# A Machine Learning Approach to Incorporating Industry Mortality Table Features Into a Company's Insured Mortality Analysis

**SOCIETY OF ACTUARIES**

# A Machine Learning Approach to Incorporating Industry Mortality Table Features Into a Company's Insured Mortality Analysis

| **AUTHOR** | Marc Vincelli, ASA | **SPONSOR** | Financial Reporting Section |
|---|---|---|---|
| | | | Modeling Section |
| | | | Product Development Section |
| | | | Reinsurance Section |

# CONTENTS

# Acknowledgments

# A Machine Learning Approach to Incorporating Industry Mortality Table Features Into a Company's Insured Mortality Analysis

This paper introduces a novel framework for leveraging the "architecture" of an industry mortality table within a company's predictive analytics-based insured mortality analysis. The author shows that by reverse-engineering the industry mortality table into a series of higher-dimensional features and then using those features as inputs to a nonlinear predictive model (in this case a neural net), a company can better model relationships between mortality cells across the full spectrum of ages and durations when faced with sparse experience data. One potential application of this approach is in the initial calibration of an industry mortality table, via its learned features, to a company's own experience.

## Section 1: Introduction

Best estimate insured mortality is most commonly represented by a series of tables (mortality tables) reflecting the mortality rates for policyholders with various combinations of issue age (age at policy issue), duration (time since policy issue), gender and smoker status. Other factors, such as face amount band, may also be modeled. The generally accepted approach to deriving the structure and content of these tables is via an in-depth experience analysis that involves populating the tables with raw mortality rates calculated from carefully prepared observed experience, extrapolating/extending rates to cells with little or no experience, smoothing the rates, assessing the appropriateness of the results (via Actual-to-Expected ratios (A/Es), monotonicity checks, etc.), and iterating back through the cycle as necessary.

While the aforementioned first principles approach has proved effective in industry mortality table development, it can be challenging for individual companies with limited data and/or experience study resources to implement. Alternatives available to companies in developing mortality assumptions include (i) industry table adjustment approaches (e.g., factor-based adjustments or credibility blending) and (ii) predictive modeling approaches. It is this second alternative that we explore herein.

The use of predictive models to develop best estimate insured mortality rates across the full spectrum of ages and durations has been met with limited success. Zhu et al. (2015) fit various logistic regression models to U.S. insured mortality experience yet ultimately elected not to publish the details of their modeled tables due to limitations with the fit. One of the main challenges that companies face when looking to leverage the predictive modeling approach is that it has no preconceived notion of what a mortality table should look like, leading to issues in fit where experience is sparse and to a lack of appreciation for generally accepted mortality relationships. Furthermore, the predictive modeling approach has historically been unable to leverage the degree of fine tuning, professional judgment and data credibility underlying an

industry table. The *2015 Valuation Basic Table* U.S. industry study alone benefited from the expertise of more than 30 different individuals and consulting firms, as well as the experience data from more than 50 insurance companies.

The author used neural nets as the primary predictive analytics/machine learning approach within this paper, both to reverse-engineer the industry mortality table and for out-of-sample prediction testing. The ability of neural nets to model integer variables (such as attained age and duration) as high-cardinality categorical variables in high-dimensional continuous space (called "entity embeddings" or simply "embeddings" herein), complemented by their ability to handle highly nonlinear behavior, were key to this decision. Furthermore, the author benchmarked the out-of-sample results from neural nets fitted with embeddings to two main effects logistic regression models based on the set of traditional features[1] and to a neural net fitted exclusively with the set of traditional features. In all cases, mortality rate was the target variable of interest.

We selected the U.S. 2015 Valuation Basic Table, age-nearest primary tables (referred to as "2015 VBT" herein), as the industry basis for this project, given its recency and widespread use. The primary tables capture four dimensions in the select period (gender, smoker status,[2] issue age, duration) and three dimensions in the ultimate period (gender, smoker status, attained age). The expectation is that the practitioner could extend the methodology herein to other industry or reference tables as required.

The author completed core data preparation and modeling work in Python v3.6.8 (Anaconda Distribution) and leveraged a combination of the following well-known analytical libraries: NumPy, Pandas, SciPy, SciKit-Learn, StatsModels and Keras.

An inventory of all files accompanying this report can be found in Appendix A.

The remainder of this report is structured as follows:

- Section 2 (Data Preparation for Feature Extraction) details the steps the author took to prepare the 2015 VBT rates for reverse-engineering.
- Section 3 (Extraction of Industry Mortality Table Features) begins with an overview of neural nets, entity embeddings, and entity embedding neural nets. It then discusses the extraction of key features from 2015 VBT, including the design of the relevant entity embedding neural nets, the algorithm used to train these models, and the goodness of fit between the rates predicted from these models and the corresponding 2015 VBT rates.
- Section 4 (Visualization and Validation of Extracted Features) presents the results of visualizing the high-dimensional mortality embeddings in low-dimensional space using t-Distributed Stochastic Neighbor Embedding (t-SNE) and Principal Component Analysis (PCA). A key reason for visualizing the mortality embeddings is to ensure that they have internally consistent behavior; that is, to ensure that cells with more similar mortality characteristics are closer together in the embedding space. Internal consistency improves the likelihood that the mortality embeddings will be portable to related nonlinear

---

[1] For the purposes of this paper, the set of traditional features is defined to be gender as a categorical variable, smoker status as a categorical variable, duration as a continuous variable, and attained age as a continuous variable.

[2] Smoker status has three levels: smoker, nonsmoker and composite/unismoke.

predictive modeling problems, such as the modeling of company-specific best estimate insured mortality.

- Section 5 (Case Study) discusses the specifications and results for a case study designed to demonstrate the value of the mortality embeddings within a nonlinear predictive modeling context. The case study leverages U.S. insured mortality experience collected by the Society of Actuaries (SOA) Individual Life Experience Committee (ILEC) for years 2009–2015 and "sparsifies" these data to better reflect the gaps that a company may have in its own experience data. Both fit and generally accepted mortality relationships are assessed.
- Section 6 (Conclusion) summarizes the key aspects of this project and the various levels on which the material herein can be leveraged.

## Section 2: Data Preparation for Feature Extraction

This section describes the steps the author took to import and reformat the 2015 VBT rates.

### 2.1  2015 VBT DATA PREPARATION

The process used to load and reformat the 2015 VBT rates consisted of four steps.

### *Step 1: Load 2015 VBT Primary Tables, Age Nearest Basis (ANB)*

In this step, the author downloaded the relevant Excel files (2015-vbt-smoker-distinct-alb-anb.xlsx and 2015-vbt-unismoke-alb-anb.xlsx) from the SOA site (www.soa.org/experience-studies/2015/2015-valuation-basic-tables) and subsequently loaded the two-dimensional age nearest rates by issue age and duration for each of male nonsmoker (MNS), male smoker (MSM), male unismoke/unknown (MUNI), female nonsmoker (FNS), female smoker (FSM), and female unismoke/unknown (FUNI) into a series of Pandas dataframes.

### *Step 2: Validate Data Load*

Here the author validated the imported data back to the source Excel files based on hash totals. The totals in Table 1 (sums of $q_x$ per thousand) calculated from the imported data were found to match back to the source files exactly:

**Table 1**
RESULTS OF DATA LOAD VALIDATION

| Table | Hash Total | Result |
|-------|------------|--------|
| MNS | 177,590.48 | Match |
| MSM | 197,317.89 | Match |
| MUNI | 182,755.43 | Match |
| FNS | 162,014.06 | Match |
| FSM | 187,187.35 | Match |
| FUNI | 167,845.31 | Match |

### *Step 3: Reformat Data*

In this step, the author reformatted the data from a series of 2-D tables (i.e., select period rates by issue age and duration, followed by ultimate rates) to a single, columnar dataframe format. He also derived attained age as issue age + duration − 1. The top of the resulting dataframe is shown in Table 2.

**Table 2**
EXCERPT OF TRANSFORMED DATAFRAME

| | Gender | Smoker_Status | Issue_Age | Duration | Attained_Age | qx |
|---|--------|---------------|-----------|----------|--------------|------|
| 0 | M | NS | 18 | 1 | 18 | 0.69 |
| 1 | M | NS | 19 | 1 | 19 | 0.64 |
| 2 | M | NS | 20 | 1 | 20 | 0.55 |
| 3 | M | NS | 21 | 1 | 21 | 0.46 |
| 4 | M | NS | 22 | 1 | 22 | 0.45 |

*Step 4: Refine Select Period vs. Ultimate Period Boundaries*

The 2015 VBT tables have select periods that vary by gender, smoker status and issue age. Since this variability was not explicitly represented in the mortality tables imported above (i.e., all imported tables appeared to have a consistent 25-year select period based on their structure alone), the author needed to reflect this variability by appending information on select period length from the *2015 VBT* report[3] and then refining the select/ultimate period boundaries accordingly. Records outside the select period were reclassified as ultimate period records, and all duplicate ultimate period records resulting from this process were deleted. This step resulted in 9,174 unique 2015 VBT select and ultimate period records (or cells).

## 2.2  CREATION OF DERIVED VARIABLES

Once the 2015 VBT rates were loaded and appropriately reformatted, the author proceeded to define a series of derived variables for use in subsequent modeling work. These new variables included (i) risk class, derived as the concatenation of gender and smoker status; (ii) gender x duration, derived as the concatenation of gender and duration; and (iii) grouped attained age x grouped duration, derived as all relevant combinations of quinquennial attained age groupings and quinquennial duration groupings.

## 2.3  ZERO-INDEXED ENCODINGS

Many neural net software packages expect categorical variables to be represented as zero-indexed integer variables, with the first level taking a value of zero, the second level taking a value of 1, and so on. Accordingly, the variables created in subsection 2.2 were converted to zero-indexed integer variables. Similarly, gender, smoker status and duration were also re-encoded as zero-indexed integer variables. Attained age required no re-encoding because it was naturally zero-indexed.

## 2.4  TRANSFORMATION OF DEPENDENT VARIABLE (MORTALITY RATE)

As expected, the distribution of $q_x$ was found to be highly right skewed, with rates ranging from 0.06 per thousand to 500 per thousand. Since the neural net models used to reverse engineer 2015 VBT could have a hard time modeling skewed target variables, the author considered various transformations to make the target distribution more symmetric and standardized. Among the options considered was the family of Box-Cox transformations and the log-odds (logit) transform.[4] In the end, the author decided on a natural log transform, followed by a min-max transform to [0, 1], which proved effective and easy to understand. The min-max transform to [0, 1] was intended to mirror the possible range of output from the neural net models. The result of this double transform is shown in Figure 1.

---

[3] Society of Actuaries. 2018. 2015 Valuation Basic Table Report & Tables. *Society of Actuaries,* pages 37, 59, *https://www.soa.org/experience-studies/2015/2015-valuation-basic-tables* (accessed on Aug. 5, 2019).

[4] $logit(q_x) = log(q_x / (1 - q_x)) \approx log(q_x)$ for small $q_x$

**Figure 1**

TRANSFORMATION OF QX USING NATURAL LOG AND MIN-MAX TRANSFORMS TO PRODUCE A MORE
SYMMETRICAL TARGET DISTRIBUTION WITH ALL VALUES IN [0, 1]
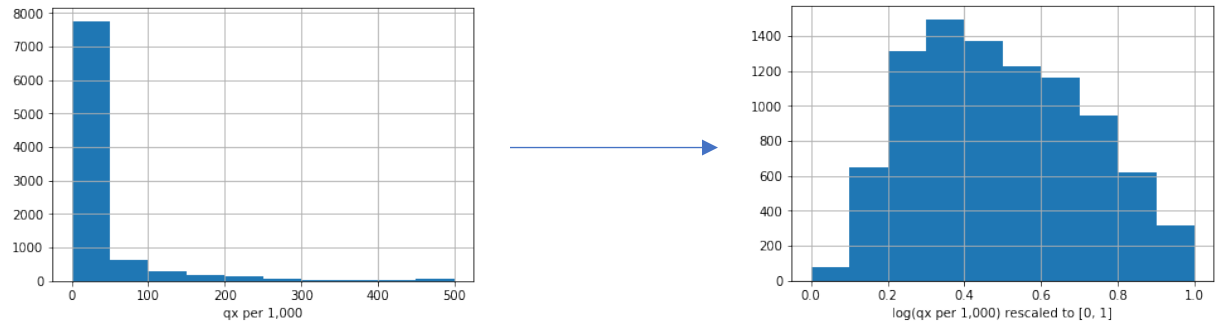
## Section 3: Extraction of Industry Mortality Table Features

This section begins with an overview of neural nets, entity embeddings and entity embedding neural nets (EENNs). The author then moves on to how he implemented these tools to extract mortality embeddings from 2015 VBT, including the principles followed in designing the relevant EENNs, the algorithm used to train these models, and ultimately the selected hyperparameters.

### 3.1 NEURAL NETS[5]

A neural net is a machine learning system that can be visualized as a network of connected "neurons" arranged in layers. Neurons in adjacent layers are connected by weights (or trainable parameters) that can be learned in a supervised manner. The first layer is called the "input layer," while the last layer is called the "output layer." Any layers in between are called "hidden layers." The neurons in the hidden and output layers typically consist of a linear summation aggregation function, followed by a nonlinear activation function (e.g., logistic function, rectified linear unit or ReLU,[6] etc.). More specifically, one can think of a neural net as a series of linear model ($f(x) = \sum w_i x_i + bias$) and nonlinear activation function pairs, arranged in a tree-like structure. The greater the number of nodes in the middle layer(s), the greater the ability of the neural net to model nonlinear patterns. It is the ability of neural nets to model highly nonlinear phenomena that makes them good candidates for modeling complex natural behavior.

Figure 2 depicts a sample neural net with four input nodes, two hidden nodes and one output node. For simplicity, the author assumed that this neural net does not include any bias (constant) terms and that both gender and smoker status are encoded as binary indicators. Note that each connection represents an independent weight, so in this example, there are a total of 10 weights or 10 trainable parameters.

---

[5] This section refers to standard feed-forward multilayer perceptron neural nets applied in a regression context.
[6] ReLU, or rectified linear unit, is the function g(x) = max(0, x).

**Figure 2**
DIAGRAM OF A SAMPLE NEURAL NET WITH ONE HIDDEN LAYER
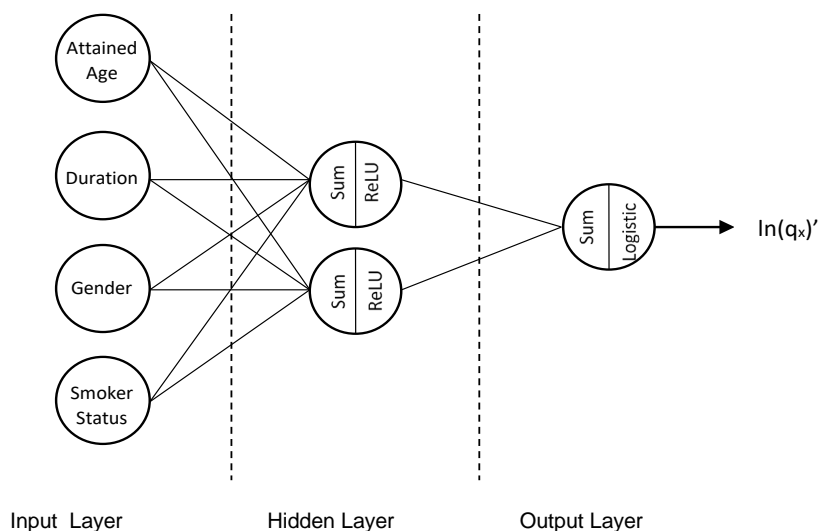


The output from this neural net, $\ln(q_x)'$, which represents $\ln(q_x)$ transformed to [0,1], can be thought of as being calculated according to the following steps: first, the inputs (e.g., attained age, duration, etc.) are multiplied by their corresponding weights and aggregated; second, the results from Step 1 are put though a ReLU; third, the results from the hidden nodes are multiplied by their corresponding weights and aggregated; and finally, the results from Step 3 are put through a standard logistic function whose output necessarily falls in [0, 1].[7] Note that the activation functions used in the various layers need not be the same.

Once the structure of the neural net has been defined—including the number of layers, number of nodes in each layer, and nature of the nonlinear activation function(s)—the network can be trained using a process called "backpropagation." During this process, the trainable weights in the network are updated in an iterative fashion in an attempt to minimize the chosen loss function. Commonly used loss functions include mean-squared error and mean absolute error.

Parameters that must be chosen by the practitioner prior to training are called "hyperparameters." All parameters defining the structure of neural net, along with parameters that define how the backpropagation process unfolds—primarily the learning rate (a measure of how quickly the weights are updated during training) and the number of epochs (the number of times that the training set is passed through the network during training)—are hyperparameters.

Unfortunately, there are no rules for determining the best hyperparameters to choose for a given problem. Practitioners tend to view neural net modeling as an art, with a lot of trial and error involved. That said, the practice of cross-validation—where the training set is split into training

---

[7]The standard logistic function takes the form $g(x) = 1/(1 + \exp(-x))$.

and validation subsets and the prediction error is measured on the out-of-sample validation subset(s)—can add rigor to the process of selecting hyperparameters by assessing the predictive accuracy of various hyperparameter combinations.

Refer to Nisbet et al (2009) and Abbott (2014) for further details on basic neural nets.

### 3.2 ENTITY EMBEDDINGS

Entity embeddings (EEs) are representations of discrete categories in higher-dimensional continuous (Euclidean) space, with similar categories being closer together.

For example, in natural language processing, word embeddings map words to a continuous semantic space. If (say) three words are represented in three-dimensional space by the vectors (2.5, 2.0, 3.1), (2.5, 1.5, 3.0) and (10.0, 10.0, 3.0), respectively, then the conclusion is that the first two words are much more similar in semantic usage to one another than either is to the third word. Similarly, in a mortality context, if two attained ages, durations or risk classes (or combinations thereof) are placed closer in continuous mortality space, the corresponding cells would be expected to have more similar mortality rates than if they were further apart.

The more complex the relationship between categories, typically the higher the dimensionality required for a suitable embedding space.

Embedding tables map each unique category of the original variable to its higher-dimensional representation. Table 3 shows what the embedding table, or "lookup" table, may look like for smoker status. If the desire is to use a richer representation of smoker status, one could replace the single variable smoker_status, with the two variables ee_ss_1 and ee_ss_2.[8] In this way, NS would be replaced by (0.429039, -0.35202), SM by (-0.62347, 0.019604), and UNI by (0.357066, 0.332035), wherever they appeared in the input dataset. Nonlinear models—such as neural nets, k-nearest neighbors, and random forests—can often take advantage of the richer structure offered by entity embeddings.

**Table 3**
SAMPLE EMBEDDING TABLE FOR SMOKER STATUS

| smoker_status | ee_ss_1 | ee_ss_2 |
|---|---|---|
| NS | 0.429039 | -0.35202 |
| SM | -0.62347 | 0.019604 |
| UNI | 0.357066 | 0.332035 |

### 3.3 ENTITY EMBEDDING NEURAL NETS

Entity embedding neural nets (EENNs) combine the features of neural nets, as discussed in subsection 3.1, with the ability to learn the entity embedding representation for one or more input categorical variables. It is important to note that these types of neural nets are structurally different from standard neural nets, since they explicitly incorporate the embedding table (or comparable lookup process) for each designated discrete variable. Whereas a standard neural

---

[8] The name ee_ss_1 stands for "entity embedding–smoker status–component 1."

net could leverage an input dataset that includes embeddings (for example, once that includes ee_ss_1 and ee_ss_2 in place of smoker_status), it cannot generate the embedding itself; that is, it cannot populate the values in Table 3, for example. If the practitioner wishes to develop an embedding representation, the individual may use entity embedding neural nets.

Figure 3 shows a sample entity embedding neural net. This neural net is similar to the one depicted in Figure 2 but includes the machinery to look up the embedding for each original variable. One other difference is that in Figure 3, it is assumed that smoker status has three levels, matching its representation in Table 3. Attained age and duration are each assumed to have embeddings with five dimensions (i.e., five components), while smoker status is assumed to have an embedding with two dimensions. There is no benefit to modeling gender with an embedding since it has only two levels; consequently, we depict only its original input.

**Figure 3**
DIAGRAM OF A SAMPLE ENTITY EMBEDDING NEURAL NET
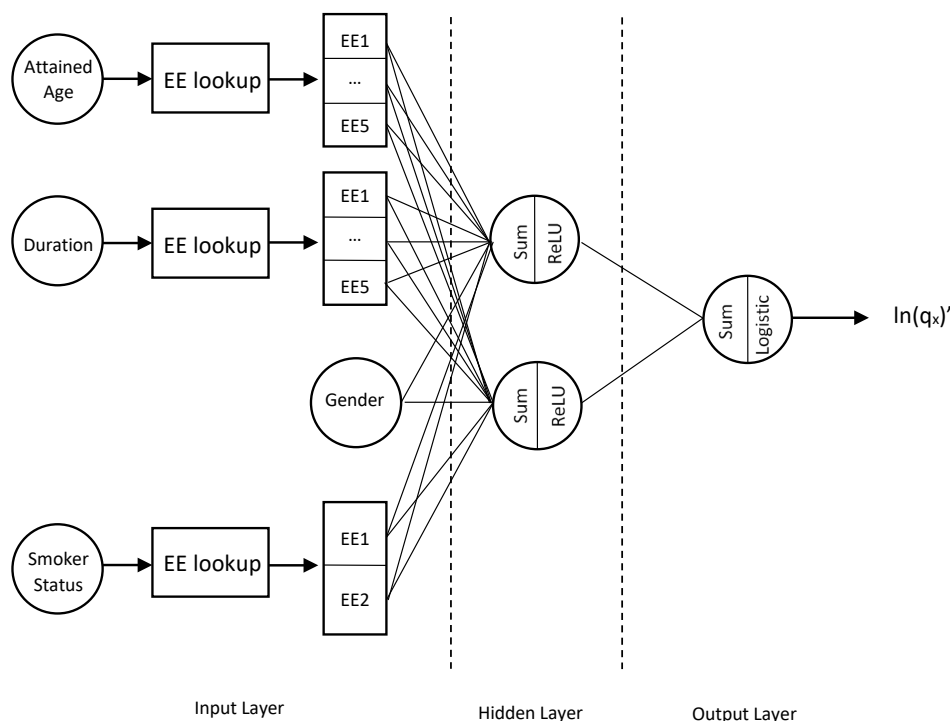


The entity embeddings can be thought of as an additional layer of weights that sits between the input layer of categorical variables and the first hidden layer. Each input record is effectively mapped to its corresponding entity embedding representation (i.e., vector of weights), before moving on to the first hidden layer. The weights in the entity embedding layer (denoted by "EE lookup" in Figure 3) can be learned in the same way as those in the other layers—that is, via

backpropagation. Consequently, for the network depicted in Figure 3, the total number of trainable parameters is equal to the number of connections (28) plus the number of values in the EE lookup tables (again, assuming no bias terms).

For further details on entity embeddings and EENNs, refer to Guo and Berkhahn (2016).

### 3.4  VALUE OF ENTITY EMBEDDINGS

Once an entity embedding neural net has been trained, the weights in the embedding layer can be extracted and reused as a powerful high-dimensional representation of the relationship between categories in related applications. In this way, and as discussed in further detail in this report, the author could (1) train an EENN to predict the mortality rates from a benchmark table, (2) extract the learned mortality embeddings representing the "architecture" of the table, and (3) reuse the learned architecture in applications expected to have comparable, though not identical, behavior. The ability to feed the learned architecture of the industry table into a new (standard) neural net and train that new neural net to calibrate this mortality architecture to a company's specific experience is one promising application of this novel framework.

### 3.5  DESIGN PRINCIPLES FOR EENNs

As discussed previously, designing and training a neural net is as much an art as it is a science. However, given that the search space is potentially unlimited, the author opted to put in place some high-level principles from the outset to guide the design of EENNs for use in reverse engineering 2015 VBT. These principles were developed based on the intended application of the derived embeddings (i.e., portability), as well as rules of thumb published by other practitioners. The five design principles that the author ultimately implemented are documented in Table 2.

**TABLE 4**
EENN DESIGN PRINCIPLES

| Principle | Description | Rationale |
|---|---|---|
| 1 | Push as much information (in this case, information about the architecture of 2015 VBT) from the hidden layer(s) to the embeddings as possible. | The embeddings are the components of the EENNs that the author is looking to extract and leverage in a predictive modeling context. The EENN itself is discarded once it has served its purpose. The richer the embeddings, the more valuable they are likely to be in a predictive modeling context. |
| 2 | Set the dimensionality for each embedding equal to min(5, (c+1)/2), rounded down to the nearest integer if applicable, where c denotes the cardinality[9] of the original underlying variable. Ensure that the total dimensionality of all embeddings is within one-tenth of the total cardinality of all original variables. | The author wanted to restrict the dimensionality of the embeddings to a reasonable number based on (i) the cardinality of the original variables and (ii) the complexity of the modeling problem. Embeddings with too high a dimensionality will be inefficient, while embeddings with too low a dimensionality will be ineffective. The formula shown is based on a rule of thumb from Jeremy Howard,[10] with a cap of five. The author chose a cap of five given the small size of the 2015 VBT dataset. |

---

[9] The cardinality of a categorical variable is the number of its unique levels. For example, if a variable has levels (smoker, nonsmoker, unismoke), it has a cardinality of three.

[10] Jeremy Howard is a founding member of fast.ai. He discusses this rule-of-thumb at the 56:50 mark in this lecture: Howard, J. 2017. Lesson 4: Deep Learning 2018. *YouTube*, Dec. 30. *https://www.youtube.com/watch?v=gbceqO8PpBg* (accessed Aug. 6, 2019).

| 3 | Keep the total number of neural net weights to within one-fourth of the total number of training records. | The author wanted the EENN to learn the architecture of the mortality table, not simply reproduce it. The intent of this principle is to force the EENN to learn generalizable characteristics. |
|---|---|---|
| 4 | Structure each subsequent layer to have fewer weights than the layer prior. | The author enforced a funnel structure on the EENN to encourage the backpropagation algorithm to transmit meaningful weight changes upstream, all the way back to the embeddings themselves. |
| 5 | Restrict the number of hidden layers to a maximum of two. | Two or fewer hidden layers normally suffice for structured data problems such as this. Very few problems outside of computer vision and complex time series analyses benefit from more than two hidden layers.[11] |

### 3.6 TRAINING ALGORITHM FOR EENNs

This subsection describes the 12-step algorithm used to train the EENNs (and, in turn, the embeddings themselves) to model 2015 VBT to within 5% mean absolute percentage error (MAPE).

1. Split dataset containing 9,174 records into training (in-sample) and validation (out-of-sample) sets; randomly assign 95% for training and 5% for validation.

   To help visualize this step, imagine the two-dimensional age and duration mortality tables for each of the six risk classes (MNS, MSM, MUNI, FNS, FSM, FUNI) being stacked one on top of another to form a three-dimensional lattice. It then creates holes in this training lattice by randomly moving 5% of records (cells) to a validation set. The training/validation process forces the model(s) to fill in these holes based on information learned from similar cells.

   The purpose of splitting data into training and validation sets was to help with the selection of reasonable hyperparameters, in the spirit of cross-validation. Note that the author used all data when determining the final embeddings, per Step 11. He used a high training-to-validation ratio due to the limited size of the 2015 VBT dataset.

   The author takes the view that a mortality table is greater than the sum of its parts. Since the author was interested in learning the entire architecture of 2015 VBT across all ages, durations, and risk classes, he explicitly avoided weighting the training sets toward one part of the table versus another.

2. Select initial EENN design hyperparameters based on the design principles from subsection 3.5.

---

[11] For example, see Heaton, Jeff. 2017. The Number of Hidden Layers. *Heaton Research,* June 1, *https://www.heatonresearch.com/2017/06/01/hidden-layers.html* (accessed Aug. 6, 2019).

3. Define and train the EENN in Python Keras. Optimizer: Adam; batch size: 128; loss: mean absolute error (MAE);[12] activation function for hidden layer(s): ReLU; activation function for output layer: Sigmoid/Logistic.

4. Record validation MAPE and training MAPE on the original (not transformed) $q_x$ basis.

5. Repeat Steps 1–4 with different random seeds for a total of five runs.

6. Average validation MAPE across runs.

7. Average training MAPE across runs.

8. Iteratively adjust the learning rate, number of epochs, number of nodes in first hidden layer, and number of nodes in second hidden layer parameters, in line with the design principles, until the average validation MAPE is less than or equal to 5%.

9. Fine-tune the EENN hyperparameters; see whether further material performance improvements are possible.

10. Confirm that the average training MAPE $\leq$ average validation MAPE $\leq$ 5%.

11. Train the EENN with the selected hyperparameters on the entire dataset.

12. Extract the embeddings.

### 3.7 SELECTED EENNs AND HYPERPARAMETERS

This section documents the parameterization and performance of three EENNs, selected on the basis of the design principles and training algorithm documented above. Each subsequent model represents an evolution of the preceding model. The author has elected to show all three models for two reasons: First, seeing the evolution of results is informative; second, the practitioner may prefer a simpler set of embeddings (say, from the first model) to a more complex set (say, from the third model). Furthermore, the author is not suggesting that the selected EENNs and their associated hyperparameters are necessarily optimal; he selected them because they aligned with his design principles, met the performance criteria outlined in the training algorithm, and were found to be effective for the intended application. Three sets of detailed embeddings were extracted from these three models.

---

[12] In this case, MAE was calculated on the transformed $q_x$ basis.

**EENN Model 1**
*ln($q_x$)' = EENN(risk class, attained age, duration)*[13]

This model includes embeddings for the main effects risk class, attained age and duration. While its embedding space is the simplest of the three models, it requires more complex hidden layers to meet the out-of-sample performance criteria.

Table 5 shows the selected EE dimensions for each variable in Model 1. Note that the total EE dimensionality (13) is less than one-tenth of the total cardinality of the original variables (153).

**TABLE 5**
SELECTED EE DIMENSIONS FOR EENN MODEL 1

| Variable | Values | Cardinality | EE Dimension |
|----------|--------|-------------|--------------|
| Risk Class | FNS, FSM, .., MUNI | 6 | 3 |
| Attained Age | 0, 1, …, 120 | 121 | 5 |
| Duration | 1, 2, …, 25, ULT | 26 | 5 |
| TOTAL | N/A | 153 | 13 |

Table 6 shows the selected structure for this model. The +1 in the number of weights calculation for the hidden and output layers represents a bias term. The number of EE weights (753) is calculated as the sum product of cardinality and EE dimension from Table 5. Note that the total number of weights (2,078) is less than one-quarter of the total number of data records (9,174). Thinking of the weights as encoding information, only 36% (753/2,078) of this network's weight information is contained in the EE layer.

**TABLE 6**
SELECTED STRUCTURE FOR EENN MODEL 1

| Layer | # Nodes | # Weights |
|-------|---------|-----------|
| EE | 13 | 753 |
| Hidden 1 | 50 | 700 = (50)(13+1) |
| Hidden 2 | 12 | 612 = (12)(50+1) |
| Output | 1 | 13 = (1)(12+1) |
| TOTAL | 76 | 2,078 |

Table 7 shows the MAPE for Model 1 when run with various combinations learning rate, number of epochs and random seed. The row highlighted in green indicates the ultimately selected model. The selection of this variant was based on a combination of performance and stability across runs. Note that the average out-of-sample MAPE for Model 1 was 3.7%, well below the 5% threshold.

---

[13] Throughout this document, ln($q_x$)' denotes ln($q_x$) transformed to [0, 1].

**TABLE 7**
MAPE FOR EENN MODEL 1 (SELECTED VARIANT HIGHLIGHTED IN GREEN)

| LR | # Epochs | MAPE on Validation (Out-of-Sample) Set | | | | | | MAPE on Training (In-Sample) Set | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Seed: 123 | Seed: 999 | Seed: 16416 | Seed: 97625 | Seed: 742354 | AVG | Seed: 123 | Seed: 999 | Seed: 16416 | Seed: 97625 | Seed: 742354 | AVG |
| 0.0001 | 5,000 | 4.7% | 3.3% | 3.3% | 3.6% | 3.6% | 3.7% | 4.3% | 2.6% | 3.0% | 3.2% | 2.9% | 3.2% |
| 0.001 | 500 | 3.9% | 3.2% | 3.7% | 6.0% | 3.9% | 4.1% | 3.8% | 2.9% | 3.1% | 5.7% | 3.8% | 3.9% |
| 0.01 | 100 | 4.6% | 4.3% | 4.7% | 5.8% | 5.3% | 4.9% | 4.3% | 3.8% | 4.3% | 5.5% | 5.2% | 4.6% |

**EENN Model 2**
*ln(q$_x$)' = EENN(risk class, attained age, duration, grouped attained age x grouped duration)*

Model 2 includes embeddings for the main effects risk class, attained age and duration, as well as for the interaction effect grouped attained age x grouped duration. It effectively takes Model 1 and adds an age-duration interaction term.

Table 8 shows the selected EE dimensions for each variable in Model 2. Note that the total EE dimensionality (18) is once again less than one-tenth of the total cardinality of the original variables (252).

**TABLE 8**
SELECTED EE DIMENSIONS FOR EENN MODEL 2

| Variable | Values | Cardinality | EE Dimension |
|---|---|---|---|
| Risk Class | FNS, FSM, .., MUNI | 6 | 3 |
| Attained Age | 0, 1, …, 120 | 121 | 5 |
| Duration | 1, 2, …, 25, Ult | 26 | 5 |
| Grouped Attained Age x Grouped Duration | 0-4xUlt, …, 35-39x1-5, 35-39x6-10, …, 115-120xUlt | 99 | 5 |
| TOTAL | N/A | 252 | 18 |

Table 9 shows the selected structure for this model. Relative to Model 1, note that this model carries significantly more weight information in its EE layer (67% = 1,248/1,849 vs 36%). The total number of weights (1,849) remains below one-quarter of the total number of data records (9,174).

**TABLE 9**

SELECTED STRUCTURE FOR EENN MODEL 2

| Layer | # Nodes | # Weights |
|---|---|---|
| EE | 18 | 1,248 |
| Hidden 1 | 20 | 380 = (20)(18+1) |
| Hidden 2 | 10 | 210 = (10)(20+1) |
| Output | 1 | 11 = (1)(10+1) |
| TOTAL | 49 | 1,849 |

Table 10 shows the MAPE for Model 2 when run with various combinations learning rate, number of epochs and random seed. The row highlighted in green indicates the ultimately selected model. The selection of this variant was based on a combination of performance and stability across runs. Note that the average out-of-sample MAPE for Model 2 was 3.3%, well below the 5% threshold.

**TABLE 10**

MAPE FOR EENN MODEL 2 (SELECTED VARIANT HIGHLIGHTED IN GREEN)

| LR | # Epochs | MAPE on Validation (Out-of-Sample) Set | | | | | | MAPE on Training (In-Sample) Set | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Seed: 123 | Seed: 999 | Seed: 16416 | Seed: 97625 | Seed: 742354 | AVG | Seed: 123 | Seed: 999 | Seed: 16416 | Seed: 97625 | Seed: 742354 | AVG |
| 0.0001 | 5,000 | 2.7% | 3.6% | 3.0% | 4.7% | 4.0% | 3.6% | 2.5% | 3.2% | 2.4% | 3.9% | 3.9% | 3.2% |
| 0.001 | 700 | 3.1% | 3.5% | 3.4% | 3.2% | 3.1% | 3.3% | 2.8% | 3.0% | 3.0% | 2.8% | 2.8% | 2.9% |
| 0.01 | 400 | 2.9% | 3.6% | 3.1% | 3.7% | 5.8% | 3.8% | 2.7% | 3.1% | 2.8% | 3.4% | 5.5% | 3.5% |

**EENN Model 3**

**ln(qx)' = EENN(smoker status, attained age, gender x duration, grouped attained age x grouped duration)**

Model 3 leverages embeddings for the main effects of smoker status and attained age, as well as for the interaction effects gender x duration and grouped attained age x grouped duration. It effectively takes Model 2 and splits duration by gender. The allowance for different select periods by gender is consistent with the structure of 2015 VBT.

Table 11 shows the selected EE dimensions for each variable in Model 3. Note that the total EE dimensionality (17) remains less than one-tenth of the total cardinality of the original variables (270).

**TABLE 11**

SELECTED EE DIMENSIONS FOR EENN MODEL 3

| Variable | Values | Cardinality | EE Dimension |
|---|---|---|---|
| Smoker Status | NS, SM, UNI | 3 | 2 |
| Attained Age | 0, 1, …, 120 | 121 | 5 |
| Gender x Duration | M1, M2, …, M25, MULT, F1, F2, …, F20, FULT | 47 | 5 |
| Grouped Attained Age x Grouped Duration | 0-4xUlt, …, 35-39x1-5, 35-39x6-10, …, 115-120xUlt | 99 | 5 |
| TOTAL | N/A | 270 | 17 |

Table 12 shows the selected structure for Model 3. This model carries 82% (1,341/1,628) of its weight information in its EE layer, versus 67% for Model 2 and 36% for Model 1. These ratios suggest that the embeddings from Model 3 carry the most information about the architecture of 2015 VBT. The total number of weights (1,628) is well below one-quarter of the total number of data records (9,174).

**TABLE 12**

SELECTED STRUCTURE FOR EENN MODEL 3

| Layer | # Nodes | # Weights |
|---|---|---|
| EE | 17 | 1,341 |
| Hidden 1 | 12 | 216 = (12)(17+1) |
| Hidden 2 | 5 | 65 = (5)(12+1) |
| Output | 1 | 6 = (1)(5+1) |
| TOTAL | 35 | 1,628 |

Table 13 shows the MAPE for Model 3 when run with various combinations learning rate, number of epochs and random seed. The row highlighted in green indicates the ultimately selected model. As with Model 1 and Model 2, the selection of this variant was based on a combination of performance and stability across runs. Note that the average out-of-sample MAPE for Model 3 was 3.6%, well below the 5% threshold.

**TABLE 13**

MAPE FOR EENN MODEL 3 (SELECTED VARIANT HIGHLIGHTED IN GREEN)

| LR | # Epochs | MAPE on Validation (Out-of-Sample) Set | | | | | | MAPE on Training (In-Sample) Set | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Seed: 123 | Seed: 999 | Seed: 16416 | Seed: 97625 | Seed: 742354 | AVG | Seed: 123 | Seed: 999 | Seed: 16416 | Seed: 97625 | Seed: 742354 | AVG |
| 0.0001 | 10,000 | 5.0% | 4.1% | 3.7% | 3.7% | 3.1% | 3.9% | 4.3% | 3.2% | 3.1% | 2.9% | 2.6% | 3.2% |
| 0.001 | 2,000 | 3.4% | 3.8% | 3.5% | 4.1% | 3.3% | 3.6% | 3.1% | 3.1% | 2.9% | 3.6% | 2.9% | 3.1% |
| 0.01 | 600 | 3.6% | 3.5% | 4.5% | 3.3% | 4.9% | 4.0% | 3.4% | 3.1% | 3.9% | 3.2% | 4.3% | 3.6% |

## Embedding Dictionaries

The embedding tables associated with EENN Model 1 and EENN Model 3 have been saved to two separate embedding dictionaries (Excel workbooks) that accompany this report. These embeddings are called the "detailed embeddings." See Appendix A for further details.

## Section 4: Visualization and Assessment of Extracted Features

Once the author selected and trained EENNs and extracted their embeddings, he proceeded to visualize the embeddings and assess their reasonableness. A key reason for visualizing the mortality embeddings was to ensure that they had internally consistent behavior; that is, to ensure that cells with more similar mortality characteristics were closer together in the embedding space. Internal consistency improved the likelihood that the mortality embeddings would be portable to similar, though not identical, nonlinear predictive modeling problems.

Since it can be difficult to visualize objects in dimensions higher than two, the author used two different dimension-reduction techniques to help visualize the high-dimensional embeddings in low-dimensional space. The first technique, t-Distributed Stochastic Neighbor Embedding (t-SNE), discussed in van der Maaten and Hinton (2008), is a nonlinear technique that uses probabilistic similarity to keep the low-dimensional representation of similar datapoints close together. The second technique, PCA, further discussed in Abbott (2014), is a linear technique that keeps the low-dimensional representation of dissimilar datapoints far apart. Within the PCA toolkit, the author focused primarily on the PC1 score—that is, the result of projecting a given embedding onto its first principal component.

The visualizations for the embeddings extracted from each of the aforementioned EENNs are now presented in turn.

### 4.1 VISUALIZATION OF EMBEDDINGS FROM EENN MODEL 1

**EENN Model 1**
*ln(qx)' = EENN(risk class, attained age, duration)*

#### Risk Class Embedding

Figure 4 shows the t-SNE plot for risk class. We see that gender is clearly separated, UNI is more similar (closer) to NS than to SM, and MNS/FNS is more similar than MSM/FSM. These relationships are consistent with those in 2015 VBT.

**Figure 4**
T-SNE PLOT FOR RISK CLASS EMBEDDING DERIVED FROM EENN MODEL 1
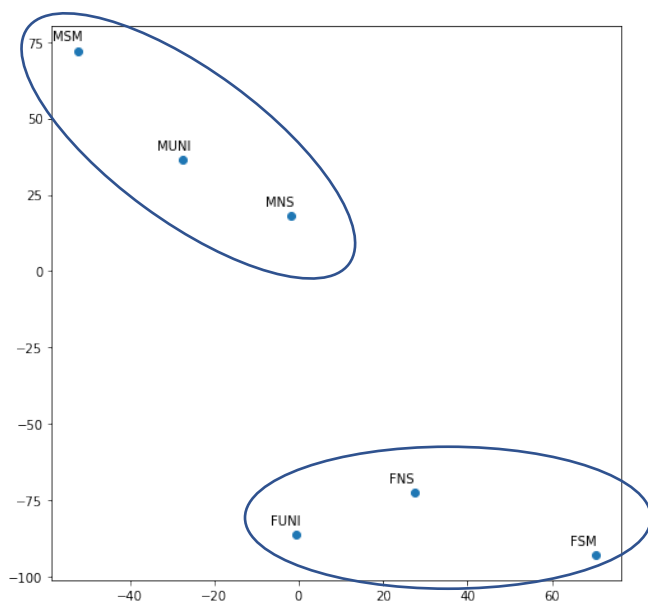


The percentage of variation explained by each of the first two principal components was as follows: PC1: 53%; PC2: 35%. Since the first principal component explains less than 75% of total variation, the author elected not to show a visualization of the PC1 score, because it paints an incomplete picture.

<u>Attained Age Embedding</u>

Figure 5 shows the t-SNE plot for attained age. The points for juveniles (in green) display irregular behavior, whereas the points for core ages (in red) and older ages (in blue) display more regular and well-ordered behavior. Interestingly, the embedding has identified the rates for ages 112 and up (omega rates) as being different enough that it has separated them into their own cluster. These relationships are consistent with those in 2015 VBT.

**Figure 5**

T-SNE PLOT FOR ATTAINED AGE EMBEDDING DERIVED FROM EENN MODEL 1 GREEN: JUVENILE (0-17);
RED: CORE (18-59); BLUE: OLDER AGES (60-PLUS)



The percentage of variation explained by each of the first two principal components was as follows: PC1: 97%; PC2: 2%. Figure 6 shows a plot of the PC1 score. Note that the behavior represented in this plot is similar to that noted above (i.e., irregular juvenile behavior; more regular and well-ordered behavior for core ages and older ages).

**Figure 6**
PLOT OF PC1 SCORE FOR ATTAINED AGE EMBEDDING DERIVED FROM EENN MODEL 1
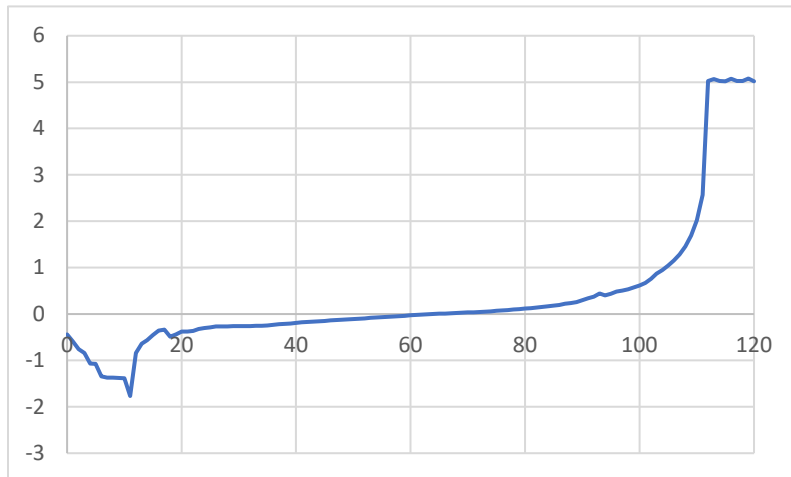


## Duration Embedding

Figure 7 shows the t-SNE plot for duration. All points are well-ordered. Note that the point labeled "ULT" represents the end of the average select period (across genders, ages, etc.).

**Figure 7**
T-SNE PLOT FOR DURATION EMBEDDING DERIVED FROM EENN MODEL 1



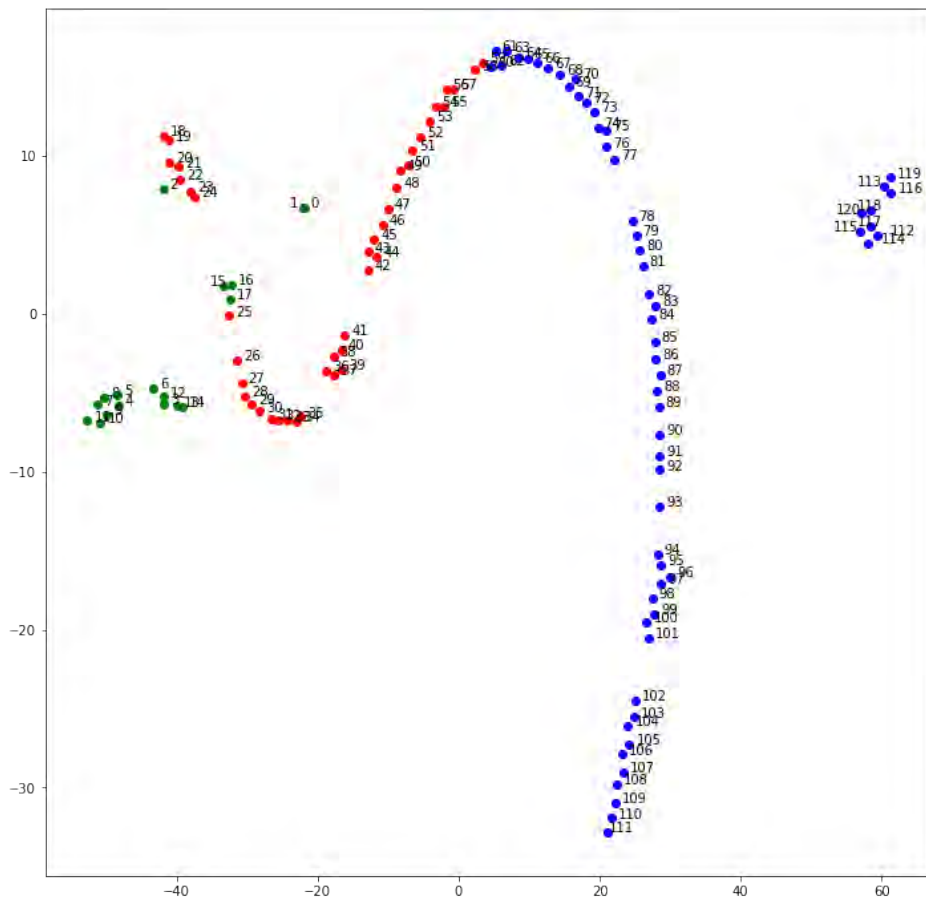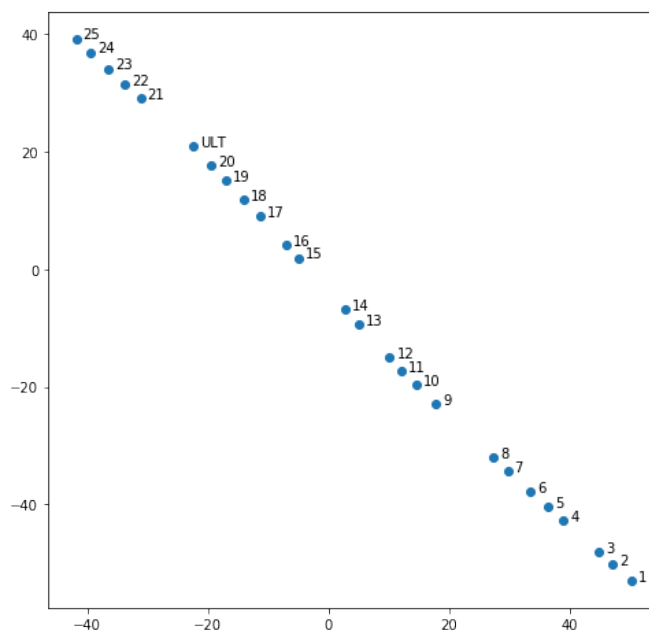The percentage of variation explained by each of the first two principal components was as follows: PC1: 43%; PC2: 30%. Since the first principal component explains less than 75% of total

variation, the author elected not to show a visualization of the PC1 score, because it paints an incomplete picture.

## 4.2 VISUALIZATION OF EMBEDDINGS FROM EENN MODEL 2

**EENN Model 2**

*$ln(q_x)'$ = EENN(risk class, attained age, duration, grouped attained age x grouped duration)*

EENN Model 2 has an interaction component in its embedding layer (grouped attained age x grouped duration). The presence of this interaction alters the interpretation of the remaining embeddings. As such, the author presented the visualization(s) for the interaction embedding ahead of those for the main effect embeddings.

It may be helpful to think of the interaction embedding as identifying an appropriate mortality neighborhood for an arbitrary individual of a particular attained age and policy duration. The main effect embeddings then adjust the location for that individual's specific risk class, attained age, etc.

### Grouped Attained Age x Grouped Duration Embedding

Table 14 shows the PC1 score organized by attained age and duration. Note that the PC1 score is generally well-ordered, in that it typically increases with attained age and duration, consistent with 2015 VBT. The percentage of variation explained by each of the first two principal components was as follows: PC1: 92%; PC2: 4%.

**TABLE 14**

PC1 SCORE FOR INTERACTION EMBEDDING DERIVED FROM EENN MODEL 2

| Att. Age | Duration 1–5 | 6–10 | 11–15 | 16–20 | 21–25 | Ult |
|---|---|---|---|---|---|---|
| 0–4 | | | | | | −1.91 |
| 5–9 | | | | | | −2.24 |
| 10–14 | | | | | | −1.96 |
| 15–19 | −0.89 | | | | | −1.30 |
| 20–24 | −0.90 | −0.78 | | | | −0.37 |
| 25–29 | −0.81 | −0.60 | −0.53 | | | −0.79 |
| 30–34 | −0.72 | −0.61 | −0.62 | −0.62 | | −1.06 |
| 35–39 | −0.46 | −0.52 | −0.46 | −0.49 | −0.44 | −0.50 |
| 40–44 | −0.38 | −0.36 | −0.37 | −0.38 | −0.44 | −0.43 |
| 45–49 | −0.46 | −0.25 | −0.24 | −0.22 | −0.23 | −0.26 |
| 50–54 | −0.28 | −0.23 | −0.14 | −0.09 | −0.06 | −0.12 |
| 55–59 | −0.19 | −0.13 | −0.07 | −0.02 | −0.07 | −0.09 |
| 60–64 | −0.17 | −0.05 | 0.01 | 0.07 | 0.04 | 0.08 |
| 65–69 | −0.09 | 0.07 | 0.17 | 0.18 | 0.21 | 0.16 |
| 70–74 | 0.12 | 0.11 | 0.22 | 0.23 | 0.31 | 0.22 |
| 75–79 | 0.21 | 0.31 | 0.33 | 0.37 | 0.43 | 0.42 |
| 80–84 | 0.35 | 0.43 | 0.47 | 0.49 | 0.54 | 0.51 |
| 85–89 | 0.35 | 0.60 | 0.55 | 0.62 | 0.74 | 0.64 |
| 90–94 | 0.74 | 0.85 | 0.90 | 0.96 | 1.04 | 0.99 |
| 95–99 | 1.02 | 0.93 | 0.99 | 1.10 | 1.20 | 1.20 |
| 100–104 | | | | | | 2.08 |
| 105–109 | | | | | | 2.89 |
| 110–114 | | | | | | 4.99 |
| 115–120 | | | | | | 9.83 |

Generally increasing with age

Generally increasing with duration

The author elected not to show the t-SNE plot for this interaction embedding because it is too complex to interpret.

## Risk Class Embedding

Figure 8 shows the t-SNE plot for risk class. As in the previous model, gender is clearly separated, UNI is more similar (closer) to NS than to SM, and MNS/FNS is more similar than MSM/FSM. These relationships are consistent with those in 2015 VBT.

**Figure 8**
T-SNE PLOT FOR RISK CLASS EMBEDDING DERIVED FROM EENN MODEL 2
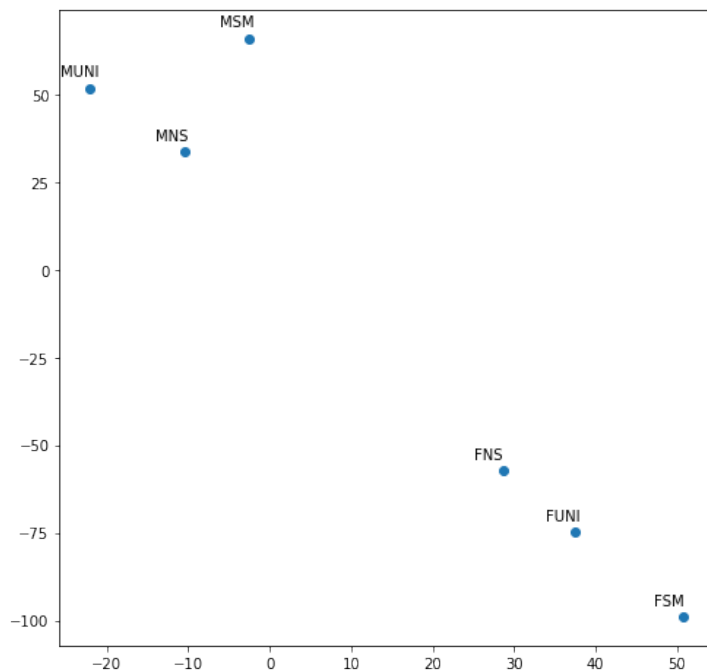


The percentage of variation explained by each of the first two principal components was as follows: PC1: 63%; PC2: 28%. Since the first principal component explains less than 75% of total variation, the author elected not to show a visualization of the PC1 score, because it paints an incomplete picture.

## Attained Age Embedding

Figure 9 shows the t-SNE plot for attained age. Note the irregular juvenile behavior, clustering of similar ages within the core and older age groups (for example, 45–49, 65–69 and 112–120), and separation between the core ages and the older ages.

**Figure 9**
T-SNE PLOT FOR ATTAINED AGE EMBEDDING DERIVED FROM EENN MODEL 2 GREEN: JUVENILE (0-17); RED: CORE (18-59); BLUE: OLDER AGES (60-PLUS)
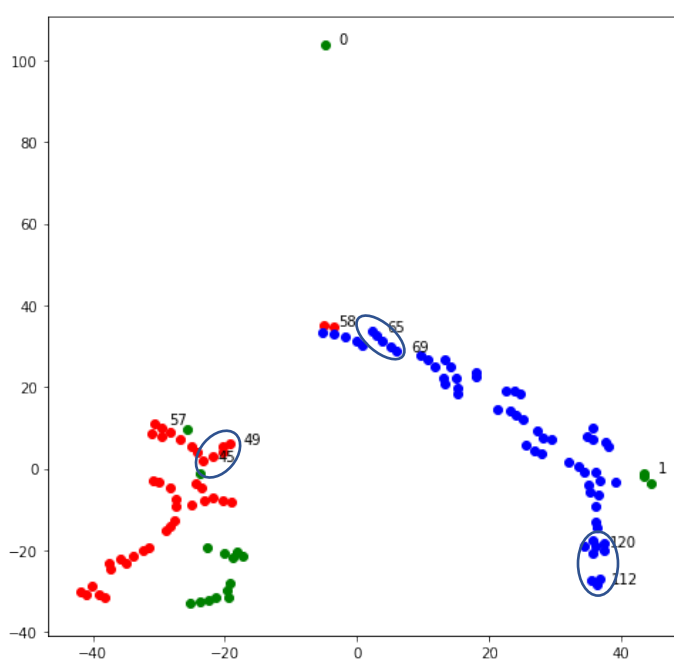


The percentage of variation explained by each of the first two principal components was as follows: PC1: 86%; PC2: 9%. Figure 10 shows a plot of the PC1 score. The saw-tooth pattern represents the embedding resetting every five years to account for the impact of the interaction embedding jumping from one quinquennial age bucket to the next. The saw-tooth pattern aside, note that the PC1 score displays a general upward trend for ages 18-plus.

**Figure 10**
PLOT OF PC1 SCORE FOR ATTAINED AGE EMBEDDING DERIVED FROM EENN MODEL 2
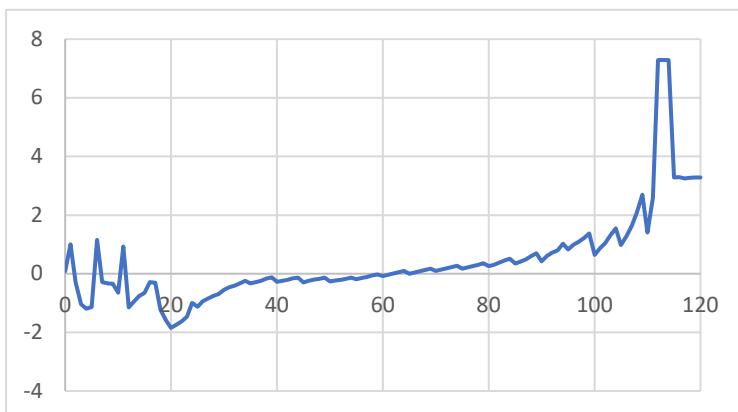


## Duration Embedding

Figure 11 shows the t-SNE plot for duration. There is evidence of clustering by quinquennial duration buckets. Note that the point labeled "Ult" represents the end of the average select period (across genders, ages, etc.).

**Figure 11**
T-SNE PLOT FOR DURATION EMBEDDING DERIVED FROM EENN MODEL 2
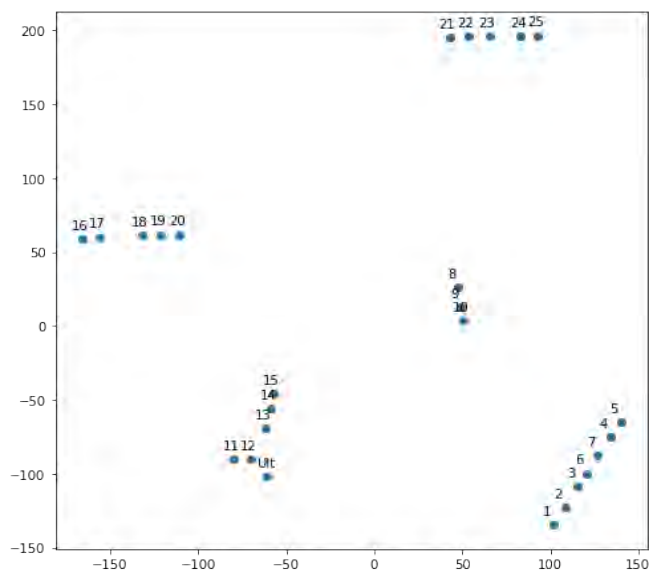


The percentage of variation explained by each of the first two principal components was as follows: PC1: 56%; PC2: 20%. Since the first principal component explains less than 75% of total variation, the author elected not to show a visualization of the PC1 score, because it paints an incomplete picture.

## 4.3 VISUALIZATION OF EMBEDDINGS FROM EENN MODEL 3

**EENN Model 3**

*ln(q$_x$)' = EENN(smoker status, attained age, gender x duration, grouped attained age x grouped duration)*

EENN Model 3 has two interaction components in its embedding layer: grouped attained age x grouped duration, and gender x duration. The presence of the first interaction alters the interpretation of the main effect embeddings for attained age and duration. As such, we present the visualization(s) for the first interaction embedding ahead of the others.

### Grouped Attained Age x Grouped Duration Embedding

Table 15 shows the PC1 score organized by attained age and duration. Note that the PC1 score is generally well-ordered, in that it typically increases with attained age and duration, consistent with 2015 VBT. The percentage of variation explained by each of the first two principal components was as follows: PC1: 91%; PC2: 6%.

**TABLE 15**

PC1 SCORE FOR AGE-DURATION INTERACTION EMBEDDING DERIVED FROM EENN MODEL 3

| | Duration | | | | | |
|---|---|---|---|---|---|---|
| Att. Age | 1–5 | 6–10 | 11–15 | 16–20 | 21–25 | Ult |
| 0–4 | | | | | | −2.40 |
| 5–9 | | | | | | −3.60 |
| 10–14 | | | | | | −3.77 |
| 15–19 | −1.96 | | | | | −1.38 |
| 20–24 | −2.09 | −1.49 | | | | −0.66 |
| 25–29 | −1.29 | −1.14 | −1.21 | | | −0.46 |
| 30–34 | −1.04 | −1.00 | −1.09 | −1.19 | | −0.74 |
| 35–39 | −0.88 | −0.86 | −0.81 | −0.88 | −0.61 | −1.11 |
| 40–44 | −0.70 | −0.67 | −0.56 | −0.65 | −0.42 | −0.96 |
| 45–49 | −0.72 | −0.42 | −0.41 | −0.42 | −0.47 | −0.79 |
| 50–54 | −0.58 | −0.39 | −0.36 | −0.38 | −0.33 | −0.48 |
| 55–59 | −0.45 | −0.19 | −0.19 | −0.20 | −0.18 | −0.13 |
| 60–64 | −0.27 | −0.10 | −0.02 | −0.03 | 0.02 | −0.05 |
| 65–69 | −0.12 | 0.18 | 0.25 | 0.23 | 0.30 | 0.22 |
| 70–74 | 0.01 | 0.29 | 0.41 | 0.44 | 0.50 | 0.44 |
| 75–79 | 0.12 | 0.33 | 0.47 | 0.55 | 0.65 | 0.57 |
| 80–84 | 0.33 | 0.63 | 0.90 | 0.87 | 0.98 | 0.90 |
| 85–89 | 0.56 | 0.82 | 1.14 | 1.11 | 1.27 | 1.17 |
| 90–94 | 0.94 | 1.30 | 1.56 | 1.49 | 1.65 | 1.57 |
| 95–99 | 1.61 | 1.71 | 1.63 | 1.56 | 1.69 | 1.65 |
| 100–104 | | | | | | 2.69 |
| 105–109 | | | | | | 3.67 |
| 110–114 | | | | | | 6.21 |
| 115–120 | | | | | | 14.07 |

Generally increasing with age

Generally increasing with duration

The author elected not to show the t-SNE plot for this interaction embedding, because it is too complex to interpret.

## Smoker Status Embedding

The percentage of variation explained by each of the first two principal components was as follows: PC1: 75%; PC2: 25%. Table 16 contains the PC1 score for each of the three smoker statuses. Note that NS < UNI < SM, and that UNI is much closer to NS than it is to SM. These relationships are consistent with 2015 VBT.

**TABLE 16**
PC1 SCORE FOR SMOKER STATUS EMBEDDING DERIVED FROM EENN MODEL 3

| NS | −0.46 |
|---|---|
| UNI | −0.32 |
| SM | 0.62 |

The author elected not to show the t-SNE plot, because it is uninformative with only three points.

## Attained Age Embedding

The percentage of variation explained by each of the first two principal components was as follows: PC1: 85%; PC2: 9%. Figure 12 shows a plot of the PC1 score. The saw-tooth pattern represents the embedding resetting every five years to account for the impact of the age-duration interaction embedding jumping from one quinquennial age bucket to the next. The saw-tooth pattern aside, note how the PC1 score displays a general upward trend for ages 18-plus.

**Figure 12**
PLOT OF PC1 SCORE FOR ATTAINED AGE EMBEDDING DERIVED FROM EENN MODEL 3



The author elected not to show the t-SNE plot for this embedding, because it is overly complex.

Gender x Duration Embedding

Figure 13 shows the t-SNE plot for gender x duration. There is evidence of clustering by quinquennial duration buckets and clear separation of males (in blue) and females (in red). Note that the points labeled "Ult" represent the end of the "average" select period (i.e., across ages) within each respective gender.

**Figure 13**

T-SNE PLOT FOR GENDER-DURATION INTERACTION EMBEDDING DERIVED FROM EENN MODEL 3
BLUE: MALES; RED: FEMALES
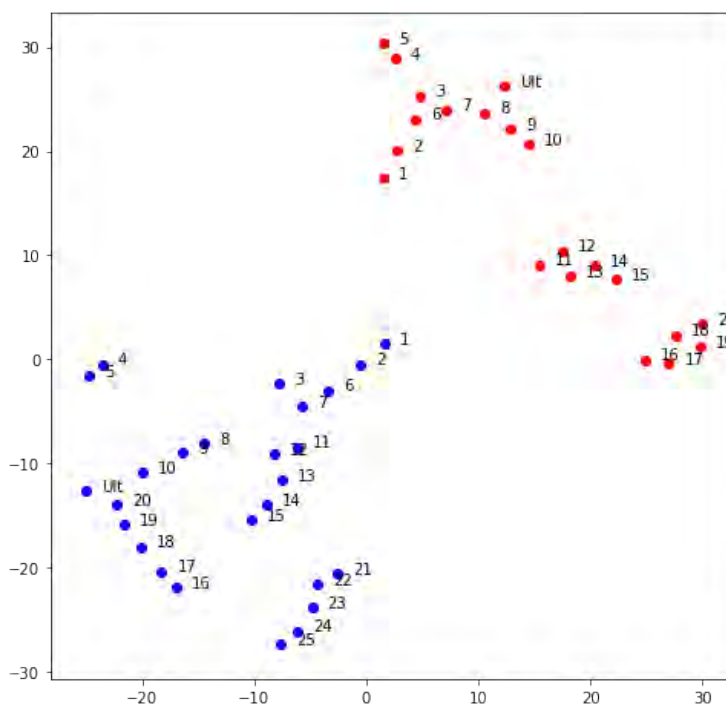


The percentage of variation explained by each of the first two principal components was as follows: PC1: 52%; PC2: 22%. Since the first principal component explains less than 75% of total variation, the author has elected not to show a visualization of the PC1 score, because it paints an incomplete picture.

### 4.4 CONSTRUCTION OF SIMPLIFIED EMBEDDINGS

A secondary use of the PCA toolkit was to generate simplified (i.e., dimensionally reduced) embeddings from the detailed embeddings. The author wanted to generate simplified embeddings because they may be more efficient in certain applications and/or predictive modeling environments.

The process used to construct the simplified embeddings was as follows:

- Where the percentage of variation explained by the first principal component was 75% or higher, the author used the PC1 score alone to represent the simplified version of the given embedding.
- Where the percentage of variation explained by the first principal component was less than 75%, the author used both the PC1 and PC2 Scores to represent the simplified version of the given embedding.

Table 17 summarizes the differences in dimensionality between the detailed and simplified embeddings.

**TABLE 17**

COMPARISON OF DIMENSIONALITY FOR DETAILED AND SIMPLIFIED EMBEDDINGS

| EENN Model | Embedding | Detailed EE Dimensionality | Simplified EE Dimensionality |
|---|---|---|---|
| 1 | Risk Class | 3 | 2 |
| | Attained Age | 5 | 1 |
| | Duration | 5 | 2 |
| 2 | Risk Class | 3 | 2 |
| | Attained Age | 5 | 1 |
| | Duration | 5 | 2 |
| | Grouped Attained Age x Grouped Duration | 5 | 1 |
| 3 | Smoker Status | 2 | 1 |
| | Attained Age | 5 | 1 |
| | Gender x Duration | 5 | 2 |
| | Grouped Attained Age x Grouped Duration | 5 | 1 |

The author has saved the simplified embedding tables associated with EENN Model 1 and EENN Model 3 to two separate embedding dictionaries (Excel workbooks) that accompany this paper. See Appendix A for further details.

# Section 5: Illustrative Case Study

In this section, the author presents an illustrative case study designed to demonstrate the value of the previously derived mortality embeddings. This case study is not intended to represent a formal experience study or rigorous table development initiative; rather, it is intended to illustrate the following features of the previously derived embeddings when calibrated to experience within a neural net environment:

- Strong fit across the full spectrum of ages and durations; and
- The ability to reflect generally accepted mortality relationships.

These features imply that one potential application of this framework is in the initial calibration of an industry mortality table to a company's own experience. It is the initial calibration, because steps like smoothing, calibrating the very ends of the table to industry experience and best practice, and monotonicity adjustments may still be required. Various factors, including the volume and coverage of training data, as well as the intended application of the output, will determine the extent to which these additional processing steps are required.

At a high level, this case study leverages U.S. insured mortality experience collected by the SOA's Individual Life Experience Committee (ILEC) for 2009–2015 and "sparsifies" these data to better reflect the gaps that a company may have in its own experience data. The author then benchmarked the out-of-sample results from neural nets fitted with the 2015 VBT mortality embeddings to comparable results from other documented models. Finally, the author projected the entire mortality table from the top-performing neural net model with embeddings for two risk classes (MNS and FNS) to demonstrate the framework's ability—within the limits of the training data—to reflect generally accepted mortality relationships.

For the purposes of this case study, the author considered the following four embedding sets:

- Detailed embeddings derived from EENN Model 1 (detailed main effect embeddings).
- Detailed embeddings derived from EENN Model 3 (detailed main and interaction effect embeddings).
- Simplified embeddings derived from EENN Model 1 (simplified main effect embeddings).
- Simplified embeddings derived from EENN Model 3 (simplified main and interaction effect embeddings).

## 5.1 CASE STUDY SPECIFICATIONS

The specifications for the illustrative case study were as follows:

- **Data:** ILEC 2009–2015 dataset.[14]

- **Data Filters:** Exclude (i) post-level term business, (ii) records from companies contributing in fewer than five of seven years, and (iii) business with issue ages above 95.

---

[14] Society of Actuaries. 2018. 2009–2015 Individual Life Insurance Mortality Experience Report & Supporting Files. *Society of Actuaries*, *https://www.soa.org/resources/research-reports/2019/2009-2015-individual-life-mortality* (accessed on Aug. 5, 2019).

- **Training/Test:** Place first four years in training and last three years in test (held out for final testing).

- **Mortality Basis:** By count.

- **Age Basis:** Age nearest (Note: For the purposes of this case study, the author combined both age nearest and age last experience without adjustment and treated the combined result as being on an age nearest basis).

- **Select Period Basis:** As defined by 2015 VBT.

- **Mortality Improvement:** Not explicitly modeled.

- **Smoker Status Adjustments:** The reported smoker status was only considered reliable for records with (i) issue year from 1980 onward and (ii) issue ages 18-plus; in all other cases, the smoker status was assumed to be UNI (unknown).

- **Performance Metrics:** A/E and weighted MAPE at a suitably aggregated level.[15]

- **Calibration Framework:** Neural nets.[16]

The principles the author used to select the neural net hyperparameters for use within the case study are as follows:

- Instead of using cross-validation, the author leveraged all available insight from the EENNs parameterized in Section 3. Those EENNs were set up for a similar type of problem (insured mortality modeling).
- The author structured the neural nets to be less complex (i.e., have fewer hidden nodes) than their closest analog in Section 3. The author wanted relatively constrained models to ensure that they leveraged the architecture encoded in the embeddings.
- The author kept the total number of trainable parameters (weights) in the neural nets to fewer than one-quarter the number of training records. That principle is analogous to EENN Design Principle 3 in subsection 3.5.

Details on the selected hyperparameters can be found in Appendix B. Note that these hyperparameters were selected for illustrative purposes only and, as such, have not been optimized.

---

[15] See the tables in Appendix C for an example of the level of aggregation; all weightings are by exposure count.

[16] In this case study, the author calibrated the mortality embeddings to actual experience using neural nets. Other nonlinear possibilities would have included k-nearest neighbor regression, gradient boosted regression trees, and random forest regression. The use of these alternative nonlinear models for calibration is a possible area for further research. The author did not find the embeddings to be effective in logistic regression models and suspected that the same would be true for other linear predictive models.

For the purposes of benchmarking fit, the author selected two main effects logistic regression models, as well as a neural net trained on the set of traditional features alone. The selection of the two main effects logistic regression models was based on the published work of Zhu et al. (2015). In their paper, they write: "The [main effects] models may be enhanced with various techniques such as adding interactions or applying variable transformations. However, our tests reveal that only minimal model performance improvement would be achieved for our [insured mortality] study data." On the basis of their detailed work with main effects logistic regression for insured mortality and the above commentary, the author decided to include logistic regression models parameterized exclusively with main effects for illustrative benchmarking purposes.

## 5.2 CASE STUDY DATA SUMMARY

This subsection provides more detail on the case study data itself, including the changes in the data as a result of the filters applied and the nature of the final training and test sets. Note that, unlike in a formal experience study, the author has used these data without detailed review and validation.

Table 18 shows the impact of the data exclusions on record count, death count and exposure count. In total, the data filters resulted in 62,154 (1.8%) death count exclusions and 10,857,483 (3.1%) exposure count exclusions.

**TABLE 18**
DATA EXCLUSIONS

| Step | Description | Observed Change in Record Count | Record Count at Step End | Observed Change in Death Count | Death Count at Step End | Observed Change in Exposure Count | Exposure Count at Step End |
|---|---|---|---|---|---|---|---|
| 1 | Import ILEC 2009–2015 dataset (7 calendar years) | 30,631,099 | 30,631,099 | 3,443,319 | 3,443,319 | 352,500,854 | 352,500,854 |
| 2 | Remove post-level term business | −1,929,331 | 28,701,768 | −35,227 | 3,408,092 | -6,940,009 | 345,560,845 |
| 3 | Remove records from companies contributing in fewer than five of seven years | −2,158,112 | 26,543,656 | -26,735 | 3,381,357 | -3,916,950 | 341,643,895 |
| 4 | Remove records with issue age above 95 | −415 | **26,543,241** | −192 | **3,381,165** | −524 | **341,643,371** |

Table 19 shows the characteristics of the training and test sets, after all data exclusions but prior to sparsification.

**TABLE 19**
CHARACTERISTICS OF TRAINING AND TEST SETS PRIOR TO SPARSIFICATION

| Dataset | Number of Mortality Cells | Death Count | Exposure Count |
|---|---|---|---|
| Training | 9,024 | 1,741,402 | 174,287,796.1 |
| Test | 9,037 | 1,639,763 | 167,355,574.5 |
| Total | N/A | 3,381,165 | 341,643,371 |

Table 20 shows the characteristics of the training and test sets following sparsification. For the purposes of this case study, the author sparsified the training set by deleting cells with fewer than 25 deaths. The intent of sparsification was to simulate the holes that an individual company may have in its own experience data. The author reasoned that by removing cells found to have limited credibility within the industry dataset designated for training, he could more realistically simulate the holes likely to be present in a typical company's experience. The other significant benefit of using a sparsified training set was that it enabled the author to get a more realistic picture of each model's true out-of-sample performance, by forcing the model to predict mortality rates for previously unseen mortality cells. This latter ability is key to mortality table modeling.

**TABLE 20**
CHARACTERISTICS OF TRAINING AND TEST SETS AFTER SPARSIFICATION

| Dataset | Number of Mortality Cells | Death Count | Exposure Count |
|---|---|---|---|
| Training | 3,674 | 1,704,519 | 147,426,516.3 |
| Test | 9,037 | 1,639,763 | 167,355,574.5 |
| Total | N/A | 3,344,282 | 314,782,091 |

### 5.3  CASE STUDY RESULTS FOR FIT

In this subsection, the author begins by presenting a summary of the overall out-of-sample performance, based on weighted MAPE, for each of the models under consideration. Recall that weighted MAPE is calculated at a grouped cell level and weighted by exposure count. The author then detailed A/E results for the top-performing neural net with mortality embeddings, as well as for the top -performing main effects logistic regression model.

Weighted MAPE

Table 21 presents a summary of the out-of-sample (test set) performance for each of the seven models in scope for the case study. Models A through C represent benchmark models, while Models D through G represent models with embeddings. All models were trained on the sparsified training set, with Models C through G using exposure count as their training sample weights. The author calculated the weighted MAPE metric for each model on the complete test set.

**TABLE 21**
ILLUSTRATIVE OUT-OF-SAMPLE WEIGHTED MAPE RESULTS

| Model ID | Description | Weighted MAPE |
|---|---|---|
| A | Logistic regression with traditional features, main effects, full model | 23.3% |
| B | Logistic regression with traditional features, main effects, subset models | 24.4% |
| C | Neural net with traditional features, main effects | 12.1% |
| D | Neural net with detailed main effect embeddings | 7.2% |
| E | Neural net with detailed main and interaction effect embeddings | 6.5% |
| F | Neural net with simplified main effect embeddings | 8.1% |
| G | Neural net with simplified main and interaction effect embeddings | 11.1% |

Based on the results in Table 21, the author made the following observations:

- Models D through G performed better than benchmark Models A through C, confirming the effectiveness of embeddings (i.e., by leveraging the structure from an industry table, the predictive accuracy was improved relative to predictive models with traditional features).
- Model E performed the best overall. Detailed A/E results for Model E are presented below.
- Of the main effects logistic regression models, Model A performed the best. Detailed A/E results for Model A are presented below.
- Model F performed similarly to Model D, suggesting that the detailed set of main effect embeddings may be overspecified (i.e., have unnecessary dimensions).
- Model G performed worse than Model E, suggesting that the simplified set of main and interaction effect embeddings may be missing key information present in the detailed set.

In addition to the results in Table 21, the author estimated the MAPE of 2015 VBT on the test set to be approximately 13%. The VBT results serve as another point of validation and act as an objective benchmark against which the calibrated results (Models D–G) can be judged.

Since the results for Models D–G were lower than this 13% baseline, the author was able to confirm that calibration of the embedding architecture to the case study data had indeed taken place. Had the embeddings or their implementation been ineffective, the author would have expected to see performance results no better than 2015 VBT, on which the embeddings themselves are based.[17]

<u>A/E Results for Model A</u>

Table 22 shows the select period A/Es for Model A, the main effects logistic regression model fitted with all four traditional features. Note the challenged fit throughout the table; only 11 grouped cells (highlighted in green) have an absolute error under 10%.

---

[17] Note that the calibration process had to bridge differences in age basis, count vs. amount basis, and time period, among others. These differences exist between the data used to develop 2015 VBT (and hence the embeddings) and the data used in the case study.

**TABLE 22**

SELECT PERIOD A/ES (BY COUNT) FOR MODEL A (GROUPED CELLS WITH AN ABSOLUTE ERROR UNDER 10% HIGHLIGHTED IN GREEN)

| Risk Class | Issue Age | Duration 1–5 | 6–10 | 11–15 | 16–20 | 21–25 | All |
|---|---|---|---|---|---|---|---|
| MNS | 18–39 | 178.8% | 119.3% | 97.5% | 99.6% | 88.0% | 101.0% |
| | 40–59 | 88.8% | 83.6% | 80.9% | 85.0% | 87.0% | 84.8% |
| | 60+ | 84.8% | 99.9% | 112.4% | 121.1% | 116.3% | 103.8% |
| | All | 93.6% | 93.6% | 92.6% | 95.8% | 87.8% | 92.4% |
| MSM | 18–39 | 242.4% | 145.6% | 133.7% | 128.2% | 108.5% | 125.6% |
| | 40–59 | 130.8% | 127.6% | 125.2% | 116.9% | 101.8% | 113.6% |
| | 60+ | 132.0% | 131.8% | 115.7% | 104.1% | 103.0% | 118.4% |
| | All | 145.8% | 131.2% | 124.3% | 116.7% | 103.7% | 117.0% |
| MUNI | 18–39 | 1269.9% | 717.9% | 332.2% | 350.4% | 232.5% | 296.1% |
| | 40–59 | 382.0% | 385.3% | 299.6% | 253.4% | 152.9% | 201.0% |
| | 60+ | 197.8% | 273.5% | 196.4% | 147.1% | 121.2% | 154.8% |
| | All | 333.2% | 309.1% | 229.4% | 185.5% | 144.6% | 180.4% |
| FNS | 18–39 | 132.4% | 113.3% | 100.6% | 103.9% | | 107.7% |
| | 40–59 | 85.2% | 86.0% | 81.6% | 82.2% | | 83.3% |
| | 60+ | 79.2% | 103.9% | 115.1% | 104.4% | | 100.9% |
| | All | 85.7% | 97.3% | 96.8% | 90.6% | | 93.4% |
| FSM | 18–39 | 175.2% | 157.7% | 132.1% | 119.8% | | 131.9% |
| | 40–59 | 138.5% | 131.2% | 125.4% | 117.2% | | 123.7% |
| | 60+ | 125.8% | 131.3% | 120.8% | 113.2% | | 122.3% |
| | All | 136.0% | 133.8% | 124.6% | 116.9% | | 124.4% |
| FUNI | 18–39 | 455.2% | 450.7% | 240.8% | 321.8% | | 316.3% |
| | 40–59 | 201.4% | 357.2% | 281.6% | 198.5% | | 229.9% |
| | 60+ | 161.8% | 223.2% | 168.7% | 134.8% | | 152.0% |
| | All | 194.2% | 241.9% | 184.5% | 146.6% | | 165.5% |

Table 23 shows the ultimate period A/Es for Model A. While perhaps better overall than the results for the select period, these ultimate period results show a pronounced curvature (higher A/Es at the ends).

**TABLE 23**

ULTIMATE PERIOD A/ES (BY COUNT) FOR MODEL A (GROUPED CELLS WITH AN ABSOLUTE ERROR UNDER 10% HIGHLIGHTED IN GREEN)

| Risk Class | Attained Age | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0–17 | 18–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90+ | All |
| MNS | | | | 144.3 | 97.2% | 91.8% | 101.3% | 124.2% | 139.0% | 113.3% |
| MSM | | | | 127.7% | 119.0% | 100.5% | 98.3% | 92.5% | 76.8% | 96.9% |
| MUNI | 351.8% | 532.0% | 273.9% | 176.8% | 127.9% | 92.1% | 89.4% | 101.2% | 107.2% | 101.1% |
| FNS | | | 164.2% | 107.0% | 82.7% | 75.1% | 83.8% | 111.2% | 138.6% | 101.9% |
| FSM | | | 100.8% | 109.7% | 98.5% | 94.3% | 97.5% | 88.2% | 81.5% | 91.9% |
| FUNI | 286.4% | 253.5% | 175.3% | 138.9% | 102.4% | 80.3% | 86.7% | 101.1% | 112.4% | 100.3% |
| All | 323.0% | 411.8% | 231.8% | 147.4% | 107.0% | 88.0% | 90.1% | 103.5% | 111.8% | 101.6% |

A/E Results for Model E

This section presents analogous A/E results for Model E, the neural net fitted with the detailed set of main and interaction effect embeddings. In Tables 24 and 25, note not only the significantly improved relative fit over Model A but also the strong absolute fit. Recall that this model (as well as the others) was trained on data from only 3,674 cells from one time period and subsequently asked to predict mortality rates for 9,037 cells from a different time period.

**TABLE 24**

SELECT PERIOD A/ES (BY COUNT) FOR MODEL E (GROUPED CELLS WITH AN ABSOLUTE ERROR UNDER 10% HIGHLIGHTED IN GREEN)

| Risk Class | Issue Age | Duration | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | 1–5 | 6–10 | 11–15 | 16–20 | 21–25 | All |
| MNS | 18–39 | 106.3% | 100.4% | 93.9% | 97.8% | 97.7% | 98.2% |
| | 40–59 | 104.7% | 96.9% | 90.8% | 89.7% | 97.3% | 94.7% |
| | 60+ | 102.8% | 99.1% | 93.3% | 105.9% | 93.4% | 99.2% |
| | All | 104.1% | 98.3% | 92.1% | 95.3% | 97.3% | 96.6% |
| MSM | 18–39 | 74.6% | 62.6% | 83.7% | 99.5% | 96.9% | 88.9% |
| | 40–59 | 70.6% | 94.2% | 101.0% | 105.0% | 101.5% | 98.2% |
| | 60+ | 136.5% | 112.8% | 96.5% | 102.3% | 100.2% | 108.6% |
| | All | 85.2% | 91.7% | 96.4% | 103.2% | 100.1% | 97.3% |
| MUNI | 18–39 | 143.9% | 316.4% | 204.1% | 228.6% | 144.4% | 169.4% |
| | 40–59 | 298.0% | 284.9% | 231.2% | 223.6% | 166.6% | 196.3% |
| | 60+ | 209.9% | 220.3% | 160.9% | 141.8% | 114.8% | 142.0% |
| | All | 196.7% | 238.7% | 181.8% | 170.5% | 137.2% | 161.5% |
| FNS | 18–39 | 83.2% | 90.1% | 96.4% | 104.7% | | 95.9% |
| | 40–59 | 111.8% | 96.9% | 93.9% | 96.7% | | 97.7% |
| | 60+ | 97.4% | 94.4% | 95.3% | 104.2% | | 96.4% |
| | All | 100.0% | 94.8% | 94.9% | 100.0% | | 96.9% |
| FSM | 18–39 | 69.5% | 71.5% | 75.6% | 94.5% | | 82.2% |
| | 40–59 | 84.6% | 89.5% | 99.7% | 109.1% | | 99.7% |
| | 60+ | 129.9% | 105.4% | 105.4% | 111.5% | | 110.6% |
| | All | 96.3% | 92.3% | 97.5% | 106.4% | | 99.5% |
| FUNI | 18–39 | 100.5% | 208.0% | 135.0% | 216.4% | | 170.6% |
| | 40–59 | 200.6% | 284.6% | 261.2% | 194.3% | | 218.8% |
| | 60+ | 146.7% | 169.9% | 128.6% | 123.3% | | 129.6% |
| | All | 145.0% | 183.4% | 143.1% | 134.8% | | 142.0% |

The poor overall fit to MUNI and FUNI experience in the select period was likely driven by data limitations (data credibility and data quality), given that the corresponding results when E = 2015 VBT were found to be similarly poor.

**TABLE 25**

ULTIMATE PERIOD A/ES (BY COUNT) FOR MODEL E (GROUPED CELLS WITH AN ABSOLUTE ERROR UNDER 10% HIGHLIGHTED IN GREEN)

| Risk Class | Attained Age | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0–17 | 18–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90+ | All |
| MNS | | | | 103.5% | 104.6% | 101.3% | 100.8% | 103.8% | 111.5% | 103.9% |
| MSM | | | | 86.9% | 117.0% | 100.5% | 92.3% | 94.8% | 80.8% | 95.7% |
| MUNI | 95.0% | 102.3% | 111.9% | 97.5% | 106.6% | 98.0% | 97.8% | 99.2% | 101.0% | 99.5% |
| FNS | | | 116.6% | 105.0% | 107.8% | 97.5% | 99.9% | 99.9% | 109.6% | 102.1% |
| FSM | | | 63.1% | 95.2% | 101.5% | 98.2% | 97.7% | 90.5% | 84.1% | 93.8% |
| FUNI | 86.4% | 104.5% | 103.7% | 101.4% | 105.9% | 92.9% | 97.3% | 99.5% | 96.9% | 98.1% |
| All | 91.5% | 102.8% | 109.1% | 99.1% | 106.7% | 97.6% | 98.0% | 99.5% | 100.5% | 99.5% |

To get a better sense of the data credibility by grouped cell in the test set, refer to the distribution of deaths in Appendix C.

## 5.4  CASE STUDY RESULTS FOR MAINTENANCE OF RELATIONSHIPS

Appendix D contains the projected mortality tables for MNS and FNS from Model E fit to the sparsified training dataset. In practice, one would typically reintroduce any held-out data into training prior to such a projection; however, for consistency with results from previous parts of the case study, the author has elected not to do so.

Based on these projections, the author observed that Model E more often than not reflected the following generally accepted mortality table relationships encoded in 2015 VBT:

- For a given duration, mortality rates should not decrease with issue age.[18]
- For a given issue age, mortality rates should not decrease with duration.
- Mortality rates should not decrease with attained age.[19]
- For a given issue age and duration in the select period or a given attained age in the ultimate period, female mortality should not be greater than male mortality.

## 5.5  CASE STUDY RECAP

In this case study, the author has demonstrated that mortality embeddings derived from 2015 VBT, when calibrated to a sparse experience dataset within a neural net environment, can result in mortality projections that (1) provide a strong relative and absolute fit across the full spectrum of ages and durations and (2) more often than not reflect the generally accepted insured mortality relationships encoded in the industry table.

These features imply that one potential application of this framework is in the initial calibration of an industry mortality table to a company's own experience. Recall that it is initial calibration because steps like smoothing, calibrating the very ends of the table to industry experience and

---

[18] Only applicable for attained ages 30-plus.
[19] Only applicable for attained ages 30-plus.

best practice, and monotonicity adjustments may still be required. Various factors, including the volume and coverage of training data, as well as the intended application of the output, will determine the extent to which these additional processing steps are required.

## Section 6: Conclusion

The use of predictive models to develop best estimate insured mortality rates as an alternative to the data- and resource-intensive traditional first principles table construction approach has historically been met with limited success. One of the main challenges faced by companies looking to leverage the predictive modeling approach is that it has no preconceived notion of what a mortality table should look like, leading to issues in fit where experience is sparse and to a lack of appreciation for generally accepted mortality relationships.

In this paper, the author has introduced a novel framework for leveraging the architecture of an industry mortality table within a company's predictive analytics-based insured mortality analysis. He has shown that by reverse-engineering the industry mortality table into a series of higher-dimensional features and then using those features as inputs to a nonlinear predictive model (in this case a neural net), the author can better model generally accepted mortality relationships (1) within a predictive modeling paradigm and (2) under the constraint of sparse experience data. One promising application of this approach is in the initial calibration of an industry mortality table, via its learned features, to a company's own experience.

Recognizing that readers may be coming at this content from different perspectives and backgrounds, the author wishes to emphasize that the material in this paper can be leveraged on at least three different levels. The first level is conceptual—readers could take away the key ideas in the paper, including the notion of being able to reverse-engineer a mortality basis and reuse that structure elsewhere. The second level is application focused—readers could potentially leverage the mortality embeddings accompanying this paper in their nonlinear mortality modeling work. Instructions on how to append the accompanying embeddings to their own predictive modeling dataset can be found on the first tab of each embedding dictionary's workbook. The third level is development centric—readers could generate new embeddings for alternative mortality bases using the machinery discussed herein. The code used to develop the mortality embeddings for 2015 VBT accompanies this paper, and knowledgeable Python users could modify it to suit their needs.

# Appendix A: Inventory of Accompanying Files

## Base Data Files

| # | Filename | Description |
|---|----------|-------------|
| 1 | 2015-vbt-smoker-distinct-alb-anb.xlsx | 2015 VBT smoker distinct mortality rates |
| 2 | 2015-vbt-unismoke-alb-anb.xlsx | 2015 VBT unismoke mortality rates |
| 3 | 2015_VBT_select_period_table.csv | 2015 VBT select period lengths by issue age, gender and smoker status |

## Additional Input Files Required for Visualization of Embeddings

| # | Filename | Description |
|---|----------|-------------|
| 1 | Model 1 Embeddings.xlsx | Detailed Embeddings from EENN Model 1 |
| 2 | Model 2 Embeddings.xlsx | Detailed Embeddings from EENN Model 2 |
| 3 | Model 3 Embeddings.xlsx | Detailed Embeddings from EENN Model 3 |

## Additional Input Files Required for Case Study

| # | Folder / Filename | Description |
|---|-------------------|-------------|
| 1 | case_study_train.txt | Case study data—training set |
| 2 | case_study_test.txt | Case study data—test set |
| 3 | Detailed Embeddings - Main Effects Only | Folders containing embeddings in long form (i.e., in a form suitable for appending) |
| 4 | Simplified Embeddings - Main Effects Only | |
| 5 | Detailed Embeddings - Main & Interaction Effects | |
| 6 | Simplified Embeddings - Main & Interaction Effects | |
| 7 | 2015_VBT_df.csv | 2015 VBT in dataframe format |

## Code Files

| # | Filename | Description |
|---|----------|-------------|
| 1 | Code to Generate 2015 VBT Mortality Embeddings.ipynb | Code corresponding to Sections 2 and 3 of this document |
| 2 | Code to Visualize Results for EENN Model 1.ipynb | Code corresponding to Section 4 of this document |
| 3 | Code to Visualize Results for EENN Model 2.ipynb | |
| 4 | Code to Visualize Results for EENN Model 3.ipynb | |
| 5 | Code for Case Study.ipynb | Code corresponding to Section 5 of this document |

## Embedding Dictionaries

| # | Filename | Description |
|---|----------|-------------|
| 1 | Dictionary of Detailed Embeddings - Main Effects Only.xlsx | Detailed Embeddings from EENN Model 1 |
| 2 | Dictionary of Simplified Embeddings - Main Effects Only.xlsx | Simplified Embeddings from EENN Model 1 |
| 3 | Dictionary of Detailed Embeddings - Main & Interaction Effects.xlsx | Detailed Embeddings from EENN Model 3 |
| 4 | Dictionary of Simplified Embeddings - Main & Interaction Effects | Simplified Embeddings from EENN Model 3 |

# Appendix B: Neural Net Hyperparameters for Case Study

TABLE B.1: Neural Net Hyperparameters

| Model ID | # Epochs | Learning Rate | # Nodes | | | # Trainable Weights[20] |
| | | | Input Layer | First Hidden Layer | Second Hidden Layer | |
|---|---|---|---|---|---|---|
| C | 500 | 0.001 | 6 | 25 | 6 | 338 |
| D | 500 | 0.001 | 13 | 25 | 6 | 513 |
| E | 2000 | 0.001 | 17 | 10 | 4 | 229 |
| F | 500 | 0.001 | 5 | 25 | 6 | 313 |
| G | 2000 | 0.001 | 5 | 10 | 4 | 109 |

Others:

All Neural Net models have one output node.

---

[20] This is a consequence of the other hyperparameters, not a hyperparameter itself; this is shown for information purposes.

# Appendix C: Distribution of Death Counts in Case Study Test Data

TABLE C.1: Actual Deaths—Select Period

| Risk Class | Issue Age | Duration | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1–5 | 6–10 | 11–15 | 16–20 | 21–25 | All |
| MNS | 18–39 | 2,628 | 3,209 | 4,518 | 5,829 | 9,492 | 25,676 |
| | 40–59 | 7,528 | 12,382 | 16,834 | 16,488 | 22,810 | 76,042 |
| | 60+ | 8,096 | 12,264 | 12,508 | 8,981 | 652 | 42,501 |
| | All | 18,252 | 27,855 | 33,860 | 31,298 | 32,954 | 144,219 |
| MSM | 18–39 | 773 | 715 | 1,177 | 2,013 | 3,712 | 8,390 |
| | 40–59 | 1,612 | 2,640 | 3,817 | 5,319 | 8,623 | 22,011 |
| | 60+ | 1,179 | 1,411 | 1,504 | 1,561 | 114 | 5,769 |
| | All | 3,564 | 4,766 | 6,498 | 8,893 | 12,449 | 36,170 |
| MUNI | 18–39 | 119 | 91 | 180 | 326 | 786 | 1,502 |
| | 40–59 | 101 | 302 | 723 | 1,182 | 2,146 | 4,454 |
| | 60+ | 147 | 857 | 1,342 | 1,799 | 2,142 | 6,287 |
| | All | 367 | 1,250 | 2,245 | 3,307 | 5,074 | 12,243 |
| FNS | 18–39 | 1,460 | 2,208 | 3,263 | 4,139 | | 11,070 |
| | 40–59 | 4,319 | 6,899 | 9,210 | 9,494 | | 29,922 |
| | 60+ | 5,913 | 9,521 | 10,007 | 3,258 | | 28,699 |
| | All | 11,692 | 18,628 | 22,480 | 16,891 | | 69,691 |
| FSM | 18–39 | 214 | 321 | 556 | 1,047 | | 2,138 |
| | 40–59 | 857 | 1,377 | 2,199 | 3,587 | | 8,020 |
| | 60+ | 780 | 1,095 | 1,457 | 968 | | 4,300 |
| | All | 1,851 | 2,793 | 4,212 | 5,602 | | 14,458 |
| FUNI | 18–39 | 39 | 46 | 93 | 205 | | 383 |
| | 40–59 | 57 | 285 | 876 | 1,613 | | 2,831 |
| | 60+ | 122 | 1,353 | 3,471 | 6,104 | | 11,050 |
| | All | 218 | 1,684 | 4,440 | 7,922 | | 14,264 |

TABLE C.2: Actual Deaths—Ultimate Period

| Risk Class | Attained Age | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0–17 | 18–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90+ | All |
| MNS | | | | 822 | 7,647 | 18,500 | 28,296 | 58,950 | 19,456 | 133,671 |
| MSM | | | | 464 | 3,216 | 7,984 | 10,781 | 11,469 | 2,338 | 36,252 |
| MUNI | 847 | 4,482 | 4,295 | 6,107 | 19,172 | 55,819 | 139,671 | 288,904 | 167,086 | 686,383 |
| FNS | | | 17 | 1,466 | 7,190 | 13,923 | 23,444 | 51,922 | 27,542 | 125,504 |
| FSM | | | 8 | 559 | 2,765 | 6,819 | 11,618 | 14,030 | 4,139 | 39,938 |
| FUNI | 542 | 1,621 | 1,995 | 2,973 | 7,957 | 23,386 | 55,519 | 138,183 | 94,794 | 326,970 |
| All | 1,389 | 6,103 | 6,315 | 12,391 | 47,947 | 126,431 | 269,329 | 563,458 | 315,355 | 1,348,718 |

## Appendix D: Sample Mortality Tables Projected from Model E

This appendix contains sample mortality tables, as projected from Model E (calibrated to the sparsified training set), for the following risk classes:

- MNS
- FNS

For tables see Appendix D Tables.xlsx

## References

Abbott, D. 2014. *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst.* [Indianapolis, IN]: John Wiley & Sons, Inc.

Guo, C., and F. Berkhahn. 2016. Entity Embeddings of Categorical Variables. *arXiv*: 1604.06737v1.

Nisbet, R., J. Elder, and G. Miner. 2009. *Handbook of Statistical Analysis and Data Mining Applications.* [Burlington, MA]: Academic Press.

Society of Actuaries. 2018. 2009–2015 Individual Life Insurance Mortality Experience Report & Supporting Files. *Society of Actuaries*, *https://www.soa.org/resources/research-reports/2019/2009-2015-individual-life-mortality* (accessed on Aug. 5, 2019).

Society of Actuaries. 2018. 2015 Valuation Basic Table Report & Tables. *Society of Actuaries, https://www.soa.org/experience-studies/2015/2015-valuation-basic-tables* (accessed on Aug. 5, 2019).

van der Maaten, L., and G. Hinton. 2008. Visualizing Data Using t-SNE. *Journal of Machine Learning Research* 9:2579–2605.

Zhu, Z., Z. Li, D. Wylde, M. Failor, and G. Hrischenko. 2015. Logistic Regression for Insured Mortality Experience Studies. *North American Actuarial Journal* 19 (4):241–255.

## About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world dedicated to serving more than 32,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

**Objectivity:** The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

**Quality:** The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

**Relevance:** The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

**Quantification:** The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.