused on assessing the predictive power of consumer data in relation to other data types. In order to get valid understanding of the predictive power of consumer data and to perform head-to-head comparisons of the predictive power against other data types, multiple gradient boosted tree models were trained on various combinations of prescription, medical, credit, and consumer input. The analysis is organized into two parts: analysis of "single data source" models, and analysis of "combined data source" models.

The single data source models are simply models that were trained on one of the prescription, medical, credit, or consumer data features, along with age and gender. A control model, based on age and gender only, is included as a baseline for comparison. Starting with an analysis of models trained on each data source in isolation is meant to establish the efficacy of each data source under consideration.

The combined data source model analysis compares the predictive performance of adding consumer data to a model that already contains other data types. In particular, we built a model with prescription and medical input data to compare with a model using prescription, medical, and consumer data. Likewise, we trained and compared models on prescription, medical, and credit data both with and without consumer data.

Models were trained across all the various input data combinations described above using a common set of training parameters as a simple way to avoid advantaging any one model. The target variable used across all the models we considered was simply whether or not a death had been observed (i.e., a binary outcome).

Finally, we split the data into 75% training, 15% test, and 10% holdout in order to get valid out-of-sample model performance assessments.

## MODELING RESULTS

In this section, we review predictive performance assessments of the models described in the previous section with the ultimate goal being to better understand the impact of incorporating consumer data into mortality predictive models.

Figure 5 provides several performance metrics evaluated on test data. These metrics include the Brier score and Spiegelhalter's Z calibration statistic, as well as area under curve (AUC), for both the receiver operating characteristic (ROC) curve and the precision-recall (PR) curve. The Brier score is a general measure of model performance that is influenced by both the discrimination power and accuracy (i.e., calibration) of model predictions. Spiegelhalter's Z calibration statistic is intended to help identify whether differences in the Brier score across models are due to calibration or discrimination.

For each of the models described in the previous section, we report the resulting performance measures on *all* corresponding test data hits. However, as illustrated in Figure 1 above, there is an appreciable difference in mortality experience across the various data source hit cohorts (in particular, the relative mortality of consumer data hits deviates widely from other data hit cohorts). Consequently, it may not be clear whether model performance can be attributed to the actual predictive power of the data or the set of individuals upon which the performance metric is evaluated. To remove this ambiguity, we include a second evaluation on the set of individuals with data from each of the data sources under consideration (i.e., the "intersection" of the data sources).

### Single data source model results

When referring to results for single data source models evaluated on all corresponding data hits, we see an interesting trade-off between coverage of credit data and the predictive power of consumer data. Consumer data appears to offer stronger ability to discriminate between mortality outcomes (evident in AUC). This appears to be true even when limiting the comparison to the cohort of individuals who are hits for all data sources. Having noted that, credit data can provide predictions on a significantly larger group of people.

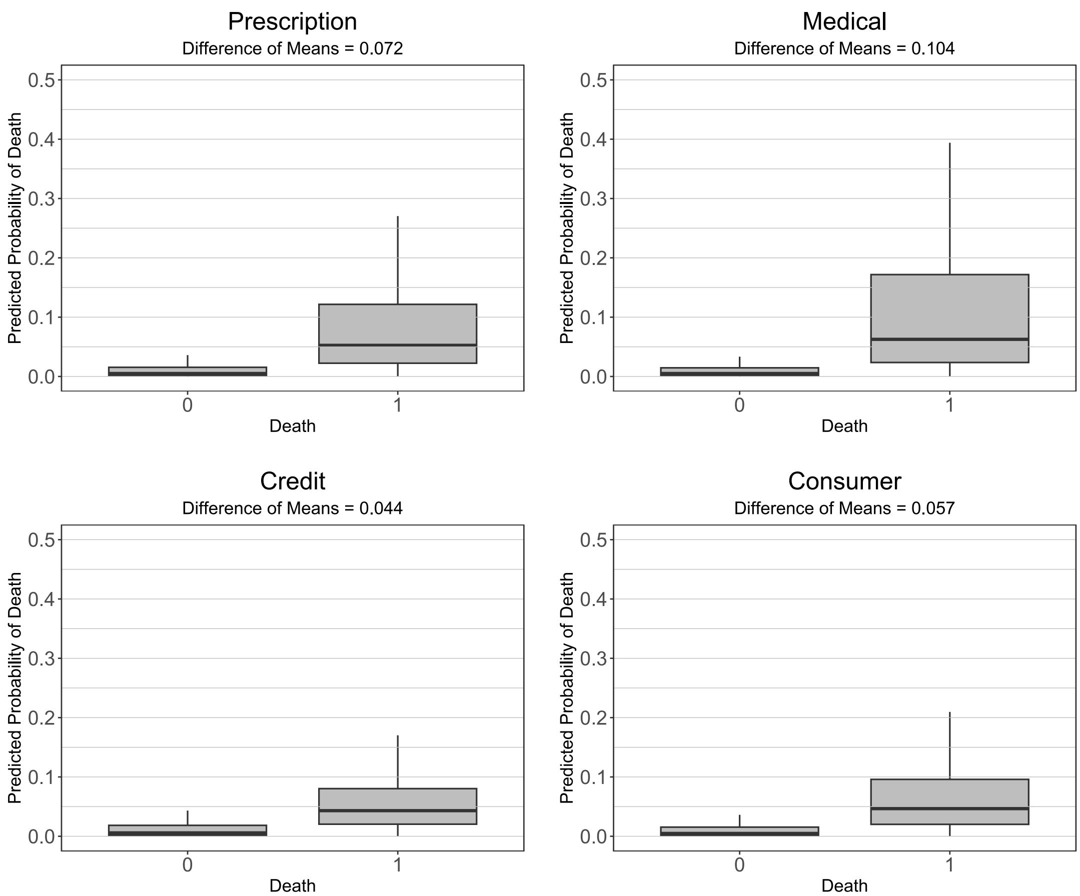
One interesting aspect of these results is the comparison between prescription data and consumer data. Looking at results when both models are evaluated on all data hits corresponding to each model, prescription data appears to achieve stronger discrimination power than consumer data. However, this reverses when limiting the comparison to the intersection of all data hits.

**FIGURE 5: EVALUATION METRICS FOR THE CONTROL VS. VARIOUS SINGLE DATA SOURCE MODELS**

| MODEL | DATA TYPE HITS | BRIER | $Z_S$ | AUC ROC | AUC PR |
|---|---|---|---|---|---|
| **Evaluated on all corresponding data hits** | | | | | |
| **Control** | 3,162,961 | 0.0155 | 19.922 | 0.7909 | 0.0695 |
| **Prescription** | 2,455,936 | 0.0139 | 0.389 | 0.8642 | 0.1375 |
| **Medical** | 2,433,341 | 0.0152 | -1.437 | 0.8778 | 0.1956 |
| **Credit** | 2,987,301 | 0.0147 | -2.687 | 0.8406 | 0.0943 |
| **Consumer** | 2,279,777 | 0.0152 | -0.210 | 0.8556 | 0.1247 |
| **Evaluated on intersection of data hits** | | | | | |
| **Control** | 1,593,346 | 0.0148 | -14.135 | 0.7851 | 0.0639 |
| **Prescription** | 1,593,346 | 0.0145 | -7.749 | 0.8374 | 0.0905 |
| **Medical** | 1,593,346 | 0.0137 | -14.740 | 0.8782 | 0.1848 |
| **Credit** | 1,593,346 | 0.0143 | 0.153 | 0.8544 | 0.1171 |
| **Consumer** | 1,593,346 | 0.0141 | -8.079 | 0.8652 | 0.1399 |

Note: A lower Brier score indicates better performance. For $Z_s$, a value closer to 0 indicates better performance. For AUC ROC and AUC PR, scores closer to 1 indicate better performance.

**FIGURE 6: DISTRIBUTION OF MODEL-PREDICTED PROBABILITIES OF DEATH BY MORTALITY OUTCOME FOR SINGLE DATA SOURCE MODELS EVALUATED ON INTERSECTION OF DATA HITS**



To get a visual comparison of the discrimination power of the various models summarized above, we include boxplots of the predicted probability of death across mortality outcome for the intersection of data hits in Figure 6. This type of boxplot display is a common alternative to the familiar ROC curves, which are difficult to distinguish for some of the models in our study, given the similarity in predictive power (as is evidenced by similar magnitudes of the AUC for the ROC). Additionally, we display difference of means for the two groups, which is often referred to as the “Yate’s discrimination slope.”

### Combined model results

In this section, we look at the gain in predictive performance that can be achieved by adding consumer data to a model with other data types. The results are summarized in Figure 7. These models are labeled according to the data elements they contain, which are abbreviated as follows: “Pr” = prescription, “Md” = medical, “Cr” = credit, and “Co” = consumer.

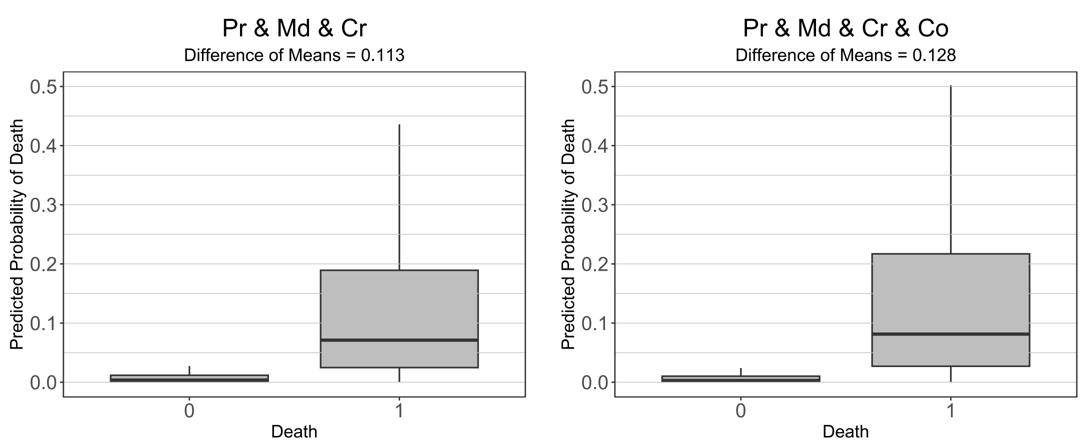
**FIGURE 7: EVALUATION METRICS FOR VARIOUS COMBINED MODELS**

| MODEL | DATA TYPE HITS | BRIER | $z_s$ | AUC ROC | AUC PR |
|---|---|---|---|---|---|
| **Evaluated on all corresponding data hits** | | | | | |
| **Pr & Md & Cr** | 3,101,726 | 0.0143 | 0.078 | 0.8837 | 0.1910 |
| **Pr & Md & Cr & Co** | 3,111,346 | 0.0142 | 0.899 | 0.8922 | 0.2069 |
| **Evaluated on intersection of data hits** | | | | | |
| **Pr & Md & Cr** | 1,593,346 | 0.0135 | -1.993 | 0.8922 | 0.2012 |
| **Pr & Md & Cr & Co** | 1,593,346 | 0.0133 | 0.916 | 0.9027 | 0.2237 |

Comparing the “Pr & Md & Cr” model results to those for the “Pr & Md & Cr & Co” models suggests that the addition of consumer data yields an improvement in all model performance metrics. Furthermore, this is true when evaluating on the *intersection* of data hits as well as when evaluating over *all* possible data hits. Moreover, the data coverage of the combined data sources goes from 3,101,726 records for the “Pr & Md & Cr” model to 3,111,346 for the “Pr & Md & Cr & Co” model; that is, we see a roughly 5% increase in data coverage from adding consumer data to a model with all other data types.

Inspection of predicted probability distribution by death outcome (again, displayed via side-by-side boxplots) supports our findings based on performance metrics. In particular, we see that the model including consumer data achieves a larger degree of discrimination between the distributions of predictions for the two outcomes.

**FIGURE 8: DISTRIBUTION OF MODEL-PREDICTED PROBABILITIES OF DEATH BY MORTALITY OUTCOME FOR COMBINED DATA SOURCE MODELS EVALUATED ON INTERSECTION OF DATA HITS**



## Top 10 feature importance in models with consumer data

FIGURE 9: FEATURE IMPORTANCE FOR THE MODEL CONTAINING ALL USED DATA TYPES

| Feature | Co Model Rank | Pr & Md & Cr & Co Model Rank |
|---|---|---|
| **Age** | 1 | 1 |
| **Bankruptcy Docket Date (diff. months)** | 2 | 2 |
| **Estimated Household Income** | 3 | 25 |
| **Investment Risk Tolerance** | 4 | 11 |
| **Gender** | 5 | 7 |
| **Percent Moved Since 2000** | 6 | 3 |
| **Use Wearable Device to Manage Health** | 7 | 5 |
| **Savers – Long Term** | 8 | 17 |
| **Insurance Amount – Whole Life** | 9 | 51 |
| **Nielsen County Size** | 10 | 33 |

In order to gain a better understanding of what aspects of consumer data might be driving the observed predictive performance, we look at the importance of features among the models containing consumer data. In Figure 9, we report the variable importance rank among consumer features ranking (according to gain) in the top 10 of all features included in the “Co” model and the corresponding rank in the “Pr & Md & Cr & Co” model. Note that both models include age and gender demographic variables.

We see that, while there is some variation in the ranking of consumer features, "Bankruptcy Docket Date (diff. months),“ “Percent Moved Since 2000,” and “Use Wearable Device to Manage Health” rank in the top 10 in both models. A few other variables are ranked in the top 25 in both models. It is remarkable that so many consumer data features rank this highly in the combined model as the input data is comprised of multiple data sources.

## Conclusions

Consumer data can be used to effectively stratify mortality risk. Specifically, we noted a consistent improvement across a variety of performance metrics, which suggests that both the ability to discriminate between mortality outcomes as well as predictive accuracy can be improved with consumer data. Additionally, several consumer variables rank high in variable importance for combined models, which include multiple data types.

Overall, consumer data offers predictive performance comparable to other FCRA-compliant input data types that may be used in predictive models of mortality risk (i.e., prescription, medical, or credit). The most obvious limitation of consumer data is the fact that it is not FCRA-compliant, thus cannot be used in underwriting. However, consumer data could prove a useful alternative for other uses such as marketing or in-force management.



**CONTACT**

Daniel Glab
daniel.glab@milliman.com

Michael Niemerg
michael.niemerg@milliman.com

# Appendix

Figures 10 and 11 present the empirical relative mortality and distribution of lives across IRT categories, split by demographic groups. While there is some variation in distribution of lives, as well as in the observed relative mortality values when comparing across demographic groups, there is a notable consistency in the rank ordering of mortality risk across IRT groups. Relative mortality tends to be higher for the “low” IRT group than for either the “moderate” or “high” groups. Similarly, the “no hit” IRT group typically has a higher relative mortality than either the “missing value” or “high” IRT groups. Other similar comparisons among IRT groups frequently hold, but there are exceptions. For example, the “U” shape exhibited in these charts breaks down for the 70+ age group.

**FIGURE 10: EMPIRICAL RELATIVE MORTALITY AND DISTRIBUTION OF INVESTMENT RISK TOLERANCE SPLIT BY GENDER**

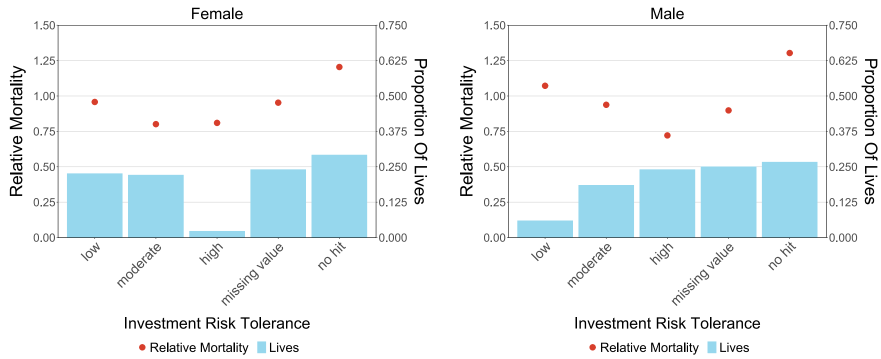


**FIGURE 11: EMPIRICAL RELATIVE MORTALITY AND DISTRIBUTION OF INVESTMENT RISK TOLERANCE SPLIT BY AGE GROUP**