Table Development

SOCIETY OF ACTUARIES

Table Development

AUTHOR	David B. Atkinson, FSA Society of Actuaries	**SPONSOR**	Experience Studies Executive Committee

REVIEWERS	Mary J. Bahna-Nolan, FSA Robert W. Beal, FSA Thomas P. Edwalds, FSA Mervyn Kopinsky, FSA Cynthia MacDonald, FSA Patrick D. Nolan, FSA C. Allen Pinkham, ASA Laurence Pinzur, FSA Marianne C. Purushotham, FSA Thomas E. Rhodes, FSA Bradley C. Spenney FSA

Caveat and Disclaimer

TABLE OF CONTENTS

Section 1: Overview

1.1 Scope

This paper documents the process of creating experienced-based tables for mortality rates, lapse rates, incidence rates, severity rates, and much more. The scope includes all experience-based rates used to predict outcomes for life, health, annuity, and pension products. The processes described in this paper can be applied to an internal study with data from a single contributor or to an industry study with data from scores of contributors.

As much as possible, this paper attempts to use generic terminology that can be understood across different products or lines of business. In instances where no common or generic terminology makes sense, the paper will attempt to note those differences.

It is hoped that this paper will serve as a reference manual for the development of future tables through documenting common and alternative practices and sharing ideas that may improve productivity or quality.

1.2 Project Planning and Oversight

The time needed to develop a table can range from hours to years, depending on the table's importance, the data available and the complexity of the underlying rates. Even for the simplest table development project, it is worthwhile to list the major steps and think through the time and resources required for each step.

For larger projects, it is common to establish upfront agreement on project oversight and periodic reviews. The bigger and more important the project, the more up-front time you will want to spend thinking through the various interdependencies, challenges, potential resource shortfalls and potential obstacles that could derail or delay the project.

To the extent possible, you'll want to think through how specific potential problems or challenges could be overcome. If you have identified any specialized resources or knowledge needed, you'll want to address how and when to acquire those missing pieces before you finalize your plan. These planning steps should result in a more realistic project plan that makes allowances for unexpected issues along the way.

Review and oversight is especially important as the table begins to take shape. As likely table dimensions emerge, they should be discussed and compared to existing tables. Implications of new dimensions (i.e., dimensions not included in any predecessor table) should be carefully considered. As rates are developed, they should be compared to rates from predecessor tables. Significant differences in rates should be reconciled and discussed and may need to have their financial impact assessed before proceeding. An important goal of oversight is to identify potential dead-ends as early in the process as possible, to minimize rework.

1.3 Table Development Flow Chart

The following chart shows how the major steps in developing a new table are connected:

```
┌──────────┐    ┌──────────────┐    ┌──────────────────┐    ┌─────────────┐
│ Develop  │ →  │ Identify     │ →  │ Populate table   │ →  │ Review and  │
│ data     │    │ table        │    │ through          │    │ adjust rates│
│          │    │ dimensions   │    │ graduation and   │    │             │
│          │    │ through      │    │ interpolation or │    └─────────────┘
│          │    │ exploration  │    │ modeling;        │        ↓    ↑
│          │    │ techniques   │    │ extend and       │    ┌─────────────┐
└──────────┘    └──────────────┘    │ project rates as │    │ Assess      │
                                     │ needed           │    │ financial   │
                                     └──────────────────┘    │ impact and  │
                                                             │ finalize    │
                                                             │ table       │
                                                             └─────────────┘
```
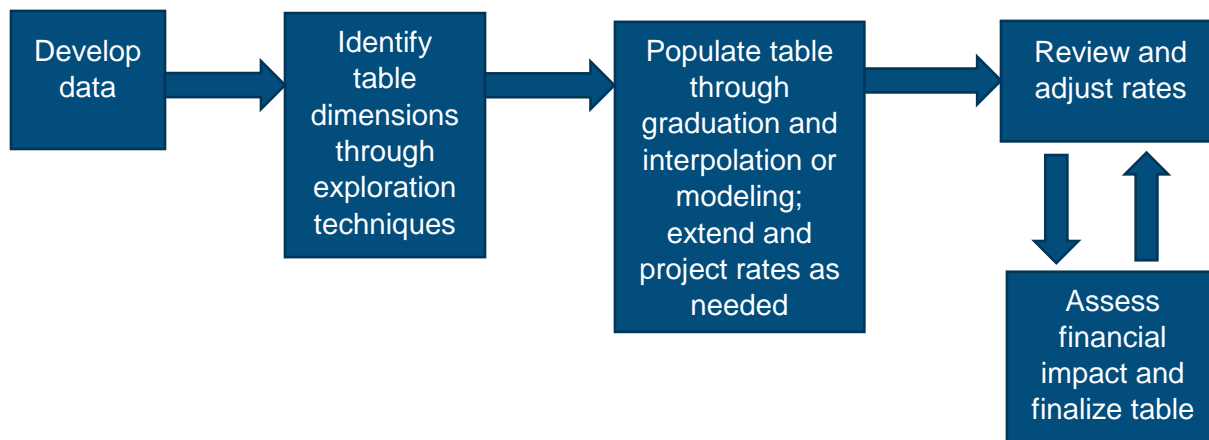
1.4 Future Table Development

We expect table development to evolve as new tools and new practices are brought to bear. We expect modeling to improve and replace some facets of the table development process. In some cases, tables could be entirely replaced by models that supply rates on an as-needed basis.

The SOA Experience Studies Executive Committee intends to sponsor periodic updates to this document to keep pace with the evolution of table development. To help keep this paper relevant, please e-mail your observations and insights to the SOA Research Department at Research@SOA.org.

Section 2: Data Development

2.1 Preliminaries

Many table development projects start with a review of one or more predecessor studies. Such reviews may influence the new table's dimensions and the source of data to be used. When consistency with a prior study is paramount, the prior study may dictate the table dimensions and source of data. If there are no predecessor studies, an early step will be to identify potential sources of data. Once data has been obtained, data analysis, described in Chapter 3, will determine the table's dimensions.

2.1.1 Clarify Purpose and Goals

It is important to understand how the table will be used, as well as the financial effect of deviations between future experience and the experience predicted by the table. Goals should be established to fulfill the table's purpose, with the table development project plan designed to meet those goals. If it is not feasible to meet the goals, development of the table should be reconsidered.

2.1.2 Data Confidentiality

When working with data collected from multiple contributors, such as pension plans or insurance companies, it is important to maintain the confidentiality of each contributor's data. No contributor should be able to access another contributor's data, either directly, indirectly or by inference. This requires that non-competitor organizations and/or independent consultants perform most, if not all, of the functions described in this chapter. Execution of confidentiality agreements may be required before access to the data is granted.

If a summary of the data is made available for further analysis, it must be ensured that no combination of variables can be used to isolate data from any one contributor. For example, if a certain combination of distribution channel and distribution method were unique to one contributor, then either the distribution channel or distribution method would have to be excluded from any data made available for further analysis. In general, data should be aggregated to prevent the ability to 'drill down' and isolate data from a single contributor.

2.1.3 Available Data

A review of currently available data must to be done to answer questions such as the following:

- How current is the available data and what has changed from the earliest year of the available data to now? For example, significant changes to an insurance company's marketing, underwriting or products, or to a pension plan's population distribution or benefit features may have occurred over the last few years. The data from before and after significant changes may need to be studied separately.

- What is the trade-off between the greater homogeneity of more recent data and the greater significance that could be obtained by including older, less homogenous data?

- How well does the available data match your needs? Is there a way to create a better match? If not, can you create a credible relationship between the data you have and the data you need?

- How credible is the available data at the anticipated level for which rates will be calculated? For example, consider a lapse study of 10,000 lapses with plans to develop 1,000 lapse rates that vary by product, gender, issue age group, and policy year. With an average of only 10 lapses supporting each rate, the majority of rates will have standard deviations of 30% or more. If the structure of the rates can be simplified to produce only 100 rates, the credibility of the average rate will be much improved.

- Should the rates change gradually and smoothly by age or by another dimension? If so, modeling or graduation of a series of observed rates may increase their collective credibility.

2.2 Data Call

If there is no existing data, or if there is a desire to update existing data, then a formal request for data, often referred to as a "data call," is needed. When structuring a data call, whether for an internal study or an industry study, there is a tendency to ask for more data items than can realistically be used. This may be justified when there are no prior studies to learn from. However, if past studies can be accessed, either by reading reports, analyzing past study results or talking to those who performed the studies, you may be able to narrow the focus on the scope of the data to be collected. For an intercompany insurance study or a study of multiple pension plans, a simpler data call will often result in more contributors to the study.

It is common to create a detailed document that exhaustively documents each data item to be gathered in the data call. This documentation would typically include the following information for each data item requested:

- An abbreviated name or identifier for the data item, such as "Age"
- A full name for the data item, such as "Attained Age"
- A format for the data item, such as a date, integer, floating point number, or character
- A definition of the data item, including how it will be used, such as "Age as of January 1, used to study mortality by age"
- A description of how the data item should be populated or calculated, such as "pensioner's age nearest birthday as of January 1 of the snapshot or transaction year"
- For data items with a list of allowable values, a definition or description of each value

A structure would be defined for storing the data to be collected, such as one or more record layouts or relational database tables. Each record layout or table would be analogous to a spreadsheet, with one column for each data item and one row for each detail record. One or more data items in each record layout or table should serve as a unique identifier for each row, thereby ensuring no duplicates.

Detail records are usually created from a regular series of snapshots (for example, once a month or once a year) or by using a "time lapse" approach that takes a snapshot after every relevant change, such as a change in attained age or benefit amount. The study would typically create a detail record for each combination of a life (associated with a pension, insurance contract or group insurance certificate) and an attained age, policy year, calendar year or other variable that changes over time. The detail records would generally cover periods of time starting no earlier than the study start date and ending no later than the study end date. Depending on the exposure methodology, detail records for the event under study, such as death, may extend to the anniversary following the study stop date.

Finally, for a relational database, table relationships should be documented. Table B can be connected to Table A by including a unique identifier in Table B that ties it to a specific row in Table A. This would make Table B a child of parent Table A. Table relationships can be illustrated in an entity-relationship diagram that resembles a family's genealogy, with younger generations arrayed beneath their parents.

2.3 Common Data Challenges

When collecting data from multiple contributors for an industry study, a variety of things may not go as planned:

- Some contributors may be able to provide data for only certain blocks of business (such as data from certain administrative systems) or certain study years.

- Some contributors may not be able to supply one or more of the requested data items for all or part of their business.

- Terms may have different meanings for different contributors, as practices can vary. For example, "simplified issue" typically encompasses a wide range of underwriting practices in terms of health questions asked and electronic data obtained from credit histories, motor vehicle records and other sources.

- Contributors may not be able to contribute their data on time due to competing priorities.

- Contributors may supply the wrong data or improperly formatted data. Data quality will vary by contributor, administrative system, and other factors.

- There are sometimes significant time lags in the reporting and recording of events, especially deaths.

- Some contributors may not have the resources to transform their data into the requested layout. To obtain a sufficient quantity of data, a decision to accept non-conforming data may be necessary.

Similar data problems can occur when data is gathered for an internal study, especially if the data is spread over multiple administrative systems.

Once the data has been received, it must undergo exhaustive checking to ensure data is valid, consistent, and credible. A future SOA Research paper on Data Processing and Reasonability may address this topic in detail.

2.4 Experience Study Calculations

Experience studies gather data for an observation period defined by study start and stop dates, often starting and ending at midnight on a December 31. The study's start and stop dates dictate the earliest and latest dates for the study's detail records, described in Section 2.2. Detail records are split into two records when they cross a **study anniversary date**, which is typically a birthday, policy anniversary or plan year-end. For studies of rates that vary significantly over the course of a policy year, such as lapse rates, it is common practice to exclude the partial policy years at the beginning and end of the study, thereby avoiding a distortion of the rates. Alternatively, a lapse rate study could develop monthly lapse rates with detail records split at "monthiversaries."

A single study could gather data for more than one event under study. For example, an employee pension study could simultaneously gather data for mortality rates and retirement rates. Similarly, an individual life insurance study could simultaneously gather data for mortality and lapse rates.

2.4.1 Age Basis

When multiple insurers or pension plans contribute data to a study, there can be differences in the **age basis** that determines how age is calculated. For example, North American insurers typically calculate an issue age as of the policy's issue date based on either the insured's age last birthday (ALB) or age nearest birthday (ANB). While the average difference between an ALB age and an ANB age is assumed to be only half a year, that difference can be significant. For example, for a mortality study, a half-year average difference in age can translate to a 4% average difference in mortality rates. There are several ways to address this distortion:

1. If the data were predominantly ALB or predominantly ANB, the data could be treated as if it were all on the predominant basis. If ANB were the predominant basis, this would add a degree of conservatism to the study. For example, if mortality rates increase at an average rate of 8% per year of age, the population was 75% ANB and 25% ALB, and ALB ages are assumed to be one-half year older than ANB ages, the resulting mortality rates would be about 1% higher than a population that was 100% ANB.

2. The date of birth and a current date, for calculating age as of a study anniversary date, or the issue date, for calculating issue age, could be used to recalculate all ages on a consistent basis, using either ANB or ALB.

3. A weighted average age could be calculated for each age, reflecting the percentages of ANB and ALB at that age. For example, if age 50 were 75% ANB and 25% ALB, the weighted average age on an ANB basis would be 50.125. When data is grouped by age to create more credible observed rates, weighted average ages could be calculated for each age group, reflecting both the distribution by age within each age group and the distribution of ANB and ALB for each age. These weighted average ages would then be used for graduation and interpolation.

2.4.2 Experience Study Summary Records

The study's detail records are grouped by study variables and totaled. The resulting summary records consist of two classes of data:

1. Study variables. These consist of potential dimensions for the table to be developed, plus some additional variables. Some table dimensions, such as issue or retirement age, gender, duration (i.e., years since issue or retirement), smoking habits, etc., may be known at the outset of the study, while other dimensions may emerge from data analysis (see Chapter 3). Some variables that do not become table dimensions, such as calendar year and data contributor, may be included to facilitate further analysis.

2. Calculated totals. Totals are calculated for each combination of study variables. Calculated totals include the number and amount of occurrences for each event under study, the number and amount exposed for each event under study and, often, the expected number and amount for each event under study. Expected number and amount may be calculated for more than one existing table.

How the event under study varies with time drives a fundamental decision: Whether the table under development will have annual, quarterly, monthly, weekly, or daily rates or some combination, such as monthly rates for three years followed by annual rates thereafter. For most rates, this question has already been answered by past studies. When there is no predecessor table, two factors will determine the time dimension:

1. How the event's rates vary across time and
2. How financial results depend on the timing of the event.

For example, a life insurance lapse study by policy month might show that policies with annual premiums have a negligible lapse rate for each of the first 11 policy months, followed by a 10% lapse rate for policy month 12. While it could be argued that the lapse rate varies significantly by policy month, more than sufficient accuracy would be achieved by applying an annual lapse rate at the end of the policy year. However, if semiannual, quarterly and monthly modes of payment were included, varying lapse rates by policy month may be preferable.

2.4.3 Count-Based Results

For each combination of study variables and time dimensions (e.g., policy year, policy month, plan year, or calendar year), the following count-based totals are calculated for each event under study, such as one set of totals for deaths and another set of totals for lapses:

- **Event Count** (e.g., the number of deaths for a mortality study)
- **Exposed Count**

The **Count-Based Rate** is then calculated as Event Count divided by Exposed Count.

There are many ways to calculate exposure. **Appendix A, Introduction to Exposure,** provides a high-level summary of some basic choices and provides a reference to a more thorough treatment of exposure.

2.4.4 Amount-Based Results

Amount-based rates reflect the benefit amount associated with pensions, insurance contracts, or group certificates. Amount-based rates are more often used than count-based rates because they are more consistent with financial results. For each combination of study variables and time dimensions, the following amount-based totals are calculated for each event under study:

- The **Event Amount** for a detail record is the sum of benefit amounts associated with the events under study that occurred, such as death claims incurred.
- The **Exposed Amount** for a detail record is the benefit amount in effect during the detail record's period of exposure multiplied by its exposed count. If the benefit amount changed during the exposure period, a weighted-average benefit amount should be calculated. Alternatively, the detail record could be split into multiple detail records, one for each period during which the benefit amount remained level.

An **amount-based rate** is then calculated as Event Amount divided by Exposed Amount.

2.4.5 Expected Results

It is often desirable to compare study results to expected results based on one or more existing tables, especially predecessor tables. For each detail record, the appropriate **Expected Rate** (or rates, if more than one expected table) is looked up and used to calculate the following:

- **Expected Count** = Exposed Count * Expected Rate
- **Expected Amount** = Exposed Amount * Expected Rate

Subtotals of actual and expected counts and amounts can be used to calculate A/E (actual to expected) ratios as follows:

- **A/E Count Ratio** = Event Count subtotal / Expected Count subtotal
- **A/E Amount Ratio** = Event Amount subtotal / Expected Amount subtotal

Actual to expected ratios can be very useful for analyzing data. In effect, they can assign a single number to a large group of data. A/E ratios can be examined to understand how experience varies by each study variable. For example, you could see at a glance how A/E ratios vary by age. This could point out important differences between the table being developed and the expected table. At the youngest and oldest ages, wildly varying ratios are not unusual and are usually due to a lack of credible data. Also, comparing a contributor's A/E ratios to its A/E ratios from a prior study could help identify potential data problems.

2.4.6 Summary

An experience study would typically calculate the following results for each detail record and summarize the results for each combination of study variables. When more than one event is being studied, the following would be calculated separately for each of the events under study:

- **Event results:** Event Count and Event Amount.
- **Exposure results:** Exposed Count and Exposed Amount.
- **Expected results: Expected Count and Expected Amount for one or more expected tables.**

Section 3: Data Analysis

Once the experience study's summary data is ready, the next step is to analyze the study's variables to determine which subset of those variables most influences – and therefore best predicts – the event under study. The "best predictor" variables become the likely dimensions of the table to be developed.

For table development projects that are effectively updates to tables that already exist, the table dimensions are likely to be similar to the dimensions of the most recent table. Perhaps the age range will be extended, the select period will be tweaked or a new predictor variable will be investigated as a possible addition. Table development projects that are essentially updates to existing tables can often be handled in a straightforward manner, as described in 3.5.1. Exploratory Analytics.

This chapter will focus primarily on table development projects that are not updates to existing tables.

3.1 Overview of Data Analysis Process

There are variations among practitioners as to the order of process, but the standard methodology for a data analysis process involves the following steps. Each step is explained in greater depth in this chapter.

3.2 Data Acquisition

Data acquisition starts by answering the following questions:

1. What is the **target variable**, such as the mortality rate in a mortality study, that will be studied and predicted?
2. What questions are to be answered by the study?
3. How will data be gathered for the study, such as internally or from many contributing companies and/or other external sources?
4. What data will be used in the analysis process?
   - Will data be collected at an individual level (such as by policy or by employee) or summarized at a higher level?
5. What data are needed for the study, such as date of death and death claim amount for a mortality study?
6. What data should be collected to better understand and predict the target variable outcomes?
7. What are current thoughts as to which variables may be the key predictors of the target variable?

Once these questions have been answered, the data layout is developed, and data can be collected.

3.3 Data Validation and Preliminary Exploration

Once all the data has been collected, the next step is to perform a rigorous data validation, including both syntax and logic checks, followed by correcting the dataset as needed for use in a complex analysis process. See Chapter 2 for more details. Additional time spent on this first step will be more than repaid by speeding up the analysis process.

This step also includes the following, which are detailed in subsections that follow.

- **Preliminary Data Exploration**: Examining 1-, 2- and 3-dimensional graphs of various combinations of variables to better understand where data is credible and where it is sparse.
- **Outlier Analysis**: Finding extreme values of data that might distort results, such as a single death claim for $30 million. Outliers could be defined as values that are, say, two or more standard deviations beyond the mean.

3.3.1 Preliminary Data Exploration

As a starting point, it can be valuable to examine the distribution of values for each variable in the study, so you understand how the data is distributed, as well as what is common and what is uncommon. Using graphs, you can examine various distributions more quickly.

Two- and three-dimensional graphs can be especially useful when trying to examine combinations of two or three variables. For example, for a select mortality study, distributions of deaths and exposure by issue age and policy year could be quickly visualized using two-dimensional graphs; these could help establish the issue age range and select period for the table to be developed. See **Appendix B, Select Period Considerations**.

Simple scatter plots can be helpful in identifying correlations among two variables, as well as identifying outliers and inconsistencies.

Often, this step will identify one or more obvious predictor variables. This step may also help identify potential predictor variables by using Excel pivot tables, query tools such as SQL, or software such as Tableau to examine how rates vary by different dimensions. The potential predictor variables identified in this step will be tested as part of modeling.

3.3.2 Outlier Analysis

An important step in data exploration is to review the **outliers,** defined as data with extreme, unusual or missing values. Subject-matter experts will be needed to weigh in as to the magnitude and appropriateness of each outlier's impact on results. For each variable, a decision should be made as to which outliers should be included and which should be excluded from the analysis. An alternative to excluding values is setting a floor or a ceiling. For example, you might cap extremely high death benefits at a lower level, so that they do not distort rates.

For missing or faulty values, it may be appropriate to replace them with an average or most common value. For example, in a pension study that analyzes mortality by salary amount for active employees, those with missing salaries may be assigned an average salary. In such a case, it may be advisable to create an indicator variable that identifies which values have been assigned. This would allow analysis to

be done with and without assigned average salaries, which could be helpful in determining whether there was an unintended bias.

3.4 Data Visualization and Preparation

This step involves a more detailed exploration of the cleaned and validated data with the goal of identifying potential groupings of variables for improving credibility or for modeling purposes. In addition, new variables that could be useful to the analysis can be created from the original study variables. This step can also include a variable reduction exercise, which is often needed where there are many potential study variables.

3.4.1 Variable Grouping

Of special interest is how the variable under study, the **target variable**, is distributed in terms of its likely predictor variables. For example, in a pension mortality study, you may want to view the distribution of deaths, exposure and mortality ratios to an established mortality table by gender, attained age, and size of pension. This could help you determine how best to group the data by size of pension, taking into consideration both the credibility of grouped results and the homogeneity of mortality ratios for pension sizes that will be grouped together. A common goal is to strike a balance between credibility and homogeneity of grouped results.

3.4.2 Variable Creation

The formats of the data fields should be examined to identify additional or alternative formatting that may better support both the variable reduction exercise, discussed below, and other modeling work. For example, for graduation, it may make sense to include both individual age and age-group variables. Another variable might be added to indicate individual ages or age groups with high, medium or low credibility, based on the underlying number of policies or pensioners.

Variables could be created to provide supplemental information related to existing variables. For example, for a study of deferred annuity withdrawals and surrenders, it may be useful to add a variable indicating the current surrender charge percentage and another variable indicating the number of years until the surrender charges end. For mortality studies, BMI (body mass index) might be calculated from, and used in place of, height and weight.

3.4.3 Variable Reduction

When many explanatory variables are collected with each observation, there may be groups of variables that are correlated. Groups of correlated variables should be statistically examined to better understand their relationships. It may be possible to eliminate one or more variables without losing much predictive ability by applying a variable reduction technique.

Common variable reduction exercises include:

- Exploratory factor analysis
- Standard factor analysis
- Principle component analysis

A detailed discussion of these techniques is beyond the scope of this paper.

After the variable reduction exercise, variables that have been deemed not explanatory and unneeded for future analysis should be removed from the database and the data should then be re-summarized. This may shrink the database by one or more orders of magnitude, thereby dramatically reducing the run time for modeling. Also, if the database includes expected values for many expected tables, additional savings can be obtained by removing expected values for those expected tables deemed no longer relevant.

3.5 Determination of Analytical Approach

At this step in the process, there is enough information about the underlying data to decide on an analytical approach. Currently, one of two analytical approaches is applied to experience study data to develop a table:

1. Exploratory Analytics
2. Advanced Analytics

3.5.1 Exploratory Analytics

Prior to about 2010, virtually all life, health, annuity, and pension table development was performed using what we will call **exploratory analytics**, which might also be called descriptive analytics. This is essentially a trial-and-error approach to identifying the variables with the most predictive power and an acceptable level of credibility. The identified variables become the table's dimensions. The next step would be graduation, which is covered in the next chapter. Since 2010, advanced analytics, described in the next section, have been increasingly used to develop significant industry tables, especially those without predecessor tables.

The exploratory analytics process examines combinations of potential predictor variables with the goal of identifying those that best explain significant variations in the observed rates. This is typically accompanied by a study of how a credibility measure (such as number of events, exposed counts or confidence intervals) varies by different combinations of predictor variables.

The quest is to find the combination of predictor variables that best achieves the following goals:

1. Includes study variables that explain significant variations in observed rates.
2. Provides a relatively high percentage of observed rates that are credible.
3. Provides credible observed rates that are relatively smooth within each array.

The key word is "relatively." Fewer than half the observed rates may be credible. "Relatively smooth" may mean a jumble of observed rates for which you can visualize a smooth curve or smooth surface that passes close to most of the rates.

To improve the credibility and relative smoothness of observed rates for graduation, data is sometimes grouped by age groups, size groups, and other groupings. However, the trade-off is a sharp reduction in the number of observed rates available for graduation.

Models, as described in the next section, can be used in lieu of, or in addition to, the above trial-and-error approach to determine the predictor variables most likely to become the dimensions of the table to be developed.

Count-based observed rates (see 3.6.5, Count- and Amount-Based Rates) are sometimes used to determine the table's dimensions, since count-based observed rates are less volatile than amount-based observed rates. However, amount-based observed rates are typically used to populate the final arrays of observed rates, since amount-based rates are more consistent with financial results.

Once the best combination of predictor variables is determined, arrays of observed rates are constructed using those variables as the table's dimensions. The arrays of observed rates would then be graduated to create smoother rates. If observed rates were calculated using age or other groupings, and rates were desired for individual ages or other values, the graduated rates would then be interpolated. Both graduation and interpolation are covered in the next chapter.

3.5.2 Advanced Analytics

Since about 2010, advanced analytics have been increasingly used to develop major industry life, health, and pension tables. The term **advanced analytics** is meant to include statistical analysis, statistical modeling, multivariate modeling, and predictive modeling.

Before the modeling process begins, the data must be divided into at least two randomly selected, independent subsets. A large training subset, typically consisting of 60% or more of the data, will be used for model development and fitting. The other subsets will be used for model **validation** to test how well the fitted model, created using the training subset, predicts results when applied to the independent **validation subsets**.

Modeling can be used to simultaneously analyze many variables to determine which are the best **predictor variables** for the target variable.

3.6 Model Creation and Assessment

This section focuses on advanced analytics and outlines a process for applying modeling techniques to develop a table.

3.6.1 Selecting Appropriate Model Forms

Not all forms of modeling are appropriate for life, health, annuity, and pension tables, which have three key requirements:

1. **Explanatory ability**: The ability to explain or interpret the results of the model so consumers and regulators can see that rates are equitable and that there has been no discrimination in the determination of rates.
2. **Predictive power**: The ability of the model to create rates that closely match observed rates.
3. **Ease of table implementation**: The ability of the model to generate arrays of rates.

The following model forms are difficult to interpret or explain and are often referred to as "**black boxes**" because the effect of individual predictor variables on rates is not apparent:

- Random forests
- Neural networks
- Support vector machines (SVMs)
- Gradient boosted machines (GBMs)

In many insurance and pension situations, a lack of explanatory ability is not acceptable. Insurance companies and pension plans often need to justify how and why rates vary. However, these "black box" models continue to be explored because of their potentially strong predictive power.

The following model forms are much more transparent and are generally the easiest to explain. GLMs have shown above-average predictive power for some insurance experience-based applications, but other model forms also show potential:

- Generalized linear models (GLMs)
- Generalized additive models (GAMs)
- Decision trees
- Multivariate adaptive regression splines (MARS)
- Penalized regression

In terms of ease of table implementation, GLMs, Penalized Regression and MARS rank the highest, while GBMs and random forests rank the lowest.

Looking at the requirements for life, health, annuity, and pension tables, this second group of model forms may be more appealing. However, the modeling analysis could employ more than one model form. For example, you might develop a GBM model to maximize predictive power, and then develop a GLM model that produces similar predictions, though not as highly predictive. The GLM model could then be used to explain the GBM model.

3.6.2 Generalized Linear Models

Because GLMs have been used more than any other model form for life, health, annuity, and pension table development, we will look more closely at the GLM family and its close cousins, Generalized Additive Models ("GAMs").

A GLM is a flexible generalization of ordinary multiple linear regression, which assumes the **target variables' errors** (i.e., the differences between observed and predicted rates) are normally distributed. A GLM allows for target variables whose error distributions are not limited to the normal distribution. For example, GLMs allow for exponential error distributions, such as the Poisson distribution, the gamma distribution, and the negative binomial distribution. When choosing a GLM model form, it is important to match the model form with the error distribution of the target variable.

GLMs use a link function to allow the estimation of an otherwise nonlinear system to occur in a linear fashion. For example, the observed number of deaths or lapses typically follows a Poisson distribution, which is an exponential distribution. Using a logarithmic (i.e., the opposite of exponential) link function creates a linear model for the estimated mortality or lapse rate.

The GLM family includes logistic regression, which allows predictor variables to be categorical. Typical categorical variables include gender and smoking habit.

Generalized Additive Models (GAMs) include GLMs as a subset. GLMs optimize linear factors applied to link functions of the predictor variables to predict values of the target variable. In contrast, GAMs allow for smooth functions in general to be applied to the predictor variables, rather than applying linear factors to predictor variables as GLMs do. GAMs use underlying patterns in the data to derive these smooth functions. The degree of smoothness can be controlled to avoid wiggly, nonsensical predictor functions.

On the other hand, a GLM's predictor variables can be transformed to create nonlinear predictor variables within the GLM. For example, rather than using age "as is," you could add predictor variables that are various functions of age, such as the logarithm of age, the inverse of age, age squared, etc. It may require much trial and error to build a family of nonlinear components that match the smooth functions automatically derived by a GAM.

3.6.3 Survival Models

Survival models are more specialized. They are focused on predicting the survival of insureds, sometimes in combination with a specific medical procedure. The more common survival models include the following:

- **Cox proportional hazard model**: Simultaneously models the effect of several factors on survival, to predict survival time for insureds.
- **Weibull Analysis**: Most often used to model and predict the life characteristics of a manufactured product, such as its mean life, reliability or probability of failure at a specific time. This can be adapted to insured lives.
- **Gompertz-Makeham**: The Gompertz-Makeham law of mortality is out of step with modern mortality studies. However, the Gompertz function offers a shape that could be useful for non-accidental and nonviolent deaths. The function starts with an asymptote at the youngest ages, where childhood mortality rates are low and fairly level, then ascends with exponential growth through adulthood, finally grading into another asymptote at the end of life, where the world's oldest people currently exhibit an annual mortality rate of about 0.6.

3.6.4 The Modeling Process

After one or more suitable model forms have been selected, the model creation and assessment steps are performed for each model form. Model fitting usually begins by including every potential predictor variable remaining after the variable reduction exercise. An opposite approach is sometimes used by starting with one or more obvious predictor variables, and then testing the addition to the model of one candidate predictor variable at a time.

Modeling is a statistical process that makes assumptions about the underlying distribution of the actual rates compared to the predicted rates. For example, if differences between actual and predicted rates were normally distributed, there is a 95% confidence level that the actual rate will fall within plus and minus two standard deviations of its predicted rate.

Once a model has been fitted to the training subset, each potential predictor variable can be analyzed to see whether its predictive power is more likely real or random chance. To answer that question, the common model forms currently used for life, health, annuity, and pension table development output a **p-value** for each potential predictor variable.

The p-value, which ranges between 0 and 1, is an estimate of the probability that there is no actual difference in the underlying rates for different values of the potential predictor variable. Said another way, it is the probability that variances in the observed rates are due to random chance rather than the effect of the predictor variable.

If the p-value is very low, you can reject the null hypothesis that the predictor variable has no effect on the rate: A p-value of less than 5% generally indicates a variable that is a significant predictor of the target rate. As the p-value increases beyond 5%, the likelihood that the variable is a significant predictor decreases rapidly.

In the model-fitting process, each time a model is revised by adding or removing a predictor variable, its predictive power should be inspected. If predictive power has declined, the last change made to the model should be adjusted or reversed. The model's p-values should also be inspected, especially for candidate predictor variables that were just added. Candidate predictor variables with p-values much higher than 5% should be removed. Variables with p-values close to 5% may be retained and reassessed as the model progresses.

Confounding can occur when an important predictive variable has been omitted. For example, looking at mortality rates by gender alone could be confounded by smoking if the prevalence of smoking for males is higher than it is for females. If smoking is added as a predictor variable, the gender relationship will change.

After multiple iterations of model fitting, the key predictor variables will be known and ranked. A common next step is to explore the potential for transformations, functions, and interaction terms. These are created from the key predictor variables to further enhance the model's predictive power, as described below.

1. **Transformation of a study variable**: For example, transforming gender from categorical values of "female" and "male" to numeric values. With modern software, this kind of transformation can be applied without having to change the input data.
2. **Function of a study variable**: For example, $(age)^{-1}$ $(age)^2$, or log(age).
3. **Interaction term:** A combination of two or more predictor variables or functions of predictor variables, such as issue age times policy year.

Interaction terms are important because two or more predictor variables can sometimes be far more predictive when combined than when included separately. For example, in a study of post-level term mortality (i.e., mortality after the end of the level-term period), it was found that, individually, both attained age and the length of the level-term period were only weak predictors of mortality rates. However, a combination of the two was a strong predictor.

Stratification is an equivalent alternative to using interaction terms. For example, instead of including an interaction term for gender and smoking status, you could create and run separate female and male models, each with a smoking variable. If the relative risk for smoking were similar in the gender-specific

models, then there would be no need to have an interaction term in a full model that included both females and males. One possible advantage to stratified models is they can be easier to understand, particularly when there are multiple interaction terms and each interaction term has multiple categories.

In addition to p-value, other statistics are useful for ranking the significance of a set of predictor variables, such as the following:

- Type 3 or Type III analysis,
- Bayesian information criterion (BIC), and
- Akaike information criterion (AIC), which is closely related to BIC.

For each refinement of the model, you will generally accept the refinement if it improved predictive power, and reject it otherwise. As improvements in predictive power diminish, you should be wary that further improvements may come at the expense of model overfit. When you suspect that point has been reached, the model should be tested against the data that was held back for validation purposes. If the model performs more poorly against the validation data than it did against the training data, then the model has been overfit and needs to be simplified.

The first attempts at model simplification should test the removal of any highly-correlated predictor variables, which can often lead to model overfit. Next, you might remove the model refinements that added the least predictive power.

Once a simplified version of the model performs well against the training data, it should be tested against the validation data. If the model performs equally well against both data sets, it can be deemed complete.

3.6.5 Count- and Amount-Based Rates

Before each iteration of the model, subtotals for the event under study and exposure are used to calculate the observed rate for all combinations of the predictor variables to be tested. For example, if gender, issue age and policy year were the only predictor variables for a mortality study, death and exposure subtotals would be used to calculate the observed mortality rate for every combination of gender, issue age and policy year occurring in the study data.

Two types of observed rates can be modeled and calculated from totals for each combination of predictor variables:

- Count-based rate = Event Count / Exposed Count
- Amount-based rate = Event Amount / Exposed Amount

The amount-based rate is equivalent to using a count-based rate weighted by amount, which can be accomplished by using the "weight" option available in most statistical modeling procedures.

A first round of modeling may be focused on count-based rates, which are less volatile than amount-based rates. A second and final round of modeling may then be performed using amount-based rates, which better reflect financial results.

Once the model for count-based rates is finalized, it becomes the starting point for the model for amount-based rates. The second model uses the predictor variables and structure created by the first model, typically requiring only fine tuning to fit the model to amount-based rates instead of count-based rates.

3.7 Final Model Selection

Once a model is completed, it should be assessed for acceptability in terms of its predictive power, explanatory ability and ease of table implementation. If deemed acceptable, it should be saved and compared to any previous versions of the model, as well as to other model forms being assessed, to determine which of the fitted models best meets the project's objectives.

As discussed at the end of "Section 3.6.1 Selecting Appropriate Model Forms," it may be that one model has the best predictive power while another model with substantially similar results has much better explanatory ability. If so, it may be possible to select the model with the best predictive power, while using the other model for explanatory purposes.

Once a model has been selected, it may be capable of producing arrays of smooth rates suitable for populating the table being developed. If the model cannot produce sufficiently smooth rates, the model's rates can be output and then graduated, as outlined in the next chapter.

3.8 Minimizing Table Dimensions

Whether using exploratory or advanced analytics, it is often desirable to minimize the number of arrays of rates that make up the table being developed. For example, the team that developed the 2015 VBT mortality table, for fully-underwritten individually insured lives, found that select mortality was best predicted by six variables. The team noted that:

> "Using the traditional issue age and policy year grid to display the mortality rates would require 96 such grids to display every combination of gender, smoker status, issue year era, and face amount band."

The team decided to drop variations by issue year era and face amount band to hold the number of arrays to a more manageable number.

3.9 Replacing Grids with Factors

For an independent predictor variable with relatively stable effects across the other dimensions in the table, it may be possible to apply a factor approach. This is often feasible when weak or insignificant interaction effects between the independent predictor variable and the other predictor variables has been observed.

Rather than create one or more extra arrays for each value of the independent predictor variable, sets of factors for that predictor variable could be developed to apply to arrays that do not include that predictor variable. The factor approach is illustrated in the following example:

> Suppose mortality rates varied by issue age, policy year, and two categorical variables, gender (female and male) and smoking habit (nonsmoker and smoker). Further suppose it was observed

that the effects of nonsmoking and smoking were relatively constant across gender, issue age, and policy year.

Rather than create an issue age/policy year array of rates for each of the four combinations of gender and smoking habit, you could develop just two arrays, for female and male "unismoker" rates. You could then develop separate nonsmoker and smoker factors to apply to the female and male unismoker arrays to reflect the effects of nonsmoking and smoking.

The nonsmoker and smoker factors could be constants or could vary by any combination of gender, issue age or policy year. However, if the factors varied by both issue age and policy year, there would little to no advantage to using factors.

While the use of such factors can reduce the number of arrays in a table, the cost is a reduction in accuracy.

Section 4: Graduation of Rates

Graduation is a mathematical process that inputs an array of rates requiring smoothing and outputs an equal-sized array of smoothed rates. There are two common situations in which graduation is used for table development:

- When modeling is not used to develop rates, arrays of observed rates are graduated. Even the largest, most credible studies produce observed rates that are not sufficiently smooth.
- When modeling is used to develop rates, but those rates require smoothing, arrays of modeled rates are graduated.

When modeling is used to develop rates and those rates are sufficiently smooth, there is no need for graduation.

Graduation is often justified by an assumption that there are underlying "true" rates that sketch a continuous curve. By smoothing out irregularities, graduation approximates these "true" rates, effectively making use of the combined credibility of the entire array of rates.

For an introduction to graduation, Chapter 1 of *Graduation: The Revision of Estimates* by Dick London is recommended.

Methods of graduation include graphically fitting a curve to input rates or solving for the parameters of a parametric formula such as those developed by Gompertz and Makeham. As those methods are rarely used, they will not be discussed further.

A preferred graduation method has the following properties:

1. Weights can be input that will cause the graduated rates to reproduce the same total events as the rates input.
2. Parameters are available to control the balance between how closely the graduated rates fit the input rates and how smooth the graduated rates are.

When performing graduation, the goal is to achieve sufficient but not excessive smoothing, which can eliminate important features of the underlying data. It can be difficult to know when smoothing is sufficient but not excessive. An approach using confidence intervals is presented later in this chapter to address this challenge.

4.1 Input Rates

Input rates are either rates produced by a model or are observed rates calculated from events and exposures. There are two types of observed rates, calculated as follows:

- Count-based rate = sum of event counts / sum of exposed counts
- Amount-based rate = sum of event amounts / sum of exposed amounts

Count-based rates are less volatile than amount-based rates and their confidence intervals are easily calculated. Therefore, count-based rates can be graduated with more confidence.

Amount-based rates produce results that are more consistent with financial results, but they can be significantly more volatile. For example, a single large claim in a mortality study may skew any rate that

includes that claim. An array of amount-based rates may include significant irregularities that are not easily removed by graduation.

One approach is to first graduate the count-based observed rates. The pattern of graduated count-based rates can then serve as model for the graduation of amount-based rates.

4.2 Insufficiently Credible Data and Grouping of Data

In virtually every experience study, there will be some observed rates that are outliers due to a lack of credible data, most often at the lowest or highest ages. For example, an observed rate with an exposure of 10 lives and an expected rate of 10% could just as easily have an observed rate of 0% or 20%. When such **incredible** (i.e., not credible) observed rates are input into a graduation, they can distort graduated rates for nearby credible observed rates. This is because the smoothness calculation gives equal weight to incredible and credible input rates. On the other hand, credible observed rates can help produce more reasonable graduated results for their neighbors with incredible observed rates. The farther incredible observed rates are from neighbors with credible observed rates, the more likely it is that their graduated rates will be distorted.

An observed count-based rate based on 10,000 events will have a standard deviation of about 1% of the rate, while an observed rate based on 100 events will have a standard deviation of about 10% of the rate. Standard deviations of 1% to 20% are typical for most observed rates. See **Appendix C, Table of Standard Deviations and 90% Confidence Intervals**. When graphed, any array of observed rates will typically show at least a few irregularities due to statistical fluctuations.

For smaller studies, results are sometimes grouped by quinquennial age groups, so that graduation can be performed using more credible observed rates. However, there is a tradeoff: Grouping results in fewer rates to graduate lowers the credibility of the graduation.

Within each age group, the distribution of exposure by age will not be uniform, especially for the youngest and oldest age groups. Rather than assume an average age equal to the midpoint of the age group (i.e., a midpoint of age 52 for the 50 to 54 age group), it is better to calculate an exposure-weighted average age for each age group. The weighted-average ages would be used both for graduation and when interpolating the graduated age-grouped rates to create rates for individual ages.

Similarly, to create more credible observed rates, results could be grouped by years, such as groups of issue years, retirement years, calendar years or policy years. For example, policy years 1 through 15 could be grouped into seven policy year groups (1, 2, 3, 4-5, 6-7, 8-10, and 11-15), such that each policy year group is more likely to have credible observed rates. Again, the more credible rates come at the cost of fewer rates available for graduation. An exposure-weighted average policy year could be calculated for each group of policy years, rather than using the midpoint.

4.3 Graduation Methods

This chapter focuses on graduation methods that solve for graduated rates, while optimizing a combination of fit and smoothness. The most commonly-used graduation methods are Whittaker-Henderson and its close cousin, Whittaker-Henderson-Lowrie. Chapter 4 and Section 8.4 of Dick London's *Graduation: The Revision of Estimates* is recommended for those seeking a thorough

understanding of Whittaker-Henderson. Tips for using Whittaker-Henderson can be found in **Appendix D, Using Howard's Whittaker-Henderson Functions**.

Another graduation method, penalized B-splines, also known as P-splines, has been used by CMI in the UK to graduate mortality tables.

4.4 Graduation Components

Graduation seeks a proper balance between fit and smoothness. Observed or modeled rates are input and graduated rates are output. The goal is to smooth the input rates, while remaining true to the overall level and general pattern exhibited by the input rates. To achieve this balance, graduation methods provide one or more parameters that control the relative weight given to fit versus smoothness.

Input Fit is calculated based on a weighted sum of the squares of the differences between graduated rates and input rates. The squares of the differences would typically be weighted by exposure or some other measure of credibility before being summed up. Note that fit is maximized when the weighted sum of the squares of the differences is minimized.

Table Fit: If it is desired to also fit the graduated rates to a table of rates (i.e., rates from an existing table or graduated count-based rates from a prior step), a second measure of fit, Table Fit, can be calculated using the same approach used to calculate Input Fit, by substituting table rates for input rates. The Table Fit sum would be multiplied by a Table Fit Weight to achieve the desired balance between Input Fit, Table Fit, and Smoothness.

Linear Transformation before Table Fit: When Table Fit is part of graduation, the table rates should be transformed so their weighted average matches the weighted average of the observed rates. Additionally, the slope of the table rates should be transformed to better match the slope of the observed rates. These two adjustments could be simultaneously handled by using a two-dimensional linear or exponential transformation.

Expected Rate Alternative to Table Fit: A simpler way to incorporate an existing table into the graduation process is to calculate actual-to-expected (A/E) ratios, with expected rates based on the existing table. Rather than graduate input rates, you would graduate A/E ratios, i.e., input rates divided by expected rates. For Whittaker-Henderson graduation, instead of using weights proportionate to exposure, you would use weights proportionate to expected events (i.e., expected rate times exposure), to ensure that graduated ratios reproduced total events.

Smoothness is often measured by the sum of the squares of the third differences of the graduated rates, although second or fourth differences may be more appropriate in some situations. Note that smoothness increases as the sum of the squares decreases. Unlike the measurement of fit, the squares that contribute to the smoothness calculation are <u>not</u> typically weighted by exposure or any measure of credibility before summing them up. This means that distorted rates often found in cells with low credibility can have an undesirably large effect on the measure of smoothness. The sum of the squares of the smoothness differences is typically multiplied by a smoothness weight to achieve the desired balance between Smoothness, Input Fit, and Table Fit.

For two-dimensional graduation, smoothness is calculated separately for each dimension and each dimension has its own smoothness weight. In other words, a two-dimensional graduation would calculate smoothness separately for horizontal rows of rates and for vertical columns of rates, weighting the horizontal results with a horizontal smoothness weight and the vertical results with a vertical smoothness weight.

Once the components of graduation – Input Fit, Smoothness, and, if applicable, Table Fit – have been calculated, the mathematical process underlying the graduation method solves for the graduated rates that minimize the sum of the weights times the measures of Input Fit, Smoothness, and, where applicable, Table Fit. Each of these goals are maximized as their measures approach zero.

4.5 The Graduation Process

The graduation process would typically follow steps such as the following:

1. Collect and populate graduation input.
2. Review and adjust graduation input as needed.
3. Set fit and smoothness parameters and run the graduation algorithm.
4. Review and evaluate the graduated rates; summarize the results of the graduation.
5. Repeat steps 3 and 4 until you have exhausted all the combinations of fit and smoothness parameters you want to test. This is a trial-and-error process that gradually leads to more acceptable results.
6. Run a final graduation using the fit and smoothness parameters that produced the best results.

Steps 1, 2, and 4 are described in more detail in the following sections.

4.5.1 Populate and Adjust Graduation Input

This step is typically performed once for each set of rates to be graduated.

When observed rates are to be graduated, the graduation process begins by querying experience study results to create arrays of observed rates and exposure for input to the graduation algorithm. The input may include additional information to gauge the credibility of graduation results, such as variance or event count.

When modeled rates are to be graduated, the same input data is needed, with the exception that modeled rates are substituted for observed rates.

Some of the cells in the input arrays may have no data. When there is no exposure for a cell, the input rate may show a divide-by-zero error, which most graduation algorithms cannot handle. Divide-by-zero errors can be replaced with input rates of zero. Alternatively, the scope of the graduation could exclude the ages at the low and high ends of the age range where insufficiently credible data is more apt to result in divide-by-zero errors. Graduation results are often improved by excluding data with low credibility. It may require some trial and error to determine the optimal minimum and maximum age for graduation.

When rows or columns are excluded from the graduation, an alternative approach is needed to generate rates for the rows or columns that were excluded. One solution may be to simply shrink the

table to exclude the ages with low credibility. Another solution may be to use percentages of an existing table to extend rates to the rows excluded from graduation. Adjustments would typically be made so graduated rates grade smoothly into extended rates for the oldest ages. Please refer to **Chapter 5, Extension of Rates**.

4.5.2 Review and Evaluate Graduation Results

We will visualize the graduated rates as one or more two-dimensional arrays. Most of the following discussion also applies to one-dimensional arrays. When working on a table with many arrays, it is worthwhile to automate the review process as much as possible.

In practice, it is not necessary to review every graduation iteration in depth, as many iterations are needed to gradually refine the fit and smoothness parameters. Once you are satisfied with high-level results, such as the graduation's output measures of fit and smoothness, a full review and a recording of key results is in order. Often, graduation is then repeated with a different approach, in a search for better results.

Once you are satisfied with the graduated results, the following reviews are in order:

1. **Ensure graduated rates reproduce the same total events as input rates.**

   An automated check should be incorporated to verify this. The verification can be done by separately multiplying all input rates and all graduated rates by the same exposure weights and summing the results. If the sum of weighted graduated rates does not equal the sum of weighted input rates, then total events have not been reproduced. There is likely an input error that must be found and corrected before proceeding.

   A secondary check can be performed to see how closely the graduated rates reproduce total events for each row and column. If some rows or columns on the edges of the graduated array poorly reproduce total events, the solution may be to exclude such rows or columns from the graduation or combine some rows or columns to increase the credibility of their input rates.

2. **Determine the percentage of graduated rates that fall within their input rates' confidence intervals.**

   Wherever possible, calculate or estimate standard deviation for each input rate and use it to calculate confidence intervals around each graduated rate. Refer to **Appendix C, Table of Standard Deviations and 90% Confidence Intervals**, and Section 4.6, Variance for Amount-Based Observed Rates.

   Next, determine the percentage of graduated rates that fall within their input rate's confidence interval and compare that to the expected percentage. For example, one would expect, on the average, that 90% of graduated rates would fall within their input rates' 90% confidence intervals.

   o If significantly more than 10% of the graduated rates fall outside the 90% confidence intervals, this may indicate the graduated rates have been overly smoothed.

   o If significantly fewer than 10% of the graduated rates fall outside the 90% confidence intervals, this may indicate the graduated rates have been overly fitted to the input rates.

3. **Review graphed results.**

Use a graphing capability to examine graduated rates in comparison to their input rates and, if applicable, to their table rates, one row or column at a time. Make a note of any rows or columns with disconcerting differences. Such a visual review can be surprisingly helpful in spotting problems.

Rather than exclude some rows or columns with unacceptable graduated rates from the graduation, an alternative may be to combine two or more rows or columns to increase the credibility of the input rates to be graduated.

If you decide to exclude or combine rows or columns, the graduation process will have to start over, with a first step of repopulating the graduation input. However, what you learned the first time through the graduation process should make the "do-over" process proceed faster.

The output from the graduation algorithm may include measures of Input Fit, Table Fit, Horizontal Smoothness, and Vertical Smoothness, in addition to the graduated rates. These output measures should be recorded for each iteration of the graduation, along with the input parameters used to control fit versus smoothness. By recording and reviewing recent sets of input parameters and their output measures, you can better estimate the input parameters that will produce the optimal output measures.

4.6 Variance for Amount-Based Observed Rates

Variance and standard deviation for count-based observed rates are easily calculated from the number of observed events and exposure for each observed rate, as shown in **Appendix C, Table of Standard Deviations and 90% Confidence Intervals**.

For amount-based observed rates, variance, standard deviation, and confidence intervals are not so easily calculated. However, if amount squared times exposure count were calculated for each life, and the sum of amount squared times exposure count were included in experience study summary records, then variance for observed amount-based rates would be easily calculated as follows:

Variance for amount-based observed rate = (observed rate) * (1 – observed rate) *

Sum of (amount squared times exposure count for each life)

When designing an experience study, the inclusion of "sum of amount squared times exposure count for each life" would add value to the study. It would allow the credibility of amount-based rates to be calculated, thereby improving modeling, graduation, and adjustment of rates.

For situations where the "sum of amount squared times exposure count for each life" is not available, **Appendix E, Estimating Variance for Amount-Based Observed Rates**, provides an approximate method of estimating variance based on exposed counts and amounts by size group.

4.7 Interpolation of Graduated Rates

If groupings of ages, years, or other variables were used to increase the credibility of observed rates input to graduation, then graduated rates must be interpolated to create rates for individual ages, years, or other variables. Interpolation may not be required when modeling is used, since some models can readily produce rates for any combination of the model's predictor variables, including individual ages and years.

4.7.1 Use of Weighted Averages Instead of Midpoints

When ages, years, and other variables are grouped, the midpoint of each grouping should <u>not</u> be used in interpolation. Instead, an exposure-weighted average should be calculated for each grouping and used for interpolation. Since exposure differs for count-based and amount-based input rates, separate weighted averages may be needed for each.

Some examples will illustrate situations where midpoints are not appropriate:

- When policy years 6 through 8 are grouped, the weighted average policy year will almost always be less than the midpoint of 7 because lapses reduce the business that persists from one year to the next.
- When retirement ages 65-69 are grouped, the weighted average retirement age will likely be less than the midpoint of 67 because of a preponderance of retirements at age 65.
- When issue ages 85-89 are grouped, the weighted average issue age will be less than the midpoint of 87 because high mortality rates at those ages mean there will likely be fewer potential purchasers at each higher issue age.

4.7.2 Common Interpolation Methods

Interpolation methods used in the insurance and pension fields include the following:

One-dimensional interpolation methods:
- **Linear** interpolation is the simplest method, which fits a straight line between the pivotal values on either side of the rate to be interpolated.
- **Cubic spline** interpolation fits a cubic curve, or third-order polynomial, using two pivotal values on either side of each interpolated value. Excel Add-ins can be found for cubic spline interpolation.
- **Log-Linear** or **Log-Cubic** interpolation applies linear or cubic interpolation to the <u>logarithms</u> of pivotal values:
    - Since the mortality rates surrounding most ages increase at a fairly constant percentage, this method is very useful for interpolating mortality rates and mortality-improvement rates.
    - Log-linear and Log-Cubic interpolation have three steps:
        1. Logarithms of the pivotal values are calculated, thereby creating "pivotal logarithms."
        2. Linear or cubic spline interpolation is applied to the pivotal logarithms.
        3. The interpolated logarithms are exponentiated to convert them to final interpolated results.

Two-dimensional interpolation methods:
- **Bilinear** interpolation applies linear interpolation to one dimension and then applies linear interpolation to the second dimension using the results of the first interpolation.
- **Bicubic** interpolation is a two-dimensional extension of cubic spline interpolation, using a 4 by 4 grid of 16 pivotal values that surround the rate to be interpolated. The resulting interpolated surface is smoother than surfaces created by bilinear interpolation. Excel Add-ins can be found for bicubic interpolation.

Section 5: Extension of Rates

A common limitation when developing a table is a paucity of data at the youngest and oldest ages. It is often necessary to look beyond the study's data to find supplemental data to extend rates to the youngest and oldest ages.

When studying old age mortality, a common problem is the underreporting of deaths, perhaps because there would have been no financial incentive to report the death for pensions while, for insurance, the beneficiaries may have had no idea that an insurance policy was in force.

5.1 Using Rates from an Existing Table

If there is a suitable industry or internal table that is considered credible, it can be used to provide guidance on the pattern or level of rates. For example, if the study has sufficiently credible data to set the overall level of rates for the youngest or oldest ages, you could use the slope from an existing table to guide the pattern of the rates. Alternatively, you could grade rates from the study's closest credible rates to appropriate rates from the existing table.

Often, a study will develop multiple arrays, such as for females and males, nonsmokers and smokers, etc. Once the youngest and oldest rates for one array have been established (say, for male nonsmokers, which may have the most data), rates for the other arrays can be set in a way that is logically consistent.

As an interim step, it may be helpful to create an array that combines all data for the youngest and oldest ages. The interim array can then serve as a pattern for developing rates for the desired arrays (such as male nonsmoker, female nonsmoker, male smoker, and female smoker).

5.2 Using Data from Other Sources

Relevant data may be available from studies performed by the government or other sources. For example, government-sponsored demographic studies, such as those produced by the U.S.'s Social Security Administration, may provide relevant data on mortality rates at the youngest and oldest ages. When deciding on an omega (the final age) for a mortality table, data on the world's 100 oldest people, easily found on the Internet, may be relevant. It may also serve as a reality check for mortality rates for the very oldest ages. Please refer to **Appendix F, Mortality Study of World's 100 Oldest People.** Medical studies may be of value when building tables related to health insurance or medical underwriting.

5.3 Using Formulas

When data is very sparse to nonexistent, formulas can be used to project rates beyond the oldest or youngest ages with credible data. The formula would typically be adjusted so it closely reproduces the rates for the nearby ages with credible data. This approach has been used for old-age mortality rates.

Section 6: Review and Adjustment of Rates

The most arduous and thankless part of table development can be the review and adjustment of rates. There are two broad categories of review and adjustment:

- First, **Creative reviews** challenge and improve the rates produced by models, graduation, and interpolation.
- Next, **Enforcement reviews** apply rules that govern relationships between rates, as agreed to at the outset of the project.

6.1 Creative Reviews

Rates produced by a model, graduation, or interpolation can be deemed deficient for many different reasons:

1. Rates can be nonsensical, such as a mortality rate less than zero or greater than one.
2. Rates can be based on data with little to no credibility.
3. Rates with moderate to high credibility may have been skewed by nearby extreme rates with little credibility.
4. Due to idiosyncrasies in the data, there may be parts of the table with non-sensical patterns of rates. Such patterns can often be discovered by using one of the following approaches:
   - By graphing an array of rates and examining one row or column at a time, you can quickly discern areas that exhibit unusual patterns, such as compression or overlapping of rates and unusual changes in slope. Alternatively, a two-dimensional array of rates could be graphed as a series of different-colored line graphs.
   - By calculating differences between rates in an array and graphing them, you can quickly discern compression and overlapping of slopes and unusual changes in second differences. Differences can be calculated (and then graphed) in three different ways: horizontally, vertically, and diagonally.
5. An array can be compared to other arrays from the same table or to comparable arrays from one or more existing tables.
   - Differences between two arrays could be calculated and graphed. A two-dimensional graph can help you visualize the relationships between the two arrays. Examining one row or column of differences at a time can pinpoint problem areas.
   - Two arrays can be compared by calculating a function of each array's rates, such as present values of survival rates, life expectancies for mortality rates, or weighted averages using a common set of weights.
   - The harder aspect may be coming to grips with whether the differences can be satisfactorily explained and, if they cannot be explained, whether the new array or the existing array is suspect.

Note that above deficiencies 2, 3, and 4 require measures of credibility. Confidence intervals for count-based rates are readily calculated from typical study data, as shown in **Appendix C, Table of Standard Deviations and 90% Confidence Intervals**.

When the rates underlying a large section of an array have very wide confidence intervals, a change in approach may be called for. An approach that groups more data before calculating rates can add to the credibility of the resulting rates.

6.1.1 Preservation of Events

"Event" refers to the event under study, such as death or recovery from disability.

Observed rates are calculated as the number (or amount) of observed events divided by the number (or amount) exposed. As observed rates are modeled, graduated, interpolated, and adjusted, it is reasonable to demand that, in aggregate or at lower levels, the resulting rates reproduce the underlying number (or amount) of events.

To check preservation of events, adjusted rates, when multiplied by their exposures, must reproduce the total number (or total amount) of events underlying those rates. An equivalent approach is to require that the difference between observed and adjusted rates, when weighted by exposure and then summed, is equal to zero.

Preservation of events can be enforced at one or more levels, such as:

- Separately for each row or groups of rows in an array.
- Separately for each column or groups of columns in an array.
- For the whole array.
- Separately for groups of arrays, such as for all-female arrays combined and for all-male arrays combined.
- For the entire table, i.e., for all arrays combined.

6.1.2 Creating Acceptable Rates

Rates identified as being deficient must be replaced by acceptable rates. Creating acceptable rates can be challenging because the reviewer may have to address multiple types of deficiencies while being careful not to distort the overall level of rates. The created rates must typically:

1. ensure "preservation of events,"
2. ensure continuity with surrounding non-deficient rates, and
3. ensure consistency with rate patterns exhibited by a related array or table.

One approach to simultaneously achieving these three requirements is shown in **Appendix G, A Method for Creating Acceptable Rates**, which applies a pattern of rates from a related array, while ensuring continuity with non-deficient rates and "preservation of events." **Chapter 5, Extension of Rates**, discusses extension of rates for the youngest and oldest ages.

6.2 Enforcing Relationships

The enforcement process involves checking for and enforcing relationships between adjacent rates and between arrays of rates. The relationships to be enforced are typically agreed to at the outset of the project. Each time rates are revised, or new rates are calculated, the enforcement process must be repeated.

The checking process can span multiple dimensions, as illustrated by the following excerpt from the 2015 VBT Report:

> **"B. Heuristic Monotonicity Constraints for Mortality Rates**
> While the Team considered multiple approaches to graduating the mortality rates in order to have the best fit with the underlying data and the Team's prospective view on mortality, none of the graduation methods considered could easily handle constraints on the model outputs such as the slope of the mortality rates. Therefore, the Team considered the following reasonable expectations to be constraints that the final select model should meet:
>
> - Above attained age 30, mortality rates should not decrease as issue age increases for the same duration, gender, and smoker status (vertical constraint);
> - Above attained age 30, mortality rates should not decrease as duration increases for the same issue age, gender, and smoker status (horizontal constraint);
> - Mortality rates should not decrease as duration increases for the same attained age, gender, and smoker status (diagonal constraint);
> - Mortality rates for males should not be lower than those for females for the same issue age, duration, and smoker status; and
> - Mortality rates for smokers should not be lower than those for nonsmokers for the same issue age, duration, and gender.
>
> In certain cases, the Team made adjustments based on their judgment to correct any violations of the above constraints. A listing of all the adjustments is shown in Appendix YY."

6.2.1 Define Relationships to Be Enforced

Before any rates are reviewed for enforcement, it is important to reach a meeting of minds as to the relationships between adjacent rates and between arrays of rates that are to be checked for and enforced, as well as an understanding of the goals and limits of rate review. If perfection is the goal, then rate review and adjustment can become a never-ending circular process.

An up-front understanding should be reached on what is expected, generally acceptable or required. For example:
- List any consistency requirements, such as "Male mortality rates should never be less than female mortality rates."
- Describe any monotonicity requirements, such as "Mortality rates should be monotonically increasing where rates commonly decrease, such as for several years after age zero and for several years after a relative maximum near age 20."
- Describe the desired level of smoothness, such as "Visually smooth when viewed on a half-page logarithmic graph."
- List any special requirements, for example:
    - Each array of mortality rates should closely reproduce total death claims incurred by the business contributing to the array.
    - No smoker or nonsmoker rates prior to age 18.
    - Smoker and nonsmoker rates should converge starting at age 86 and finishing by age 95.

A related topic is "Whether to round rates and, if so, to how many decimal places?" If rates will not be rounded, another question is "How many decimal places will be shown when rates are published?" A final question is "To how many decimal places must relationships be enforced?" For example, is a violation worthy of correction if it is not visible in the last decimal place shown or if it amounts to less than a 1 in 10,000 (0.01%) difference?

6.2.2 Define When Relationships Will Be Checked

There are many points at which relationships between rates could be checked. On one hand, it would be good to perform relationship checks early in the process to nip problems in the bud. On the other hand, it might minimize rework to wait until rates are close to final before checking and enforcing relationships.

Arrays of rates would typically have their relationships checked at many points in time:

- Relationships within an individual array of rates would typically be reviewed after any of the following occur:
    - The array's otherwise final rates, whether modeled, graduated, or interpolated, were creatively reviewed and adjusted.
    - Rates that were thought to be final are replaced by newer rates. This can happen more than once.
    - New arrays are created from an existing table, such as by:
        - Multiplying by projection factors for improvement or deterioration,
        - Multiplying by valuation loading factors,
        - Creating gender-blended tables (such as male/female blends of 80/20, 60/40, 50/50, 40/60, and 20/80), or
        - Creating an ALB table from an ANB table, or vice versa.
- Relationships between arrays of rates (such as between genders and between underwriting classes) would typically be reviewed after the relationships within each individual array were reviewed and, if necessary, adjusted.

6.2.3 Create Spreadsheets to Check Relationships

Armed with a list of required relationships or "heuristic monotonicity constraints," as well as with guidance as to any minor infractions that can be ignored, you will want to construct spreadsheet tools to automate the checking process for two reasons:

- It is very difficult, if not impossible, to accurately perform a visual check of relationships between thousands of rates, much less hundreds of thousands of rates.
- You will likely need to repeat the process each time new rates are developed, or existing rates are updated or adjusted.

6.2.4 Checking Select and Ultimate Mortality Rates

For a two-dimensional array that varies by issue age and policy year, rates would typically be examined along the horizontal axis (constant issue age, increasing policy year), along the vertical axis (constant policy year, increasing issue age), and along the diagonals (constant attained age, with offsetting changes in issue age and policy year).

For each issue age, mortality rates will generally increase with increasing policy year, except where attained age mortality rates are declining, as they often do for young adult ages. For products with significant anti-selection, such as for guaranteed issue or preneed products, mortality rates may decrease with increasing policy year in the early policy years.

For each policy year, mortality rates will generally increase with increasing issue age, except where attained age mortality rates are declining.

Mortality rate patterns along the diagonals tend to be more predictable. For rates with the same attained age, those with a lower policy year and a higher issue age generally have lower mortality rates. At the later policy years for some issue ages, rates along the diagonals can become almost constant, signifying a premature end to the select period. In such cases, it is not unusual for the rates along the diagonal to have very slight ups and downs that may need to be smoothed out.

6.2.5 Adjusting Rates to Enforce Relationships

Finding the rates that violate the required relationships is only half the battle. Once found, you will need to adjust the rates to comply, while being cognizant of two dangers:

- By fixing one relationship problem, you may be creating another relationship problem.

- Fixes to many rates, if all in the same direction, could materially raise or lower the average rate.

When fixing one problem begets another problem, it may be time to question the level at which relationships are being enforced. For example, does it really matter if there is an inconsistency in the last decimal place shown?

When fixing many rates, you should use a method that increases some rates while decreasing others. For example, if a sequence of several rates should be monotonically increasing, but instead has some ups and downs, you could adjust the rates in several different ways:

1. You could determine the average rate for the rates needing adjustment, as well as a slope based on the rates immediately preceding and following the rates needing adjustment. You could then calculate new rates based on the average rate and slope, which may not mesh with the rates immediately preceding and following.
2. You could sort the rates needing adjustment into order by increasing amount.
3. You could replace the rates needing adjustment with a linear interpolation between the rates immediately preceding and following the rates needing adjustment, which may not reproduce the average of the rates needing adjustment.

Section 7: Projecting Future Rates

Since tables are developed based on experience from the past, tables are most accurate when applied to the past. However, tables are often used to estimate future events. Rates change over time and future rates are difficult to predict. Even so, attempts at projecting future rates are often more accurate than using rates based solely on experience.

Some events seem to follow long-term trends. For example, mortality rates improved greatly during the 20th century, although the 1918-1919 influenza epidemic and two world wars caused significant but temporary upticks in mortality. The rate of 20th century mortality improvement varied greatly by country, calendar year, age, and gender, making it difficult to use past trends to forecast future trends. In some cases, mortality improvement was more pronounced for certain year-of-birth cohorts; this has been referred to as the "cohort effect."

While most actuaries, demographers, and health professionals anticipate the trend toward longer life spans will continue, there is growing concern about increasing adverse effects from obesity, sedentary habits, opioid use, and pollution.

7.1 Data Considerations for Projecting Future Rates

For the projection of event rates into the future, it is very important to have a source of historical data that meets both of the following criteria:

- The data is based on a consistent underlying population.
- The data is of consistent quality over time.

This is often a difficult hurdle to overcome. For example, in the U.S., the only source of mortality data with consistent quality over very long periods of time is population mortality data published by the U.S. government. However, mortality trends for insured populations have been markedly different from those of the general population, as have mortality trends for various population subgroups.

In addition, year-to-year changes in population mortality tend to be quite volatile, making it difficult to separate the "signal" from the "noise." Grouping of data, such as using five-year averages or five-year moving averages, can be used to reduce, but not eliminate, the noise.

7.2 Connecting Cause and Effect

When past changes in event rates are analyzed at a more detailed level, it is sometimes possible to connect past changes with their causes. For example:

- Liberalization of disability insurance benefits and underwriting requirements have been linked to significant increases in disability rates.
- The availability of new or alternative products that credit higher interest rates is often associated with higher surrender rates for products that credit lower interest rates.
- The widespread introduction of different individual life insurance premium rates for nonsmokers and smokers, starting in the U.S. in the late 1970s, led to higher mortality for products that offered aggregate rates, i.e., the same premium rate for nonsmokers and smokers. Aggregate rates became popular for smokers and unpopular for nonsmokers. The same "market anti-selection" process was repeated with the widespread introduction of

preferred underwriting classes in the early 1990s and, later, refinements involving multiple preferred classes.

While such analysis can be quite helpful, there are always multiple changing conditions operating at the same time. It is difficult to impossible to isolate the effects of multiple changing conditions. For example:

- The U.S. mortality rate for heart disease declined significantly from 1985 until 2015, driven by factors such as a reduction in tobacco usage, the increased use of statins to lower cholesterol levels, and widespread improvements in cardiac procedures, including catheterization, stents, and bypass surgery.
- During the same period, the U.S. experienced an obesity epidemic, with the percentage of overweight, obese, and morbidly obese people soaring to over two-thirds of the population. In recent years through 2017, burgeoning opioid overuse added significantly to mortality rates for young to middle-age adults.
- By 2015, the tide had turned: The U.S. mortality rate for heart disease rose for the first time. Reductions in heart disease and smoking, though still spreading, had already reached the vast majority of the population. Meanwhile, the rising use of opioids and the long-term effects of obesity were increasingly taking their toll. Had there been an ability to isolate the various competing effects, the turning of the tide in 2015 might have been predictable.

7.3 Three Types of Projections

It is typical to designate the "as of" date of the table being constructed as the middle year of the data underlying the study. When rates are projected beyond that "as of" date, one or more of the following processes may be used to project rates into the future:

1. **A projection of rates from the midpoint of the historical data to the time the rates will first be used**, such as the first year for new insurance policies, new annuities, new pensions, or new forecasts. This type of projection can be based on recent trends that often cover a good part of the projection period. The table resulting from such a projection will have the same dimensions as the table before projection. This type of projection was used by the SOA to create the 2015 Valuation Basic Table, which was based on unloaded mortality rates projected from the midpoint of the study data to 2015.

2. **A projection of rates beyond the time when rates will first be used, assuming each rate has a single rate of improvement or deterioration that applies to it in every future year**. A different annual percentage change can be assigned to every rate in the table, effectively doubling the size of the table rather than adding a dimension of time. While simpler to create and use than a table with an added dimension of time, the results are highly constrained by each rate having a single, unchanging trend factor. For example, Scale AA, developed in conjunction with the 2000 Retirement Plan (RP-2000) tables, had a single improvement rate of 0.014 for male age 65, which meant the male age 65 mortality rate for each year was calculated as 0.986 times the prior year's rate, ad infinitum.

3. **A projection of rates beyond the time when rates will first be used, assuming each rate has a rate of improvement or deterioration that can vary by future year.** The table resulting from such a projection has one more dimension than the table before projection, due to the added dimension of time. More recently, projected mortality tables have added a calendar-year dimension to the original table. This more complex process attempts to better model observed mortality trends such as cohort effects. Depending on the intended use of the projected rates, some conservatism may be added to the projection, such as by assuming that a positive trend will shrink over time.

Examples of the third type of projection include the following:

a. Continuous Mortality Investigation (CMI), a UK-based organization that carries out research into mortality and morbidity experience, first reported cohort effects in December 2002 in their Working Paper 1. Since then, they have developed sophisticated projections that include a cohort, or year of birth, element. https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/cmi-working-papers/mortality-projections/cmi-wp-1

b. A report titled "Global Mortality Improvement Experience and Projection Techniques," sponsored by the SOA, was published in 2011. https://www.soa.org/research-reports/2011/research-global-mortality-improve/

c. In 2012, the SOA published projection Scale BB, a two-dimensional approach to projecting pension mortality improvements that incorporated the advanced version of the CMI model. https://www.soa.org/experience-studies/2012/research-mortality-improve-bb/

d. In 2014, the SOA began publishing annual mortality improvement scale updates with MP-2014. The most recent update, MP-2017, showed U.S. population mortality increased 1.2% between 2014 and 2015. https://www.soa.org/experience-studies/2014/research-2014-mp/; https://www.soa.org/experience-studies/2017/mortality-improvement-scale-mp-2017/

e. The Canadian Pension Mortality table, published by the CIA in 2014, incorporated two sets of attained age, gender-distinct improvement rates: One set for 2011 and another set for 2031, with the improvement rates for intervening years calculated using linear interpolation. http://www.cia-ica.ca/docs/default-source/2014/214013e.pdf

Section 8: Financial Impact

8.1 Considerations

The financial impact of an industry table could be examined at an industry, company, or pension plan level. The financial impact of a company or pension plan table would be examined at the company or pension plan level. The following questions should be considered when estimating the financial impact of a newly developed table.

Questions for life, health, or annuity products:

- At an industry or company level, will the new table affect reserves used for financial reporting, such as statutory, GAAP, PBR, or IFRS reserves, or minimum statutory nonforfeiture values?
- At a company level, will the new table affect:
    - Best-estimate or economic reserves?
    - Statutory surplus, GAAP equity, or economic capital?
    - Expected cash flows or cash-flow testing?
    - The company's plans, including financial plans?
- At a company level, will the new table be used to set premiums or dividends for new or existing products? If so, what effect will the new table have on premiums or dividends?

Questions for pension products:

- For pension plan sponsors, such as employers, will the new table affect minimum pension contributions required by the government or the annual pension cost that must be reflected in financial reporting?

8.2 Internal Table versus Industry Table in Life Insurance

The financial impact of the new table may be significantly different depending on whether it is an internally-developed table or a table prescribed by regulators, such as, in the U.S., a table prescribed by the NAIC's Standard Valuation Law (SVL) or Standard Nonforfeiture Law (SNL). Regardless of whether the new table is an internal table or a prescribed table, a company should take steps to quantify, or at least understand, its financial implications.

In the U.S., the financial impact of a prescribed table may be limited to statutory minimum reserves and nonforfeiture values as promulgated by the SVL or SNL. A new valuation or nonforfeiture table could also affect the pricing of products for companies that expect products to achieve a specific return on investment based on statutory results. A new industry table could also affect the timing of the expected release of a company's statutory surplus, thereby leading to modifications to its dividend scale.

The impact of a new table on an industry is difficult to model and predict since the impact can vary materially by company. In some situations in the U.S., the SOA or AAA working group responsible for developing a new valuation table will develop an Excel spreadsheet that allows companies to compare reserves and net premiums at a granular level, i.e., for specific policy or claim characteristics, between the new table and the table it is replacing. For example, the working group that developed the 2013 Individual Disability Income Valuation Table provided such a spreadsheet, which allowed companies to

compare active life reserves, claim reserves and net premiums based on the new table and its predecessor, the 1985 Commissioner's Individual Disability A table.

A new internal table may have a larger financial impact on a company than would a new industry table, potentially affecting premiums, dividend scales, cash flows, premium deficiency reserves, model office projections, future earnings, and financial plans.

8.3 Impact on Reserves for Financial Reporting

In the U.S., a new industry valuation mortality table will affect a company's statutory reserves after the table has been adopted by the NAIC and various states. Companies that have robust capabilities for projecting financial results will need to update their projections to reflect the new industry table.

A new industry valuation table may not directly affect GAAP, PBR or IFRS reserves, since these reserves typically incorporate companies' best-estimate actuarial assumptions with some provision for adverse deviations. However, when companies review their best-estimate actuarial assumptions, the new valuation table, net of any loading, may be an important reality check.

8.4 Impact on Best-Estimate Actuarial Assumptions

8.4.1 Life and Health Insurance

A new internal table could lead to revisions in best-estimate actuarial assumptions used for cash-flow testing, GAAP, PBR, and IFRS reserves, premium rates and dividend scales. The new table may require a company to set up or take down statutory premium deficiency reserves and could affect the results of GAAP recoverability testing.

A company should quantify the financial impact of the new table and determine whether that impact is acceptable or manageable. The results may direct a company to revisit the methodologies and assumptions used to develop the new table, especially if the new table is producing unintended consequences.

A company should examine the differences between a new internal table and the one being replaced, as this may create an understanding of the source of any unintended consequences. If the change is unfavorable, the company should investigate whether certain demographic characteristics associated with its business in force are causing the differences or whether they are attributable to company practices such as underwriting or claims management. The results of the investigation could guide the company to revise its target markets or company practices.

8.4.2 Pensions

The creation of a new internal or industry mortality table for pension plans could impact the best-estimate mortality assumption used to calculate the plan sponsor's pension cost and balance sheet liabilities for financial reporting and, in the U.S., Funding Target Liability and Minimum Required Contributions under ERISA (the Employee Retirement Income Security Act of 1974) and PPA (the Pension Protection Act of 2006).

Because the income statement, balance sheet, and cash flows of the plan sponsor will all be affected, actuaries should be sure to quantify the effect of using the new mortality table on these measurements

and ensure that clients are well-aware of the implications of the change. A helpful way to estimate the impact on liability change, if re-running the full plan valuation is too cumbersome, is to compare annuity factors developed using both the old and new tables at various ages. Once the approximate change in liability has been estimated, an actuary can then estimate the approximate change in the plan's pension cost and Minimum Required Contributions.

8.5 Impact on Life and Health Insurance Premium Rates

A new industry table could affect the expected returns associated with future new business from currently issued products and ultimately lead to a change in premium rates or product features. A new internal table is even more likely to affect expected returns and the calculation of premiums for future new business. A new internal table that reduces expected profitability could lead to premium increases. In that case, it is imperative for the company to scrutinize the factors causing higher premium rates and either accept the change or take the necessary corrective actions.

Section 9: Finalization of the Table

Finalization of the table includes the creation of additional tables that can be derived from the developed table, as needed, plus the final stages of important processes that operate throughout the development of a table: Namely, project review, oversight, and documentation.

9.1 Creation of Derived Tables

Depending on what is needed, additional tables can be derived from the developed table for any combination of the following five, and possibly other, dimensions:

1. With or without projected trend factors
2. With or without valuation loading factors
3. For one or more age definitions, such as age last birthday, age nearest birthday, age next birthday or other age variations, such as half-ages or quarter-ages
4. For gender blends of female and male rates, such as 80/20, 60/40, 50/50, 40/60, and 20/80
5. For "relative risk" versions of tables, used to accommodate multiple underwriting classes, such as the Relative Risk tables associated with the 2015 VBT and the Preferred Structure tables associated with the 2017 CSO

When a combination of more than one of the above dimensions is needed, the usual process is to address one dimension at a time. For example, to create a table with both projected trends and valuation loads, you would first create a table with projected trends, and then apply valuation loads to create a table with both projected trends and valuation loads. With this sequential process in mind, we will address common constraints and then address each of the five dimensions separately.

9.1.1 Common Constraints

When deriving a new table, you may need to enforce constraints such as the following:
- Agreed upon minimum and maximum rates
- Agreed upon relationships between rates (See 6.2, Enforcing Relationships)

9.1.2 Projected Trend Factors

Projected trend factors are used to create rates that are more applicable for future use. Projected trend factors are applied by multiplying each of the table's rates by the appropriate projected trend factor.

As discussed in 7.2, Two Types of Projections, a projection for one fixed period results in a new table with the same dimensions as the predecessor table. However, a projection for a series of future years adds a calendar-year dimension to the resulting table. When combined with a select and ultimate table, such a projection effectively creates a new select and ultimate table for every issue year.

Rather than create such a series of tables, a more common approach would be to use the original table and the projection factors to calculate projected rates on demand. Such an "on demand" approach would need to be enhanced to apply the applicable common constraints.

9.1.3 Valuation Loading Factors

Valuation loading is used when developing valuation tables for life insurance products; the intent is to create higher reserves that will be adequate for a high percentage of companies, including many with worse than average experience. Therefore, for life and health insurance products, loads increase mortality or claim rates while, for annuity products, mortality loads decrease mortality rates. Valuation loading factors are applied by multiplying each rate of the unloaded table by the appropriate valuation loading factor.

Valuation loading factors can range from a single factor to a multi-dimensional array of factors. The dimensions of the valuation loading factors are always a subset of the table's dimensions. As a result, the loaded table will always have the same dimensions as the underlying table.

Loadings are often expressed as the percentage to be added to or subtracted from the applicable rates. For example, a load of 20% for life insurance mortality would translate to a loading factor of 1.20, while a load of 20% for annuity mortality would translate to a loading factor of 0.80. Please refer to **Appendix J, Valuation Loading**, for more information on this topic.

9.1.4 Rates for Additional Age Definitions

In North America, most industry life and annuity tables and all pension tables are developed on an age-nearest-birthday (ANB) basis. ANB rates are then used to create rates on an age-last-birthday (ALB) basis, where desired. Both ALB and ANB age bases are used in health insurance tables according to factors such as how underlying health insurance policies were typically priced, the age basis of other tables used, and general industry practices.

Appendix H derives the formulas for two methods of converting rates from ANB to ALB and from ALB to ANB. One method, the "UDD conversion method," based on the assumption of a uniform distribution of deaths, has long been used by the SOA for life insurance mortality table development. The other method, the "Geometric conversion method," uses geometric averages, is simpler and more intuitive, and produces comparable results.

These two conversion methods were compared based on their round-trip error, i.e., the difference between the original rate and the rate resulting from converting the rate from ANB to ALB and back again. The one-way error, i.e., the error from a single conversion, from ANB to ALB or from ALB to ANB, is assumed to be half the round-trip error. The findings were interesting:

> Below age 90, the UDD conversion method produced errors, as a percentage of mortality rates, that averaged 0.01% less than those from the Geometric conversion method, i.e., an average improvement of 1 part per 10,000. For ages 90 and above, the results were reversed, with the average error for the UDD conversion method averaging 0.4% larger than that for the Geometric conversion method, a difference of 1 part per 250. This result is not surprising, as the assumption of uniform deaths becomes increasingly inaccurate as mortality rates climb past 0.1.

9.1.5 Gender-Blended Rates

Some jurisdictions consider the use of rates that vary by gender as discriminatory. To comply with such jurisdictions, it is commonplace to develop and use tables that reflect the expected or actual mix of males and females. For example, to create a table with a 60/40 blend of female and male rates, each blended rate would be calculated as 60% of the female rate plus 40% of the male rate.

9.1.6 Relative Risk Rates

Relative risk rates result from splitting the rates for an underwriting group, such as female nonsmoker, into multiple underwriting classes, such as female nonsmoker preferred plus, female nonsmoker preferred, and female nonsmoker standard. Typically, each set of relative risk rates would be calculated by applying a grid of percentages to the rates for the underwriting group. Separate grids of percentages would typically be developed for every combination of gender, smoking habit, and relative risk category.

Relative risk rates, when needed, are usually generated before any of the other four types of rates are generated.

9.2 Project Documentation

The documentation process starts at the outset of the table development project. All key findings and important decisions should be documented as they occur. Unless significant, it is not usually necessary to document a dead end, where an idea was pursued and then discarded. However, every step that led to the final tables should be documented. The narrative should convey how each important step led to the next step. Attention should be given to aspects that were out of the ordinary or where unexpected results were encountered.

Documentation should be organized and summarized with two key audiences in mind:

1. A peer review at the end of the project, whether by an internal resource for a company project or by an outside expert for an industry project.
2. Future review of the table development project to better understand the results, or to apply aspects of the project to a future project.

9.3 Project Oversight

Project oversight starts at the outset of the table development project. The oversight should be focused on pursuing and balancing two often-conflicting goals:

1. Delivering a quality result and
2. Completing the project in a timely fashion.

For a company project, oversight may be as simple as one or more people keeping their boss up to date on the project's progress, problems, and key decisions. Ideally, the person performing project oversight has experience with table development and can add value along the way, such as by clearing roadblocks and suggesting alternatives.

For an industry project, it is common to form a team of people with appropriate expertise drawn from several companies. Team members often focus on different parts of the table development process, while keeping all team members up to date on challenges and decisions. Such a team can be difficult to

manage, because most team members have demanding jobs that keep them very busy, making it hard for them to find time to contribute to the project. A key aspect of project oversight for such a team is recognizing and addressing the parts of the project that are not getting the attention they need.

Whether a company or industry project, a good practice is for the project team to review results, conclusions, and decisions with those responsible for project oversight at the end of each stage of the table's development.

Appendix A—Introduction to Exposure

For a complete treatment of exposure, please refer to the Society of Actuaries' 2016 paper entitled "**Experience Study Calculations**."

Annual Rates, Monthly Rates and Exposures

Most experience studies develop **annual rates**, which are typically calculated as O divided by E where:

O = The total number of occurrences of the event under study during the year

E = Annual Exposure

Annual exposure, also known as "actuarial" exposure, is calculated as follows. Note that "year" could refer to a calendar year, year of age, policy year, or plan year:

1. Lives that persist through the current year are exposed for the full year, resulting in an exposure of 1.

2. Lives with an occurrence of the event under study during the current year are exposed for the full year, resulting in an exposure of 1.

3. Lives that terminate during the current year for any reason other than the event under study are assigned exposure equal to the fraction of the year before termination.

The above description does not address what to do when the current year crosses the study start or stop date, which creates partial years of exposure that can distort results. This issue is examined in detail in a 2017 SOA paper titled "**Experience Study Rate Errors**."

Annual exposure adjusts for the study start and stop dates as follows:

1. Lives are not exposed prior to the study start date. This results in partial years of exposure for all lives in force at the study start date.

2. Lives are not exposed beyond the study stop date, with one exception: Lives with an occurrence of the event under study in the partial year prior to the study stop date are exposed for a full year, resulting in an exposure of 1.

A variation of annual exposure, called "distributed exposure," uses a different approach to calculating exposure for lives with an occurrence of the event under study that occurred during the partial year preceding the study start and stop dates. Distributed exposure may have some advantages when results are to be studied by calendar year.

The calculation of annual exposure implicitly assumes the Balducci Hypothesis, i.e., that the partial year rate from any date during the year to the end of the year is proportional to the full-year rate:

$$_{1-t}q_{x+t} = (1 - t)q_x$$, where "x" is the age at the beginning of the year

and "$1 - t$" represents the partial year remaining

Although this assumption results in rates for a fraction of a year that increase as the year progresses, the annual exposure method accurately calculates rates when used for full years. However, when used for partial years at the beginning and end of the study period, material distortions emerge under the following conditions:

1. The rate is 10% or more and

2. The annual increase in rates is significantly less than the rate itself.

For example, if the true monthly rate were a flat 1.05413%, equivalent to a 12% annual rate, annual exposure, which uses the Balducci Hypothesis to calculate partial year rates, would:

- Overstate the rate for the partial year at the beginning of the study by 2%, i.e., 102% of the 12% annual rate.
- Understate the rate for the partial year at the end of the study by 2%, i.e., 98% of the 12% annual rate.

However, if the true monthly rate were increasing at an 12% annual rate, there would be no errors. If the true monthly rate were increasing at an 8% rate, the errors would be about one-third of those shown above. Additionally, when applied to a study consisting of many years of experience, the errors may be largely offsetting.

Some experience studies, such as lapse and disability studies, employ **monthly rates** when rates change significantly from month to month. Monthly rates are typically calculated as the total number of events under study that occurred during the month divided by **monthly exposure**. Monthly exposure is calculated identically to annual exposure with "month" substituted for "year" in the above description of annual exposure.

Daily Rates and Exposure

When the annual rate is greater than 10%, is assumed to be continuous, and is increasing exponentially at an annual rate significantly less than the rate itself, the use of annual exposure will understate the rate. The use of daily rates and daily exposure can produce more accurate results in this and other cases. Also, daily rates are more readily applied to short periods of time measured in days.

The average **daily rate** can be calculated as O divided by E, where:

O = The number of occurrences of the event under study during the rate period, which could be a month, a year or some other period.

E = Daily exposure

Daily exposure is calculated for each life as the number of days for which each life was exposed to the event under study during the rate period:

1. For persisting lives, daily exposure would equal the number of days in the rate period. If the rate period is a calendar year, the number of days could reflect leap years and non-leap years, or use an average of 365¼ days.

2. For terminating lives, including terminations due to the event under study, daily exposure would equal the number of days from the beginning of the rate period to the termination date.

The cumulative daily rate for a period of X days can be calculated as $1 - (1 - \text{daily rate})^X$.

In practice, daily rates are virtually equal to the force of decrement, as there is very little discernable difference between daily and continuous compounding.

Fractional Rates and Exposure

Daily rates are the **fractional rates** with the shortest possible period useful for insurance and pensions. Fractional rates can be calculated for other periods, such as quarterly and monthly.

Unlike the monthly rates discussed previously, which vary from month to month, a fractional monthly rate applies for a 12-month rate period. (In practice, fractional monthly rates are rarely, if ever, used.) The rate period for a fractional rate would typically be a full year.

The fractional rate for a rate year is calculated as O divided by E, where:

O = The number of occurrences of the event under study during the rate year.
E = Fractional exposure

Fractional exposure is calculated as the number of fractional periods to which each life was exposed to the event under study during the rate year:

1. Lives that persist for the full rate year are exposed for a full year, resulting in a fractional exposure equal to the number of fractional periods per year. For a fractional quarterly rate, a full year's exposure would be 4.

2. The event under study is not always a termination. For example, the event could be a utilization of medical services, a partial withdrawal or an election of an option. Typically, such events would either receive a full year's exposure or, if they were to terminate for any reason, would receive exposure through the end of the period in which they terminated.

3. Lives that terminate during the year due to the event under study are exposed to the end of the fractional period in which the termination occurred. Fractional exposure would equal the number of full fractional periods for which the life was in force plus 1 to expose the life to the end of its last fractional period. For a life that terminated due to the event under study halfway through its third quarter, its quarterly fractional exposure would be 3.

4. Lives that terminate for any reason other than the event under study are exposed to their date of termination. Fractional exposure would be the number of full fractional periods for which the life was in force plus the fraction of the period the life was in force during the period in which it terminated. For a life that terminated halfway through its third quarter, for any reason other than the event under study, quarterly fractional exposure would be 2.5.

The probability of surviving for one fractional period is $1 - \text{fractional rate}$. The cumulative rate for X periods is $1 - (1 - \text{fractional rate})^X$.

Appendix B—Select Period Considerations

Individually-underwritten insurance often experiences lower claims because of the underwriting process; underwriters are said to "select" the better risks. While some of the effects of selection may linger on for decades, most of the effects wear off over a much shorter period. For products that lack individual underwriting, there can be an opposite effect: Anti-selection can result in much higher than normal first year claims, with part of the anti-selection effect wearing off over a few years.

"Select" rates refer to rates that vary by issue age and policy year. "Ultimate" rates refer to rates that vary by attained age. The "select period" is the last policy year for select rates, after which ultimate rates kick in.

As the select period is lengthened, the added select rates are increasingly less credible, since fewer policies persist as each policy year go by. However, the discontinuity between select and ultimate rates is reduced by lengthening the select period.

When select periods were first introduced for individually-underwritten life insurance products, both genders, and all adult issue ages, used the same select period. The tendency was to truncate the select period prematurely for the older issue ages. This resulted in large discontinuities at the end of the select period for the older issue ages.

Over the years, the trend has been toward reducing the select to ultimate discontinuity by stretching out the select period. The less credible select rates in later years can be smoothed by grading between earlier, more credible select rates and ultimate rates at the end of the select period. At the same time, a goal of reproducing the study's total claims can be enforced. This has resulted in select periods that can vary significantly by gender and issue age.

Appendix C—Table of Standard Deviations and 90% Confidence Intervals

Standard Deviation as a Percentage of Observed Rate ("q")
(See Note 1 below)

Number of Events or Decrements:

q	10	30	100	300	1,000	3,000	10,000	30,000	100,000
0.001	31.6%	18.2%	10.0%	5.8%	3.2%	1.8%	1.0%	0.6%	0.3%
0.01	31.5%	18.2%	9.9%	5.7%	3.1%	1.8%	1.0%	0.6%	0.3%
0.1	30.0%	17.3%	9.5%	5.5%	3.0%	1.7%	0.9%	0.5%	0.3%
0.3	26.5%	15.3%	8.4%	4.8%	2.6%	1.5%	0.8%	0.5%	0.3%
0.5	22.4%	12.9%	7.1%	4.1%	2.2%	1.3%	0.7%	0.4%	0.2%

90% Confidence Interval (+ and -) as a Percentage of Observed Rate ("q")
(See Note 2 below)

Number of Events or Decrements:

q	10	30	100	300	1,000	3,000	10,000	30,000	100,000
0.001	52.0%	30.0%	16.4%	9.5%	5.2%	3.0%	1.6%	0.9%	0.5%
0.01	51.8%	29.9%	16.4%	9.4%	5.2%	3.0%	1.6%	0.9%	0.5%
0.1	49.4%	28.5%	15.6%	9.0%	4.9%	2.8%	1.6%	0.9%	0.5%
0.3	43.5%	25.1%	13.8%	7.9%	4.4%	2.5%	1.4%	0.8%	0.4%
0.5	36.8%	21.2%	11.6%	6.7%	3.7%	2.1%	1.2%	0.7%	0.4%

Note 1:

The binomial mean for rate q and exposure n (the number of trials) is nq.

The number of events or decrements, d, is equal to the rate times exposure; that is, $d = nq$.

The binomial variance is npq, where $p = 1-q$. Standard deviation is SQRT(nq) = SQRT(dp).

The standard deviation as a percentage of the mean is SQRT(npq)/nq = SQRT(p/nq) = SQRT(p/d).

If q is small, the standard deviation can be approximated as 1/SQRT(nq) or 1/SQRT(d).

Note 2:

The 90% confidence interval is approximated using the Normal distribution, which results in a confidence interval that is plus and minus 1.645 standard deviations.

Appendix D—Using Howard's Whittaker-Henderson Functions

R.C.W. (Bob) Howard, FSA, FCIA, has developed functions that perform Whittaker-Henderson (W-H) graduation. W-H graduation functions for Excel and VBA, as well as their related documentation, are available at http://www.howardfamily.ca/graduation/index.html. The purpose of this appendix is to supplement Howard's documentation for those new to W-H and Excel arrays.

For those comfortable with VBA, Howard recommends the VBA version of his graduation functions, which he finds much more straightforward, easier to read, and with none of the difficult to use Excel arrays.

It is recommended that you read and reread at least the first three sections of Howard's documentation, "WHGrad.doc," paying special attention to Section 2, Lowrie's Variation. This is only about two pages of material.

Downloading Howard's .dll File

When downloading files from Howard's site, the .dll file should be saved to your C:\Windows directory. If you cannot save it directly, try the following sequence:

1. Download the .dll file
2. Left click on the download button on your browser to show a list of downloaded files
3. Right click on the .dll file and right click the "Open Containing Folder" choice, which will give you a list of downloaded files
4. Left click on the .dll file and right click the "Copy" choice
5. Left click on the C:\Windows directory and then right click the "Paste" choice

Using Excel Arrays

Howard's graduation tool comes in the form of an Excel array function, which can be difficult to work with. With one wrong keystroke, you can botch the input of a long, complex series of key strokes needed to create an array of graduated rates. You must repeat the process every time you change the size of the array of rates to be graduated. A solution to this problem is to store the series of keystrokes in a nearby cell, then copy and paste them each time you need to recreate the array.

You can't store the exact character string, as Excel will try to use it as a function. Instead, store the string with an "x" placed immediately before the equal sign. This will allow the string to be stored and the "x" will remind you that it needs to be left behind when copying and pasting.

As an example, suppose you stored the following character string: "x=gradrange(mort, wtdexp, order, h, grate)" where:

- "x" is a dummy value used to stop Excel from interpreting this as an array function.
- "gradrange" is the name of Howard's array function for graduating rates.
- "mort" is the name of an array of input mortality rates that are to be graduated.
- "wtdexp" is the name of an array of normalized exposure weights, calculated as each rate's exposure divided by total exposure and multiplied by the number of rates being graduated. Normalized exposure has two advantages:
  - Total deaths and the average age at death are preserved.

- o It roughly equalizes the size of the fit and smoothness factors required to balance fit and smoothness.
- "order" is the name assigned to the order of the differences used when calculating smoothness, measured as the sum of the squares of the nth differences (n = order).
- "h" is the name of the input parameter that adjusts fit versus smoothness: lower values of h give more emphasis to fit, while higher values of h give more emphasis to smoothness. In the limit, as h approaches infinity, graduated values will be the least squares fit of a polynomial of order n-1 to the input rates and the sum of the squared n^{th} differences will be zero.
- "grate" is the name of the constant geometric rate at which a component of the rates increases; this can be useful for middle age and older age mortality rates, which tend to roughly follow a geometrically increasing pattern.

Notes:

1. Rather than define names for each of the above, you can specify cell locations. However, names are easier to reuse and may be subject to fewer mistakes.
2. Two-dimensional graduations have separate horizontal and vertical input parameters for order, h, and grate.

Because it is so easy to make a mistake, the following spells out, in excruciating detail, the steps to cut and paste a stored character string to create an Excel array:

1. Click on the cell with the stored character string, using your cursor and the shift key to highlight all but the leading "x", then press Ctrl C to copy.
2. Press the Esc key. This will get you out of the previous cell. Very important!
3. Use your cursor and shift key to highlight the entire range of the array to be created.
4. Click the editing box above the columns, then press Ctrl V to paste the character string there. This will highlight the array and its input items in different colors. The first time you see this, you should make sure the colored boxes match up with your intended input.
5. Hold down the Shift and Ctrl keys, then press the Enter key.
6. Your array should now be fully populated with graduated rates!

Balancing Fit and Smoothness

It is crucial to strike a proper balance between smoothness and fit. Confidence intervals can be employed to test whether graduated rates have been overfit or underfit to the observed rates:

> For example, if a two-dimensional graduation involved 200 rates, you would expect on average that 180 observed rates would fall within their 90% confidence intervals. If 195 or more observed rates fell within their confidence intervals, you might strongly suspect the graduated rates were overfit to the observed rates. If only 160 observed rates fell within their confidence intervals, you might strongly suspect the graduation oversmoothed the observed rates.

Another useful tool is to graph both the input rates and the graduated rates, then visually check smoothness versus fit. If working with rates like mortality rates that can range through several orders of magnitude, it is best to use a logarithmic graph or show rates as a percentage of an existing table. When performing a one-dimensional graduation, it is easy to view input and graduated rates on one graph. A two-dimensional graph is usually too noisy to review. A better approach is to set up a one-dimensional graph that allows you to view input and graduated rates for one horizontal or vertical cross-section at a

time. You could also graph the differences between input and graduated rates. Visually, your goal should be to smooth out most of the irregularities in the input rates, while preserving the overall shape implicit in the input rates.

Graduation Strategies

The smoothness component of a Whittaker-Henderson graduation of order n seeks to minimize the sum of the squares of the nth differences of the graduated rates. By using successively higher values of h, the graduated rates will converge on the polynomial of order n-1 that is the least squares fit to the input rates; a polynomial of order n-1 has n^{th} differences of zero. For example, if order were n=2, the result of using a very high value of h would be that graduated rates would follow a straight line, which has second differences of zero.

Using a higher order gives the graduation more degrees of freedom to achieve a balance between smoothness and fit. On the other hand, the use of a lower order sometimes results in a shape that more naturally fits the rates.

A full range of mortality rates from age 0 to age 100 can be difficult to graduate, as the youngest and oldest ages often suffer from a lack of credible data. Therefore, a common strategy is to focus the graduation on the middle ages, such as ages 30 to 90, and use other methods for the youngest and oldest ages, including, perhaps, separate graduations.

When performing a two-dimensional graduation, you will want to experiment with different combinations of polynomial orders. For a select and ultimate mortality graduation, you may find that you need a 4th-order polynomial to match the complex pattern by advancing age, while a 2nd- or 3rd-order polynomial may be sufficient to match the select pattern.

When working with two-dimensional graduation, you may want to balance smoothness so both dimensions have about the same smoothness score. However, if one dimension is larger than the other, it might make sense to target smoothness scores proportional to the dimensions.

Grouping, such as by age groups, may be used to increase the credibility of observed rates. However, grouping of mortality rates can be problematic because many mortality tables exhibit geometric increases of 5-10% per year of age. It may be possible to determine an approximate rate of geometric increase for each age group. Then, data for the ages within each age group might be combined in a manner that reflects the group's geometric average rate.

When grouping, you should be aware that Howard's graduation array function requires a minimum number of rows or columns of observed rates, which is based on the order of the differences used to calculate smoothness for that dimension. The minimum number of rows or columns is equal to two times the order plus one.

Observed rates in the top and bottom few rows (and sometimes in the left-most and right-most columns) may have low credibility resulting in extreme rates that distort the graduation of more credible rates that are nearby. This is because rates with low credibility get the same smoothness weight as rates with high credibility. Where nearby rates are significantly distorted by extreme rates, you should consider removing the responsible rows or columns from the graduation.

Graduation Production

Experience study summary data, described at the end of Chapter 2, can be used to populate an Excel pivot table. The pivot table can then be used to create graduation input for any desired array. At a minimum, graduation input should include observed rates and the exposure for each rate. Event count and amount should also be input so you can examine how closely graduated and adjusted rates reproduce actual events.

An additional data item that is very useful for graduation is the sum of amount squared times exposure, calculated per life, which enables the calculation of confidence intervals for amount-based observed rates. See 4.6, Variance for Amount-Based Observed Rates.

To automate the process, you can structure your pivot table to produce rows and columns of data that are designed to be exported to a graduation workbook through one big cut and paste. You can then build your graduation workbook around the "cut and pasted" data. The graduation sheet can be expanded to include validity tests, such as how closely graduated and adjusted rates reproduce actual events.

A valuable tool to add is a grid that shows each outlier, defined as a graduated rate that falls outside a confidence interval placed around its observed rate. If the confidence interval (CI) is expressed as plus and minus CI, then showing a grid of Outlier Percentages, as calculated below, will let you see at a glance where outliers are barely outside or far outside the confidence interval:

>   If Observed Rate > Graduated Rate + CI,
>   Then Outlier Percentage = (Observed Rate – Graduated Rate – CI)/CI
>   Else If Observed Rate < Graduated Rate – CI,
>   Then Outlier Percentage = (Observed Rate - Graduated Rate + CI)/CI
>   Else Outlier Percentage = blank.

Additionally, the number of outliers can be used to calculate the overall percentage of rates outside their confidence intervals, which can be compared to the expected percentage of outliers. For example, if 40 out of 200 rates fell outside their 90% confidence intervals, 40 actual outliers would be compared to 20 expected outliers, resulting in a ratio of 200% of expected. Ratios significantly above 100% indicate oversmoothing. Ratios significantly less than 100% indicate overfitting.

As you work on graduating each new set of rates, you will find that some sets have more or fewer cells lacking credible data. In many cases, this will force you to resize the graduation array to work with different numbers of rows or columns, a process that is difficult to automate.

Graduation Statistics

Howard supplies a "gradstat" array function that he very briefly documents in item 3 of his documentation: Section 8, VBA Utilities for Graduation. The gradstat array is horizontal, in contrast to the one-dimensional "gradrange" array, which is a vertical column. The gradstat array can be created with a length of three to five elements, allowing you to shorten the array when only three or four elements are applicable.

The first gradstat statistic is the calculated value of the function being minimized. The calculation is based on the next four statistics along with their adjustment (e.g., "h") factors. Formulas for these terms are given in Sections 1.3, 2, and 3 of Howard's documentation. We will refer to these terms as scores; low scores are preferable. The five scores are:

1. Overall score, equal to the sum of the following four scores times their adjustment factors.
2. Input fit score, which measures the fit of graduated rates to input rates; equal to the sum of the squares of the differences between input and graduated rates.
3. Horizontal smoothness score, equal to the sum of the squares of the nth differences, where n = order for horizontal rates.
4. Vertical smoothness score, equal to the sum of the squares of the nth differences, where n = order for vertical rates (only used for two-dimensional graduations).
5. Table fit score, which measures the fit of graduated rates to an existing table, equal to the sum of the squares of the differences between table and graduated rates (only used when a table with matching rows and columns is specified).

When performing a two-dimensional graduation, it is necessary to adjust the h factors for both horizontal and vertical smoothness. If both smoothness measures use the same order, the number of rows is roughly equal to the number of columns, and the same level of smoothness is desired for both, then it may be appropriate to use the same h values for both. In such a case, the h values could be increased or decreased in tandem to achieve the desired balance between fit and smoothness. This is often the case for select and ultimate mortality rates, which tend to be continuous and smooth over both dimensions.

However, for more discontinuous rates, such as lapse rates, the desired smoothness by issue age may be quite different from that by policy year. You may want to use different orders to calculate the two dimensions' smoothness.

Appendix E—Estimating Variance for Amount-Based Observed Rates

When graduating amount-based observed rates, it is very helpful to understand the credibility associated with each observed rate. Observed rates with low credibility should be prevented from distorting results.

Credibility can be inferred from standard deviation, which is the square root of variance. Unfortunately, the calculation of variance for amount-based rates requires data that is not often available at the point at which graduation or modeling occurs.

The observed rate, q_r, is calculated as the sum of the event amount, $EventAmount_x$, for all lives x contributing to observed rate q_r, divided by the sum of exposed amount, $ExposedAmount_x$, for the same lives:

$$q_r = (\Sigma\, EventAmount_x) \,/\, (\Sigma\, ExposedAmount_x)$$

The complement of the observed rate, p_r, also known as the probability of survival when q_r is a decrement, is calculated as $1 - q_r$.

Variance for Amount-Based Rates

The variance contribution to q_r from life x is calculated as:

$$Variance_x = (Amount_x)^2 (ExposedCount_x)\, q_r\, p_r,$$

where $Amount_x$ is the benefit or other amount associated with life x. $ExposedCount_x$ is equal to 1 for a full year of exposure and less than 1 for a partial year of exposure.

Variance for observed rate q_r is calculated by summing the individual variances for all lives x that contribute to the observed rate:

$$
\begin{aligned}
Variance_r &= \Sigma\, Variance_x \\
&= q_r\, p_r\, \Sigma(Amount_x)^2 (ExposedCount_x).
\end{aligned}
$$

While q_r and p_r are readily calculated from event amount and exposure amount, the sum of $(Amount_x)^2 (ExposedCount_x)$ is <u>not</u> often available. However, with foresight, it could be calculated at the seriatim level and accumulated along with event counts, event amounts, exposed counts, and exposed amounts.

The remainder of this appendix is focused on developing an approximation of the variance for an observed rate, assuming subtotals of exposed count and exposed amount are available by size group.

Simple Variance Approximations for Amount-Based Rates

Simple variance approximations can be obtained by assuming average sizes are widely shared:

1. If the average size associated with most observed rates is close to the overall average size, variance for each observed rate could be roughly approximated by assuming that all lives have amounts equal to the overall average size, $AvgSize_{OA}$:

$$Variance_r = (AvgSize_{OA})^2 \; q_r \, p_r \; \Sigma(ExposedCount_x)$$

2. A slightly more refined approach would be to assume that each observed rate's average size, $AvgSize_r$, is the amount for all lives that contributing to the observed rate:

$$Variance_r = (AvgSize_r)^2 \; q_r \, p_r \; \Sigma(ExposedCount_x)$$

Allocating Exposed Counts

A rough estimate of variance can be obtained by assuming the overall average size for the study is the amount for every life in the study. A better estimate of variance can be obtained by assuming the average size for an observed rate is the amount for each life contributing to the observed rate.

If exposure counts and amounts are available by size group, a better estimate of total variance could be obtained by allocating each size group's exposed count between its two most common sizes, while preserving the average size, as follows:

- For each size group, select amounts on either side of the size group's average size that either are known to be or are likely to be the most common amounts in that size group.
  - For example, an individual life insurance size group that ranges from $100,000 to $249,999 typically would have most of its lives concentrated at $100,000, with the second most common amount at $200,000.
- Distribute each size group's exposed count between the two most common amounts such that the size group's total exposed amount and average size are preserved; this sets up two linear equations with two unknowns that are easily solved for. This process creates an approximate distribution by size within each size group.
  - Continuing the above example, if that size group's average size were $110,000, then 90% of the exposed count would be allocated to the $100,000 size and the other 10% would be allocated to the $200,000 size. The weighted sum of amounts squared would be $90\%(\$100,000)^2 + 10\%(\$200,000)^2 = 130\% * 10^{10}$.
  - If the second size were $150,000 instead of $200,000, then 80% would be allocated to the $100,000 size and 20% to the $150,000 size. The weighted sum of these amounts squared would be $125\% * 10^{10}$.

The 90/10 distribution, which used amounts of $100,000 and $200,000, is probably as extreme as the distribution is likely to be, yet it's weighted sum of amounts squared is not that much higher (130% versus 125%) than the more conservative 80/20 distribution, which used amounts of $100,000 and $150,000. If all policies had a size of $110,000, the resulting percentage would be 121%, which is not that much lower than 125% or 130%.

Comparing the results of the 90/10 and 80/20 distributions, 131% and 125%, to the simple estimate based on assuming all amounts were equal to the average size, 121%, we find that the best estimate as

a percentage of the simple estimate would likely range between 103.3%, equal to 125%/121%, at the low end, and 107.4%, equal to 130%/121%, at the high end.

Averaging the low end and high end could yield an improved estimate of variance as 105.4% of the simple estimate. Similar analysis could be performed for observed rates with various average sizes and for multiple size groups. In the end, it may be found that variance can be well-estimated as a percentage of the simple estimate.

Appendix F—Mortality Study of 100 Oldest People

On Wikipedia, a list of the 100 verified oldest people who have ever lived is maintained. You can find it by searching for "100 verified oldest people." For the people who were on the list on December 12, 2017:

- The earliest and latest birth years were 1870 and 1903.

- The youngest and oldest ages were 114 years, 93 days and 122 years, 164 days.

- Only two people had lived beyond the age of 117; one died at 119 years, 97 days; and the other died at 122 years, 164 days.

- The 15 oldest people were females. Only six of the 100 oldest people were male.

- Almost half of the 100 oldest people were from the United States, about one-quarter were from Japan, and about one-fifth were from Western Europe.

- Only eight of the 100 oldest people were alive, ranging in age from 114 years, 118 days to 117 years, 26 days.

- The age of each living person is updated only after they have been verified to still be alive. This process works well for the calculation of exposure.

- A mortality study was performed using daily exposure starting at age 114, 93 days (the youngest age) and ending at age 117, 365 days. The number of deaths and annualized mortality rates by age are shown in the table below:

Age	Deaths	Rate
114	54	0.666
115	24	0.590
116	8	0.482
117	4	0.606
Total	**90**	**0.622**

- The annualized mortality rate for ages 114 to 117 combined was 0.622, with a 90% confidence interval of 0.539 to 0.705. Surprisingly, the low end of the 90% confidence interval was somewhat higher than the maximum 2015 VBT mortality rate of 0.500.

Appendix G—A Method for Creating Acceptable Rates

This appendix has been included with the hope of stimulating readers to develop better methods of adjusting rates. It describes a relatively simple approach for replacing rates deemed deficient, according to one of the four tests described in 6.1, Creative Reviews, with adjusted rates that:

1. Closely reproduce the number or amount of total events, for example, total number of deaths or total amount of death claims,
2. are continuous with nearby acceptable rates, and
3. are consistent with rate patterns exhibited by a related array or table.

The approach constructs a series of arrays, all equal in size and shape to the original array of rates, which contain both acceptable and deficient rates. There are seven such arrays:

1. The array of unadjusted rates, which includes "deficient" rates that require adjustment.
2. The adjusted array of rates, in which deficient rates are replaced by rates that move successively closer to achieving the above three requirements with each iteration. This array is calculated as array 4 * array 5.
3. Three arrays for calculating how closely adjusted rates reproduce the number or amount of total events:
   a. An array of exposures
   b. An array of observed events (e.g., deaths or death claims)
   c. An array of differences in events, calculated as array 3b – array 2 * array 3a, i.e., observed events – adjusted rates * exposure, which equals actual events minus adjusted events. To closely reproduce total events, the sum of array 3c should be a small percentage of total events, such as 0.5%.
4. An array of rates that is a combination of the rates deemed acceptable and rates from a related array or related table that will be used as a pattern to create adjusted rates. Different colors should be used to differentiate the acceptable rates from the related rates.
5. An array of ratios, with all ratios initially set to 1.0 for the first iteration. This array will be used to recalculate array 2 in each iteration as array 4 * array 5. The ratios corresponding to rates deemed acceptable should remain equal to 1.0.
   a. In the second iteration, the ratios for related rates should be adjusted so ratios times related rates grade smoothly to nearby acceptable rates.
   b. In the third and later iterations, ratios for related rates can be further adjusted to achieve preservation of events. To that end, it may be advantageous to insert interpolation formulas in each column that interpolate ratios between the first and last ratios applied to related rates.

Appendix H—Converting Mortality Rates between Age Nearest and Age Last Birthdays

Age Zero

This appendix presents formulas based on two different assumptions, neither of which apply to age zero. Because of the relatively high mortality rate in the days after birth, followed by a much lower mortality rate for the rest of the first year of life, a formulaic approach to converting between ANB and ALB is not viable. Instead, it is common practice to borrow ratios of ALB to ANB age zero mortality rates from prior studies. Perhaps, as part of a future study, age zero mortality could be studied by age of death measured in days.

UDD Conversion Method

The UDD conversion method assumes a uniform distribution of deaths (UDD) which results in partial year mortality rates that are proportional to time exposed:

(1) $_tq_x = t\, q_x$.

The ALB mortality rate for age x can be calculated from the half-year mortality rates for the second half of ANB age x and the first half of ANB age $x+1$.

Using UDD, we can calculate $qn_{x+½, ½}$, the mortality rate for the second half of ANB age x, as one minus the probability of surviving from the middle to the end of the year of age x:

(2) $qn_{x+½, ½} = 1 - (1 - qn_x) / (1 - ½\, qn_x)$.

Also using UDD, we can calculate the mortality rate for the first half of age $x+1$ as:

(3) $qn_{x+1, ½} = ½\, qn_{x+1}$.

The mortality rate for ALB age x, ql_x, is calculated as one minus the product of the probabilities of surviving the second half of ANB age x and the first half of ANB age $x+1$:

(4) $ql_x = 1 - (1 - qn_{x+½, ½}) (1 - qn_{x+1, ½})$.

Substituting formulas (2) and (3) into formula (4), we can express the ALB rate for age x in terms of the ANB rates for ages x and $x+1$, as follows:

(5) $ql_x = 1 - ((1 - qn_x) / (1 - ½\, qn_x)) (1 - ½\, qn_{x+1})$, which leads to:

(6) $ql_x = 1 - (1 - qn_x) (2 - qn_{x+1}) / (2 - qn_x)$, which in turn leads to:

(7) $ql_x = (2 - qn_x - 2 + qn_{x+1} + 2\, qn_x - qn_x\, qn_{x+1}) / (2 - qn_x)$,

which finally simplifies to:

(8) $ql_x = (qn_x + (1 - qn_x)\, qn_{x+1}) / (2 - qn_x)$.

Using the UDD assumption once again, we can develop the equivalent formula for calculating ANB mortality rates from ALB mortality rates. From UDD we have:

(9) $ql_{x-½, ½} = 1 - (1 - ql_{x-1}) / (1 - ½ ql_{x-1})$ and

(10) $ql_{x, ½} = ½ ql_x$.

The ANB mortality rate for age x can be calculated as one minus the product of the probabilities of surviving the second half of ALB age $x - 1$ and the first half off ALB age x:

(11) $qn_x = 1 - (1 - ql_{x-½, ½}) (1 - ql_{x, ½})$.

Substituting formulas (9) and (10) into formula (11), we have:

(12) $qn_x = 1 - (1 - ql_{x-1}) / (1 - ½ ql_{x-1})) (1 - ½ ql_x)$,

which simplifies to:

(13) $qn_x = (ql_{x-1} + (1 - ql_{x-1}) ql_x) / (2 - ql_{x-1})$,

which is analogous to the ANB to ALB conversion formula (8).

Geometric Conversion Method

While "Uniform distribution of deaths" is a reasonable assumption for most ages, it increasingly distorts the conversion of rates, both from ANB to ALB and from ALB to ANB, as mortality rates climb above values of 10% in combination with mortality rates increasing at rates lower than the mortality rates.

An alternative assumption is half-year probability of survival rates that decrease at a constant geometric rate within each two-year span This assumption generally performs acceptably, except at ages where there is a significant change in the rate of mortality increase. It tends to outperform UDD as mortality rates climb above values of 10% in combination with mortality rates increasing at rates lower than the mortality rates. A geometrically decreasing probability of survival assumption results in simple and intuitive formulas based on geometric means.

We will assume that half-year probability of survival rates decrease at a constant geometric rate over ANB ages 1 and $x+1$ and use the following notation:

- R will denote the constant ratio between half-year survival rates within a two-year age span.
- S_i will denote the survival rate for the ith half-year, where i runs from 1 to 4.
- pn_x will denote the probability of surviving ANB age x.
- pl_x will denote the probability of surviving ALB age x.

From these definitions, we have:

(1) $S_i = S_1 R^{i-1}$.

This results in: $S_2 = S_1 R^1$, $S_3 = S_1 R^2$, and $S_4 = S_1 R^3$.

Annual probabilities of survival can be calculated as products of half-year survival rates:

(2) $\quad pn_x = S_1 \, S_2$,

(3) $\quad pn_{x+1} = S_3 \, S_4$, and

(4) $\quad pl_x = S_2 \, S_3$.

Using formulas (1), (2) and (3) to substitute for pn_x and pn_{x+1}, we have:

(5) $\quad \sqrt{(pn_x \, pn_{x+1})} = \sqrt{(S_1 \, S_2 \, S_3 \, S_4)} = \sqrt{((S_1)^4 (R)^6)} = S_2 \, S_3 \ = pl_x.$

We have shown above that the probability of surviving through ALB age x, pl_x, is equal to the geometric mean of the probabilities of surviving through ANB ages x and x+1, pn_x and pn_{x+1}, that is:

(6) $\quad pl_x = \sqrt{(pn_x \, pn_{x+1})}.$

Translating formula (6) to use mortality rates instead of survival rates, we have a simple and intuitive formula for calculating ALB mortality rates in terms of ANB mortality rates:

(7) $\quad ql_x = 1 - \sqrt{((1 - qn_x)(1 - qn_{x+1}))}$.

Similarly, assuming that half-year probability of survival rates decrease at a constant geometric rate over ALB ages x-1 and x, we can prove that the probability of surviving through ANB age x, pn_x, is equal to the geometric mean of the probabilities of surviving through ALB ages x-1 and x, pl_{x-1} and pl_x:

(8) $\quad pn_x = \sqrt{(pl_{x-1} \, pl_x)}.$

Translating formula (8) to use mortality rates instead of survival rates, we have a simple and intuitive formula for calculating ANB mortality rates in terms of ALB mortality rates:

(9) $\quad qn_x = 1 - \sqrt{((1 - ql_{x-1}) \, (1 - ql_x))}$.

Appendix J—Valuation Loading

In the U.S., the NAIC promulgates mortality and other tables that are used for statutory valuation and nonforfeiture values. In 2016, new legislation cleared the way for principle-based reserves (PBR), which allows insurers with credible experience to add their own experience to the calculation of statutory reserves for certain life insurance products. Valuation loadings applied to PBR assumptions based on company experience are beyond the scope of this appendix.

Before the NAIC adopts a new valuation table, various steps are performed to ensure the new table will produce adequate reserves for a super-majority of the insurers that will use the table. This appendix describes those steps in general terms.

Industry experience studies solicit contributions of data from a cross-section of insurers. Often, contributors range from some of the largest to some of the smallest writers of the business covered by the study.

The valuation loading process involves the four tables shown below, which will be discussed in the following sections:

Valuation Loading Tables		
	No Loading	**With Loading**
No Projection	Experience Table	Loaded Experience Table
With Projection	Valuation Basic Table	Commissioners' Valuation Table

Valuation Basic Table

The product of an experience study will be referred to as the **experience table**, which is based solely on the experience study's data with no projection of rates to a future date and no valuation loading.

The experience table is assumed to have an effective date equal to the mid-point of the study data. The **Valuation Basic Table (VBT)** will typically have an effective date that is beyond the mid-point of the study data. The VBT is developed by applying improvement or deterioration factors to the experience table to update rates from the experience table's effective date to the VBT's effective date.

Coverage Ratios

Each contributor's experience is compared to the experience table to determine a **coverage ratio** for each contributor. A contributor's coverage ratio is calculated as A divided by B, where:

A. is the contributor's actual claims contributed to the study and
B. is the contributor's expected claims, based on the experience table.

The contributor coverage ratios are then ordered from low to high to help select the **regulators' coverage ratio**. A typical goal is a regulators' coverage ratio that will equal or exceed the coverage ratios of, say, 70% to 80% of the contributors.

It is not unusual for coverage ratios for a few small contributors to be very much out of line with the coverage ratios for the rest of the contributors; this is often attributed to random fluctuations within a small sample. Sometimes, such contributors are excluded from the coverage ratio analysis.

The coverage ratios can also be analyzed in combination with each contributor's exposure and claims. Regulators may consider selecting a regulators' coverage ratio that equals or exceeds the coverage ratios of contributors that account for, say, 70% to 80% of total exposure or total claims.

Loaded Experience Table

The regulators' coverage ratio is used to structure **loading factors** (or loading percentages, equal to loading factors minus one). In the simplest case, a single loading factor might be set equal to the regulators' coverage ratio. In other cases, the final loading factors may vary by attained age and other factors. For example, for a mortality table, the loading factors may grade towards 1.0 as age approaches 100.

Once loading factors have been selected, they are tested. Testing is done by applying the loading factors to the experience table to create a **loaded experience table**. Expected claims are then calculated based on the loaded experience table and compared to each company's actual claims. If the regulators' coverage goal is not met, then the loading factors must be adjusted up or down and the process repeated until the coverage goal is met.

Commissioners' Valuation Table

Once loading factors and the loaded experience table have been finalized, the **Commissioners' Valuation Table (CVT)**, can be created. This is the table that will be used to calculate statutory reserves in the future. The CVT has its own effective date, which may be beyond the VBT's effective date. The CVT is created by applying improvement or deterioration factors, which may be the same as those used to create the VBT, to the loaded experience table to update claims experience from the effective date of the experience table to the effective date of the CVT.

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 24,000 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 non-actuaries 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.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org