SOCIETY OF ACTUARIES

# Long Term Care Experience Basic Table Development

April 2015

SOCIETY OF ACTUARIES

# Long Term Care Experience Basic Table Development


**AUTHORS**

Vincent Bodnar, ASA, MAAA
Chief Actuary
LTCG

Ben Williams
Senior Consultant
Towers Watson

Matt Morton, FSA, MAAA
Senior Consulting Actuary
LTCG

Tim Wood, FSA, MAAA
Senior Consultant
Towers Watson

# TABLE OF CONTENTS

## Preface: Information added to the report subsequent to the 7/23/2015 release

**July 2015 Update:**

- **Commentary added to Appendix A:**

  The modelled claim termination rates were developed as a rate of termination per 0.5 months of exposure. The data manipulation required for GLM ensured that individual claim exposure periods did not transcend more than one policy duration, claim duration  or calendar year. This allowed for the true effect of these items to be measured within the GLM model development.

  As a result, the GLM model output shows the rate of claim termination for each 0.5 months of claim exposure. The Appendix A Microsoft Excel models illustrate the development of the model itself with two final claim termination rate columns on the summary tabs. The first column shows the output of the claim termination rate per 0.5 months of exposure and the second column shows the claim termination rate per full month of claim exposure.

# Long Term Care Experience Basic Table Development

## SECTION 1: ACKNOWLEDGEMENTS

The authors would like to extend our thanks to all participating companies for making this project a success. Without your support, such research projects would not be possible. Your contributions have led to the development of these basic tables for long term care morbidity assumptions. A list of the participating companies is in Section 5 of this report. We would like to thank the steering committee for their support, guidance, direction and feedback throughout the project. The members of the committee are:

- Sheryl Babcock
- Barry Koklefsky, Vice-Chair
- Susan Oberman Smith
- Eric Perry
- Eric Poirier
- Jon Prince, Chair
- Steve Schoonveld
- Maureen Shaughnessy
- Bruce Stahl
- Kevin Waterman
- Perry Wiseblatt
- Bob Yee

We would like to thank Cynthia MacDonald, Erika Schulty and Muz Waheed from the SOA for their leadership and coordination of the project.

We would like to thank the staff members of LIMRA and MIB for their contributions to the project.

This report was prepared by Ben Williams and Tim Wood of Towers Watson and Vincent Bodnar and Matt Morton of the Long Term Care Group.

## SECTION 2: BACKGROUND AND SCOPE

In a letter of engagement dated January 14, 2014, Towers Watson Delaware, Inc. (“Towers Watson”) was retained by the Society of Actuaries (“SOA” or “you”) to assist in a multi-phase project with an objective of creating tables that represent the experience for long term care (“LTC”) business, utilizing data gathered by the SOA from LTC carriers in the industry.

Phase 1 of this project was to assess the quality of data submitted to the SOA by certain LTC carriers (the “Participants”) to determine its suitability to support the creation of tables that represent the experience. The recommendation at the conclusion of Phase 1 was to move forward with experience studies for claim incidence, claim termination and benefit utilization.

In Phase 2, we developed aggregate databases of experience results for each of the studies, which were made available on the SOA website along with a report describing the approach to developing the aggregate databases and definitions of each of the data elements provided therein.

Phase 3 of the project was to produce basic rate tables for claim incidence, claim termination and benefit utilization from the data collected and the analysis performed in Phases 1 and 2.

Rather than utilizing the traditional methodologies to develop the tables, predictive modelling techniques were used to develop a multiplicative model to represent the basic table for each of the assumptions. Section 3 of this report sets forth the basic tables for each of the assumptions, provides instructions on how to use the basic tables and limitations and caveats to the basic tables. Section 4 of this report describes the predictive modelling process and the assumptions made in building the predictive models for each of these long term care assumptions.

This report is intended to provide certain actuarial information and analyses that may assist a qualified professional in interpreting experience data and developing model assumptions for long term care.

It’s important to note that these models were built as a reflection of the data provided by the participants. Companies included or not included in the study may have experienced different morbidity experience. In addition, usage of these models may/may not be applicable to similar products like LTC riders on life insurance policies or potential product design changes in the future.

This report is considered a statement of actuarial opinion under the guidelines promulgated by the American Academy of Actuaries. Vincent Bodnar and Matthew Morton, the consulting actuaries contracted by Towers Watson who developed this report, are members of the American Academy of Actuaries and meet the Qualification Standards of the American Academy of Actuaries to render the opinions contained herein.

# SECTION 3: BASIC TABLES

To understand the drivers of claim incidence, claim termination and benefit utilization, we developed predictive models, which were fit from the data collected from the participants. Rather than expressing the basic tables in tabular format, they are presented as multiplicative models where the value (e.g., the number of claims expected per year of exposure for a specific risk profile) is the product of a base rate and factors for each of the product and policyholder characteristics that were determined to be significant drivers.

In total, eight models were fit on the actual experience. These were:

- Claim incidence
    - Active life incidence rate
    - Total life incidence rate
- Claim termination
    - Total claim termination rate (recovery and death combined), including claim type and diagnosis
    - Total claim termination rate (recovery and death combined), excluding claim type and diagnosis
    - Claim termination rate due to death, including claim type and diagnosis
    - Claim termination rate due to death, excluding claim type and diagnosis
- Benefit utilization
    - Benefit utilization, including claim type and diagnosis
    - Benefit utilization, excluding claim type and diagnosis

Each of these models and analyses are described in the following sections. Microsoft Excel-based models are also included showing the detailed values resulting from the analysis.

## 3.1 CLAIM INCIDENCE

Claim incidence is a rate of claims per annual exposure. The list of claim incidence models provided in Microsoft Excel can be found in Appendix A.

The incidence models were developed on the aggregated data. In each model, the basic table, or components of the table, may or may not be applicable to specific individual companies. Although the table generally fits well in aggregate for each of the participating companies, individual relationships may vary by company. Additional information and examples can be found in Section 3.1.5.

It is important to note that the user should include all factors utilized in the model. Removing factors or relationships may yield inaccurate results. It is not recommended that any factors be removed.

### *3.1.1 DATA*

The table below summarizes the data used to develop the claim incidence model.

| | Exposure | | |
|---|---|---|---|
| **Gender** | **Total Lives** | **Active Lives** | **Claim Count** |
| Female | 8,845,238 | 8,630,328 | 116,779 |
| Male | 5,991,446 | 5,905,355 | 55,502 |
| **Total** | **14,836,684** | **14,535,684** | **172,281** |

The data represents the contributions of the participants over the period January 1, 2000 to December 31, 2011. This data is consistent with the incidence aggregate database produced by the Society of Actuaries in 2015.

- Total life exposure is calculated as the number of days during the exposure period (January 1, 2000 to December 31, 2011) between the policy effective date and termination date. There is no adjustment for the period of time on claim.

- Active life exposure is calculated as the total life exposure reduced for the period of time on claim.

- All exposure was calculated based on the exact number of days exposed. No adjustments were made to exposures for individuals who started a claim in a given period.

- The claim count for each policy was provided within the policy table received from each participant. Any claims that started prior to the beginning of the exposure period (1/1/2000) were removed from the claim count as the incidence rate is intended to measure the probability of a new claim occurring.

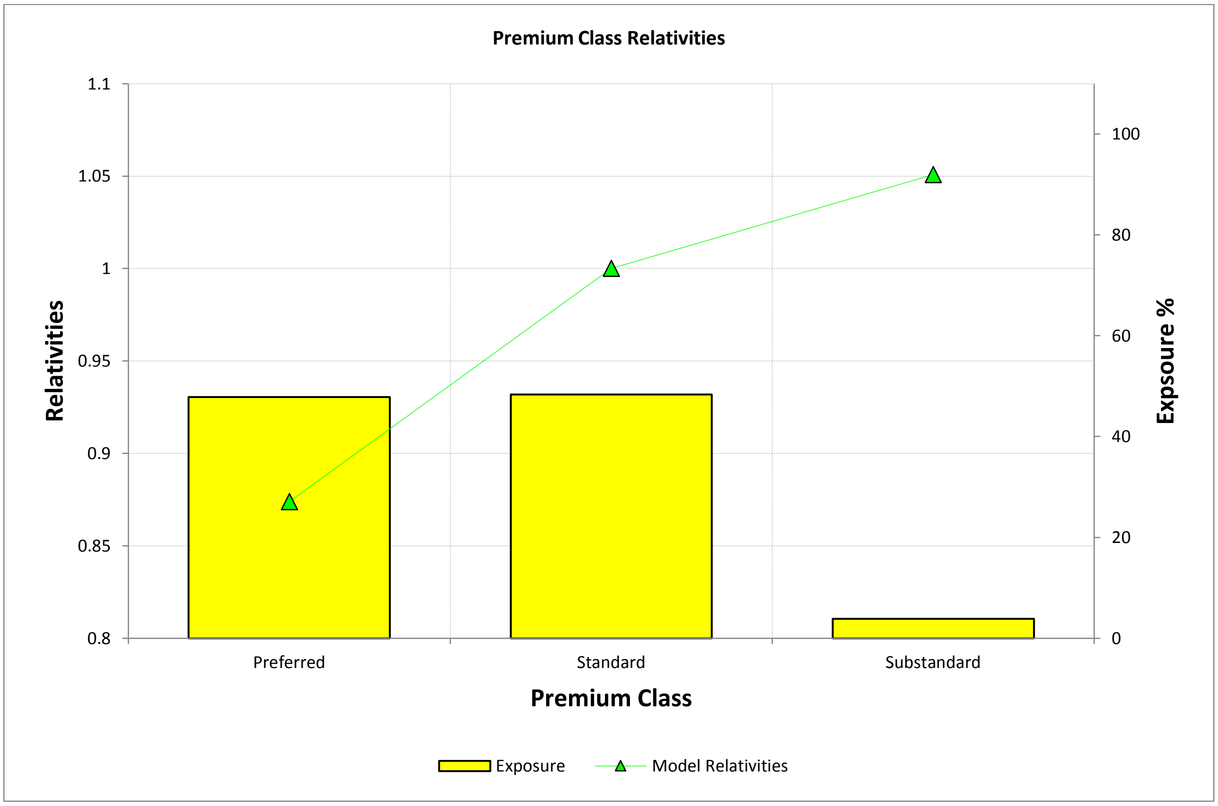
- For policies identified with more than one claim, aggregation of claims occurred when service dates overlapped or were within six months of each other. This six-month test was applied to each of the studies to ensure a common definition of a unique claim.

### *3.1.2 PREDICTORS*

To use the models, the following product and policyholder data is necessary as these are the components used in the claim incidence models. It is important to understand that the basic table produced in each model may not be appropriate for a risk profile that does not have the data necessary for certain variables, such as gender, marital status and premium class. Other variables, such as region and benefit period, have levels indicating when this data is unknown.

| **Product Characteristic** | | **Policyholder Characteristic** | |
|---|---|---|---|
| **Underwriting Type** | *Full Underwriting or Other* | **Issue Age** | |
| **Elimination Period** | *in Days* | **Premium Class** | *Preferred, Standard or Substandard* |
| **Benefit Period** | *Limited, Lifetime or Unknown* | **Marital Status** | *Married or Single* |
| **Coverage Type** | *NH, HHC, ALF, Comprehensive, Other* | **Region** | *Northeast, Midwest, South, West or unknown* |
| **Max Daily Benefit** | *in dollars* | **Gender** | *Male or Female* |
| **TQ Status** | *Tax Qualified, Non-Tax Qualified, Other* | | |

As an illustrative example of the results of this analysis, the graph below shows the exposure and relativities associated with each level of premium class. The relativities isolate the true effect of each level of premium class only by standardizing for the effects of other factors. For example, this means that, if the effects of all other variables are removed, substandard premium class will have an incidence rate approximately 5% higher than standard premium class.




The analysis indicates that most of the variables have effects that will interact with other variables – mostly attained age or policy year - such that the factor will vary based on these criteria. In the case of region, the effect varies by maximum daily benefit.

Benefit period is the only variable in the claim incidence model whose effect does not interact with another variable.

Other variables in the data were reviewed. These included, but were not limited to, policy distribution method, historical rate actions on policy and claim payment style. These data items were reviewed for completeness and impact on incidence. In some cases, the data was not considered reliable and, in other cases, the impact on incidence was not statistically significant.

### 3.1.3 SAMPLE CALCULATION

For the risk profile in the table below, we will show how the final incidence rate is determined from the tables in the accompanying Microsoft Excel files listed in Appendix A for several policy years.

| Product Characteristic | | Policyholder Characteristic | |
|---|---|---|---|
| **Underwriting Type** | *Full Underwriting* | **Issue Age** | *65* |
| **Elimination Period** | *0 days* | **Premium Class** | *Standard* |
| **Benefit Period** | *Limited* | **Marital Status** | *Single* |
| **Coverage Type** | *Comprehensive* | **Region** | *Northeast* |
| **Max Daily Benefit** | *$100* | **Gender** | *Female* |
| **TQ Status** | *Tax Qualified* | | |

For each attained age/policy year for this policy issued at age 65, the calculated claim incidence rate is the product of all the factors shown below corresponding to the base rate and the values for each component.

| **Attained Age** | **65** | **66** | **67** | **69** | **71** | **74** |
|---|---|---|---|---|---|---|
| **Policy Year** | **1** | **2** | **3** | **5** | **7** | **10** |
| Base Incidence Rate | 0.00039 | 0.00039 | 0.00039 | 0.00039 | 0.00039 | 0.00039 |
| Elimination Period | 2.7523 | 2.6835 | 2.6165 | 2.4874 | 2.3646 | 2.1918 |
| Benefit Period | 1.0580 | 1.0580 | 1.0580 | 1.0580 | 1.0580 | 1.0580 |
| Tax Qualified Status | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Gender | 1.0000 | 1.1682 | 1.3781 | 1.9663 | 2.8713 | 5.1286 |
| Coverage Type | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Region | 0.9381 | 0.9381 | 0.9381 | 0.9381 | 0.9381 | 0.9381 |
| Premium Class | 1.0000 | 1.4301 | 1.8856 | 2.6944 | 3.1776 | 3.2926 |
| UW Type | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Marital | 2.1133 | 2.0587 | 2.0046 | 1.8982 | 1.7943 | 1.6438 |
| **Final Incidence Rate** | 0.0023 | 0.0036 | 0.0053 | 0.0097 | 0.0150 | 0.0236 |

Note that these rates represent exposure on a total lives basis. Similar results on an active lives basis were also produced.

### 3.1.4 ASSUMPTIONS IN MODEL

Exposure was calculated based on the number of days a policy is exposed. As a result, partial annual exposures exist. This contrasts to a method where every exposed policy receives a full exposure.

Certain assumptions were required to create the model, including, but not limited to:

- Factors are related multiplicatively and claims occur as a rate of exposure.
- Flattening of attained age factors at age 97. Minimal data existed in the dataset above age 97.
- Certain characteristics contained unknown values. In some cases, these were treated as a separate level.
- Definition of incurred date:
    - Minimum service date of claim less the elimination period.
- Definition of unique claim
    - Payment must have been made.
    - Services within six months are combined into a single claim.

### 3.1.5 SAMPLE MODEL DISCUSSION

Certain results of the model were discussed in depth among the steering committee and other team members. An example of this is the impact of benefit period on incidence rates.

The final model has a slight difference between limited and lifetime benefit periods. The expectation prior to the study was that lifetime benefit periods would have materially higher incidence rates than limited benefit periods. Though this relationship was observed in some of the participating companies, not every participant's data exhibited this relationship. When all of the data was aggregated, the relationship of higher incidence for lifetime benefit period was no longer present.

This was an example of the strengths of the model, but illustrates its limitations as well. The models and relationships therein were developed based on the aggregate data without regard for the relationships inherent in specific companies' data. For this reason, the same relationships may not exist for each company in the LTC industry.

There was significant discussion around this topic that provided potential explanations for the differences of this relationship from expectation. A (non-exhaustive) sample of potential drivers of the benefit period relationship in this model include:

- The mix of companies participating in the study; and
- Differences in underwriting protocols by company.

This relationship is an example of a potential difference between an individual company's result and the aggregate model. Other differences could be length of underwriting differences by policy duration. Some companies may have differences for 5 or 10 or 25 years, for example, based on their underwriting.

The model was created based on the aggregate data and reviewed in aggregate by company, which yielded reasonable results. The relationships inherent in the model, as discussed in this report, may not be applicable for every company in the LTC insurance industry.

## 3.2 CLAIM TERMINATION

The probability that an existing claim will close in a given month is referred to as the claim termination rate. Claim termination typically occurs due to either recovery or death. We fit separate models of termination rate due to any reason, and owing to death, and separate models including all information available, and including all information except for information on claim type or diagnosis. The models excluding information on claim type and diagnosis can be used to project termination rates for future claims, or claims for which this information is unavailable.

The models were developed on the aggregated data. For each model, the basic table, or components of the table, may or may not be applicable to specific companies. Although the table generally fits well in aggregate for each of the participating companies, individual relationships of characteristics may be different by company.

It is important to note that the user should include all factors utilized in the model. Because the removal of factors or relationships may yield inaccurate results, it is not recommended that any factors be removed.

The list of claim termination models provided in Microsoft Excel can be found in Appendix A.

### *3.2.1 DATA*

The claim termination models were fit on approximately five million observations. Each observation recorded the characteristics (including, but not restricted to, the policy characteristics listed in the previous section) of a number of claims in a given month. In fitting the model, the weight was the number of claims open in the month, and the response was the proportion of those claims that closed.

Claim exposures were calculated as the number of days the claim was exposed to the possibility of termination. As a result, partial monthly exposures exist. This contrasts with a methodology whereby, if a claim is exposed in a month, it would receive a full exposure. Claims that were listed as benefit exhaustions were not considered terminations (i.e., only deaths and recoveries were considered). However, full exposure was given to these claims during the experience period.

| | **Exposure** | |
|---|---|---|
| **Gender** | **Disabled Months** | **Claim Terminations*** |
| Female | 3,119,379 | 86,660 |
| Male | 1,347,469 | 57,782 |
| **Total** | **4,466,848** | **139,442** |

*Note that the count of claim terminations does not include claims that were either open at the end of the exposure period or exhaustions.

It's important to note that policies receive zero exposure for time during the elimination period. As an example, a claimant with a 90-day elimination period would receive zero exposure for the first three claim duration months. The first month of claim exposure would occur in the fourth claim duration, after satisfaction of the elimination period. It is worth noting that, by using this method, a policy that is reported before the satisfaction of the elimination period and ends before the elimination period is complete is not considered in this analysis.

### 3.2.2 PREDICTORS

Similarly to the incidence model described previously, in order to apply the termination models to a given claim, a number of characteristics of the product, policyholder and claim are necessary. These are:

| Product Characteristic | | Policyholder Characteristic | |
|---|---|---|---|
| **Benefit Payment Style** | *Cash, Indemnity, Reimbursement, Other* | **Incurred Age** | |
| **Benefit Period** | *Limited, Lifetime or Unknown* | **Marital Status** | *Married or Single* |
| **Coverage Type** | *NH, HHC, ALF, Comprehensive, Other* | **Gender** | *Male or Female* |
| **Daily Benefit** | *in dollars* | | |
| **TQ Status** | *Tax Qualified, Non-Tax Qualified, Other* | | |

| Claim Characteristic | |
|---|---|
| **Claim Type** | *NH, HHC, ALF, Other* |
| **Diagnosis** | *See Appendix C for list* |
| **Claim Duration** | *in months* |
| **Policy Duration** | *in years, when claim incurred* |

Note that claim type and diagnosis are not required in order to apply the model, which excludes information on claims.

### 3.2.3 SAMPLE CALCULATION

The termination models are applied similarly to the incidence models, as described in 3.1.3.

### 3.2.4 ASSUMPTIONS IN MODEL

Certain assumptions were required to create the model, including, but not limited to:

- Factors are related multiplicatively.
- All claim termination models exclude claim exhaustion as a reason for termination. The exposures for these claims were included until the date of exhaustion.
- The models are reliant on the coding of claim termination. No specific validations were completed to verify or audit the coding of death, recovery and exhaustion.
- Certain characteristics contained unknown values. In some cases, these were treated as a separate level.

## 3.3 BENEFIT UTILIZATION

Some long term care policies reimburse claimants on actual incurred benefit amounts subject to a maximum daily benefit and some policies have indemnity policy form language. For the policies that have reimbursement provisions, the amount of the reimbursement relative to the maximum daily benefit is referred to as benefit utilization. This is a key morbidity assumption for modeling long term care policies with reimbursement provisions. For policies with indemnity policy form language, this assumption is typically not utilized in the same manner.

We fit separate models, including all information available, and including all information except for information on claim type or diagnosis. The models excluding information on claim type and diagnosis can be used to project termination for future claims, or claims for which this information is unavailable. For each model, the basic table, or components of the table, may or may not be applicable to specific companies. Although the table generally fits well in aggregate for each of the participating companies, individual relationships of characteristics may be different by company.

It is important to note that the user should include all factors utilized in the model. Because the removal of factors or relationships may yield inaccurate results, it is not recommended that any factors be removed.

The list of benefit utilization models provided in Microsoft Excel can be found in Appendix A.

### *3.3.1 DATA*

The benefit utilization models were fit on approximately five million observations. Each observation recorded the characteristics of a number of claims in a given month. In fitting the model, the weight was the total monthly benefit dollars available, and the response was the ratio of benefit dollars paid to benefit dollars available.

| Benefit Utilization | | |
|---|---|---|
| **Gender** | **Claim $** | **Disabled Months** |
| Female | $6,099,111,677 | 2,390,828 |
| Male | $2,481,763,731 | 939,216 |
| **Total** | **$8,580,875,408** | **3,330,044** |

### 3.3.2 PREDICTORS

Similarly to the incidence models described previously, in order to apply these models to a given claim, a number of characteristics of the product, policyholder and claim are necessary. These are:

| Product Characteristic | | Policyholder Characteristic | |
|---|---|---|---|
| Benefit Payment Style | Cash, Indemnity, Reimbursement, Other | Incurred Age | |
| Benefit Period | Limited, Lifetime or Unknown | Marital Status | Married or Single |
| Coverage Type | NH, HHC, ALF, Comprehensive, Other | Gender | Male or Female |
| Daily Benefit | in dollars | | |
| Inflation Rider | Without inflation protection, Other | | |

| Claim Characteristic | |
|---|---|
| Claim Type | *NH, HHC, ALF, Other* |
| Diagnosis | *See Appendix C for list* |
| Claim Duration | *in months* |
| Policy Duration | *in years, when claim incurred* |
| Calendar Year | |

Note that claim type and diagnosis are not required in order to apply the model that excludes information on claims.

### 3.3.4 SAMPLE CALCULATION

The benefit utilization models are applied similarly to the incidence models, as described in 3.1.3.

### 3.3.5 ASSUMPTIONS IN MODEL

Certain assumptions were required to create the model, including, but not limited to:

- Factors are related multiplicatively.
- Certain characteristics contained unknown values. In some cases, these were treated as a separate level.
- Benefit utilization rates are capped at 100% in the user Excel model. Some actual experience cells showed values over 100%. These were not capped in the analysis of the data as there may be contractual benefits that are paid without being subject to the daily or monthly maximum.

### 3.4 USAGE

Each of these models were developed specifically for the data collected in this study. Because the data captured is diverse in characteristics and experience and sourced from many different companies, all attributes could not be fully considered. Some examples of attributes that could not be captured that may / may not influence morbidity rates include:

- Claim management process
- Data collection process
- Underwriting questions or philosophy
- Provider / network discounts
- Specific policy form language
- Sales / distribution channel

This is an illustrative, but not comprehensive, list of potential drivers of morbidity experience that are not captured in the data. Any application of these models should keep this fact in mind.

In addition, these models were based on long term care policy experience. As a result, using these models for other insurance products (either currently in force or future product designs) may not be appropriate.

Finally, these models were fit on historical data. High ages and durations were grouped because of comparatively small volumes of data. Application of these models to high ages or durations, or to future years, may require extrapolation of these model results or adjustments to the models' base levels.

## SECTION 4: PREDICTIVE MODELING APPROACH

We illustrate the development of a predictive model using the example of the claims incidence model for total exposure. The procedure was similar for the other models.

Once the initial review of the data was complete, we loaded data from the combined incidence tables into Towers Watson's proprietary Generalized Linear Modelling (GLM) software, Emblem. For the incidence model, the response variable for the predictive modelling exercise was number of claims / total years exposed and the weight was total years exposed. Emblem was used for all of the following phases: univariate analysis, multivariate analysis and model validation.

### 4.1 UNIVARIATE ANALYSIS

Univariate analysis was performed prior to continuing to the GLM or multivariate phase. During this step, we reviewed raw incidence rates and exposures summarized by the different variables (duration, marital status, etc.). We looked for any areas where the exposure distribution or the incidence rates were not consistent with our expectations.

As an example, the graph below shows the incidence data summarized on a univariate basis by Premium Class. Incidence in each level is shown relative to incidence in the most representative level, Standard. We note that the pattern does not conform to our expectations, as observed incidence is lower for Substandard than for Standard. Investigation of the data confirmed that this is correct. In the following section, multivariate analysis shows us that this pattern can be explained by a different mix of business between Standard and Substandard.


### 4.2 MULTIVARIATE ANALYSIS

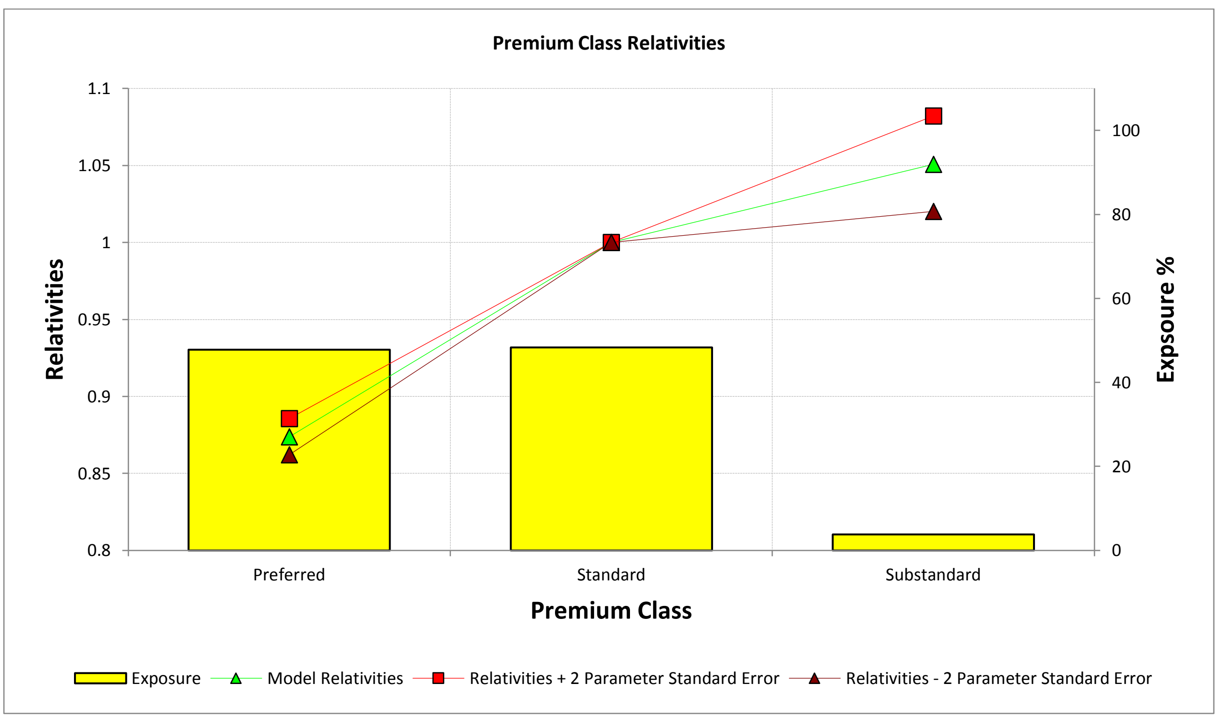
The model fitted was a Generalized Linear Model, or GLM. The technical details of GLMs are discussed in greater detail in Appendix B. The model was fit using a Log link function and Poisson error structure. This combination is widely used in the life insurance industry to model mortality and in the property and casualty insurance industry to model claim frequency. One of the advantages of using a Log link function is the model is multiplicative (the different parameters in the model are multiplied together). The model was fit over a subset of the overall data set (a randomly-determined two-thirds) and the remaining data was reserved for model validation.

The output of the model was a base claim incidence rate and a series of relativities that vary across the possible levels or values for each variable (or predictor) included in the model. Each variable has one level defined as the base, which has a relativity of 1.00. The base level is either the level with the most exposure, or some other level of importance. The impact on incidence of other levels of the variable are measured relative to this base level. The model also includes interactions, which are effects that require the combination of different variables (e.g., the effect of attained age may vary by marital status). Factors or interactions can be simplified during the modelling process by grouping levels or through the fitting of a curve. Simplification is carried out in order to reduce the number of parameters in the model (principle of parsimony) and to smooth effects. Curve-fitting is a predictive modelling technique used for ordinal variables, especially those with a large number of levels. An x-value is associated with each level of the variable, and some function of the x-value is included in the model. The most widely used example is a polynomial; rather than include one parameter in the model for each level of the variable, we include one or more coefficients of the powers of x.

Broadly speaking, in this phase we determined which factors have a significant impact on incidence, and how these factors relate to each other. In particular, the effect of each factor on incidence is determined, controlling for the impact of the other variables included in the model. Included in this phase were decisions around which variables should be included, excluded, simplified or interacted with other variables. These decisions were guided by a number of statistical measures including Chi-Squared percentages, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and analysis of standard errors of parameters, as well as Balance Tests and Consistency Tests, all of which were performed in Emblem.

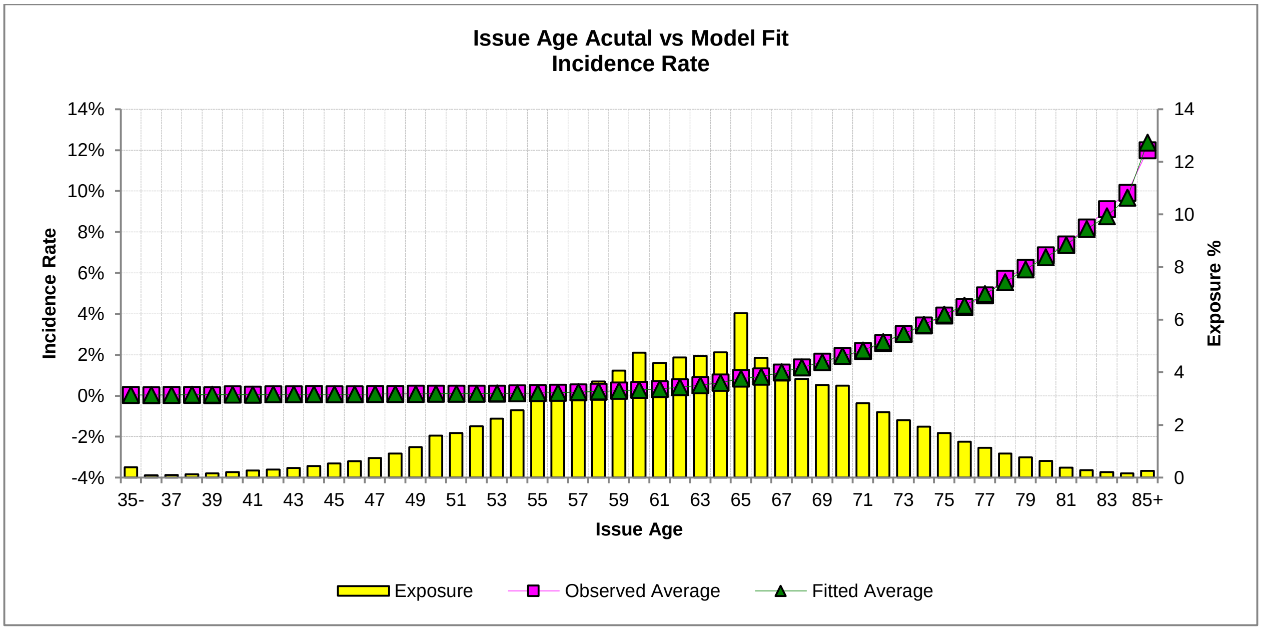
AIC, BIC and Chi-Squared percentages were used to make judgments on which variables to include in the model. AIC deals with the trade-off between the goodness of fit of the model and its complexity. BIC is similar. AIC and BIC tests were performed when adding variables or making other changes to the model to determine whether the improvement in the goodness of fit was sufficient to justify the additional complexity of the model. Similarly, Chi-Squared tests were used to judge the difference between nested models.

The graph below illustrates the standard error for the variable premium class. All premium classes have narrow standard errors. A large standard error for a level would indicate that limited confidence can put in the relativity for that value of the variable, and might suggest that the level should be grouped with the base or another appropriate level.

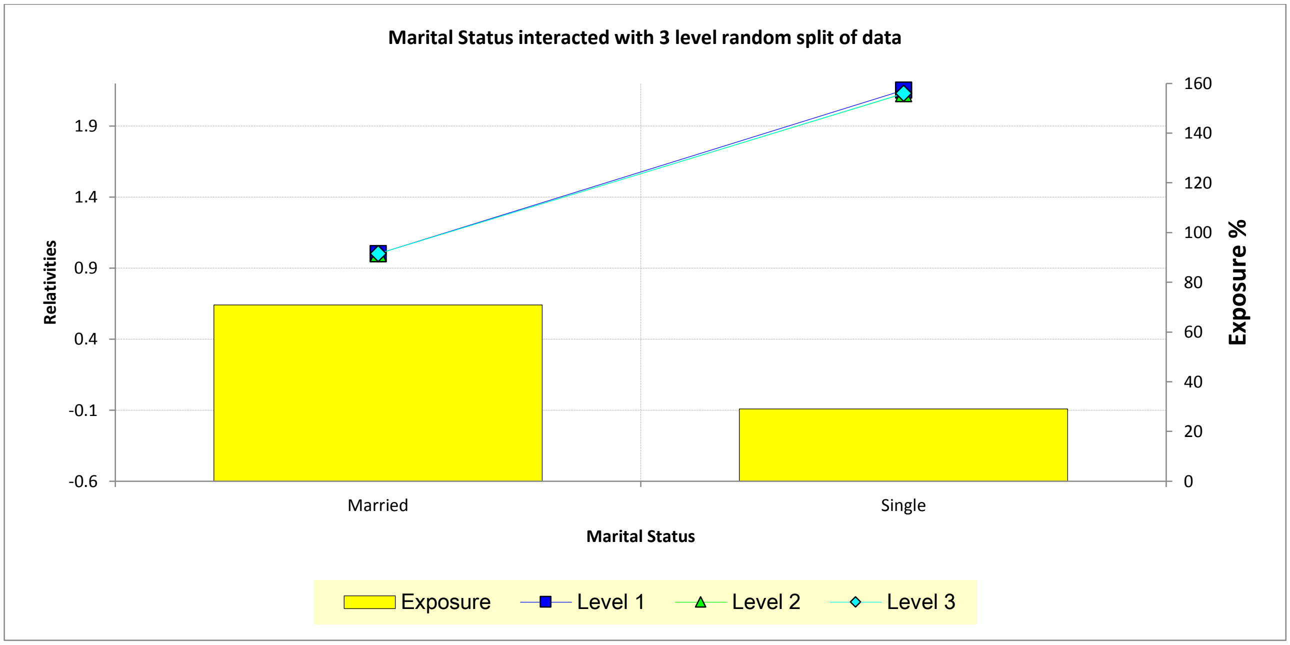


This graph can be compared to the graph of observed incidence summarized by Premium Class in the previous section. Once the effects of other factors are taken into account, Substandard is shown to have higher incidence than Standard, all else being equal. This means that the lower observed incidence for Substandard is owing to a different mix of business from Standard risks.

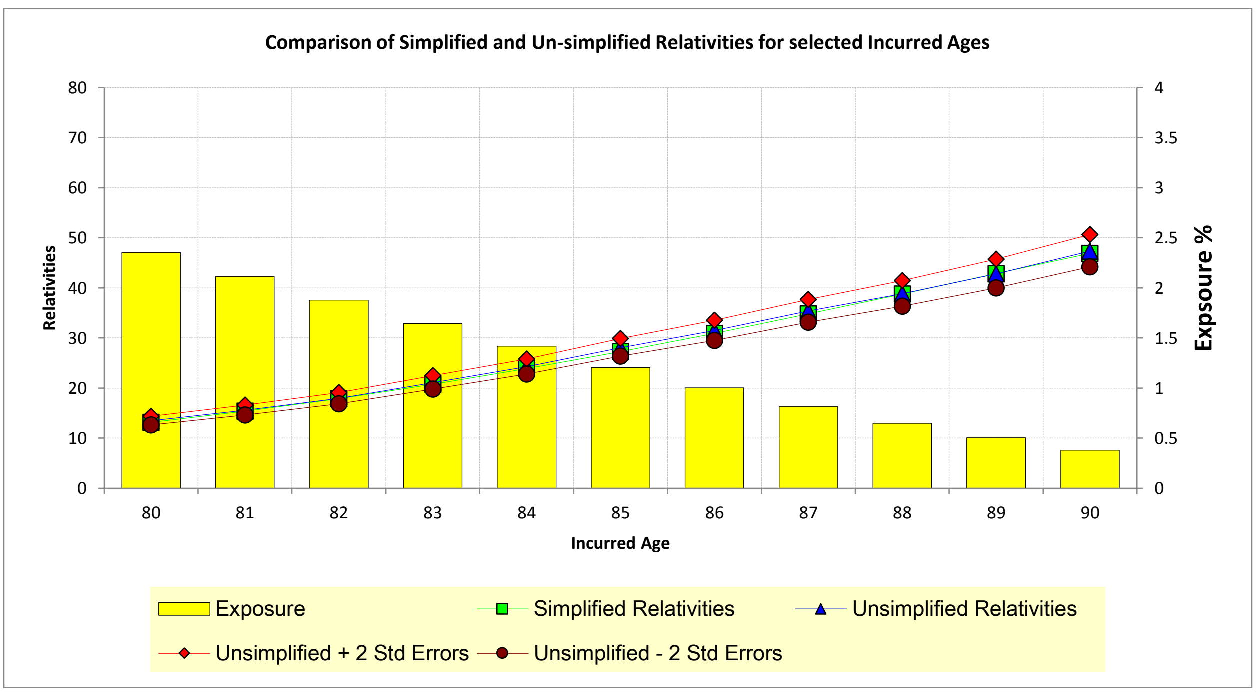
Balance tests were used and reviewed throughout the model-fitting process. Balance tests compare observed values against predicted values according to the model (fitted values) on a univariate basis. Important differences between observed and predicted incidence suggest the factor should be considered for inclusion in the model, or that the factor simplification is not appropriate. An example of a balance test is shown below for issue age, a variable that was not explicitly included in the incidence model.



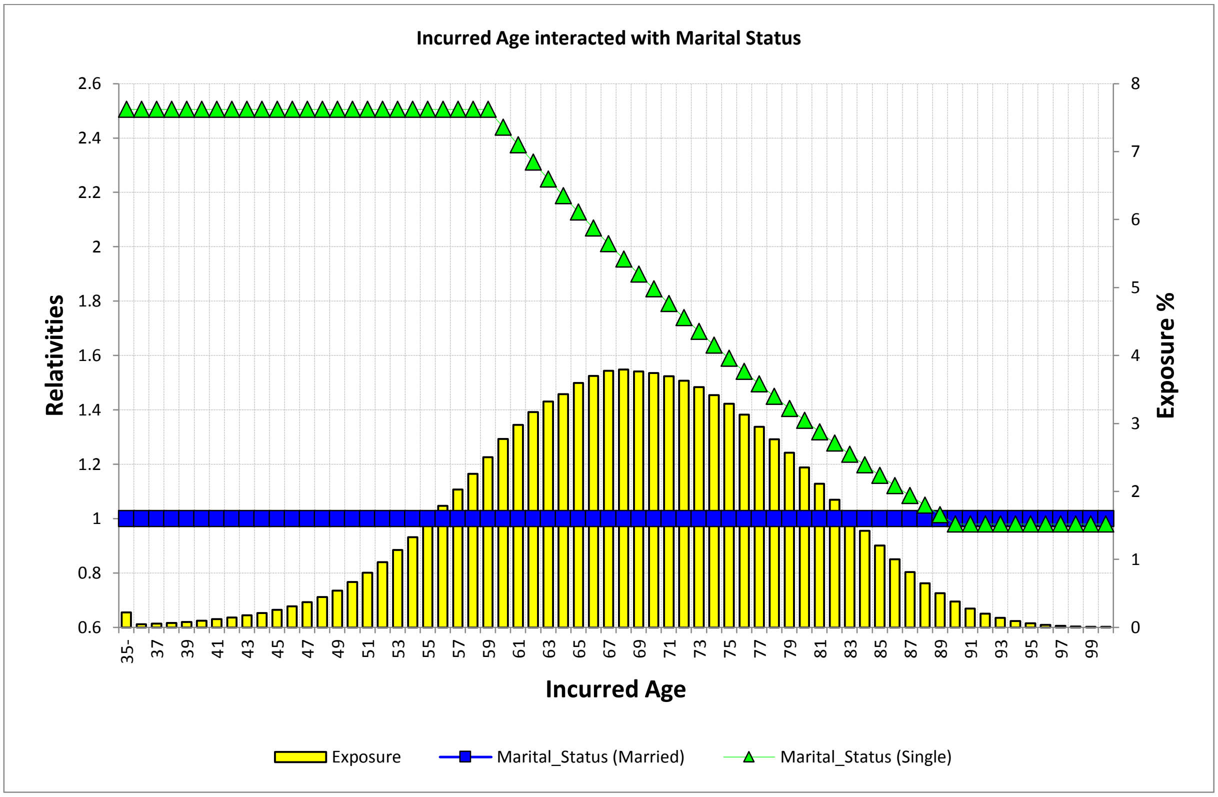
Tests were also performed to ensure that patterns were relatively stable across a random split in the data. The graph below shows how the effect of marital status was consistent across a three-way random split in the data.




An example of a simplification appears below. This shows how the original un-simplified Incurred Age relativities are replaced by a curve that closely follows the original pattern, and does not stray outside the envelope defined by the standard errors.



An example of a simplified interaction appears below. This shows how the effect of incurred age differs for different levels of marital status. At lower ages, singles have an incidence rate more than twice that of marrieds, but this ratio decreases with age until the rates are approximately equal.



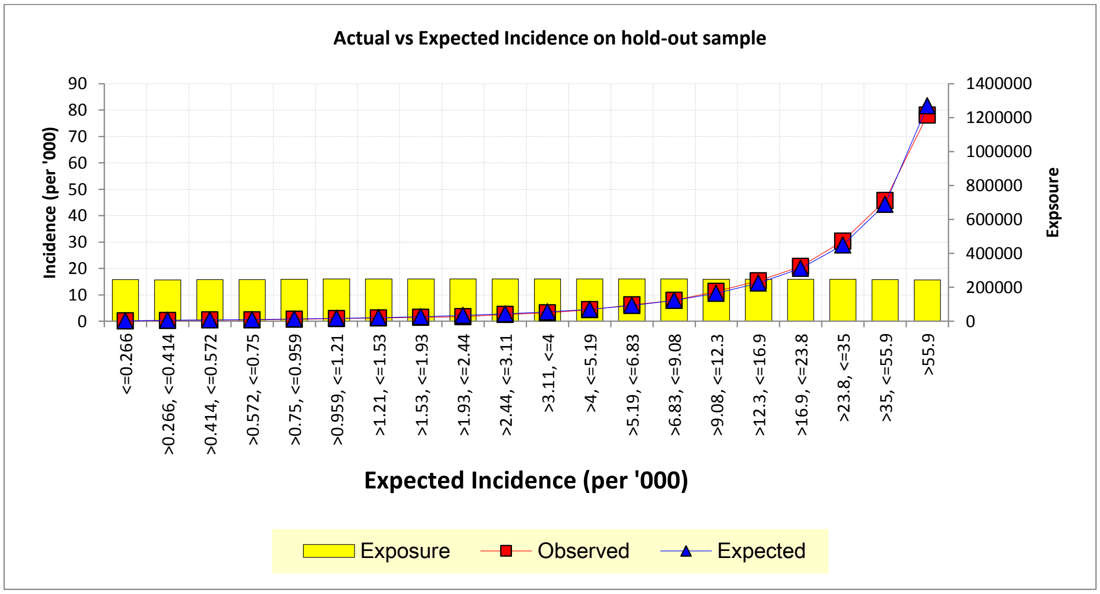

At the completion of the modelling phase and corresponding statistical review, the model included the following variables and interactions.

- Benefit Period
- Benefit Dollars interacted with Region
- Duration interacted with Premium Class
- Duration interacted with Underwriting Class
- Duration interacted with Tax Status
- Incurred Age interacted with Coverage Type
- Incurred Age interacted with Elimination Period
- Incurred Age interacted with Gender
- Incurred Age interacted with Marital Status

### 4.3 MODEL VALIDATION

Once the multivariate analysis was complete, the data not used in the multivariate analysis (i.e., one-third of the observations, or the “hold out data”) was used for validation purposes. For each observation in the hold out data, the expected incidence according to the model was calculated. The actual and expected incidence were summarized by the factors in the data, as well as by expected incidence, and compared. There were no significant differences between actual and expected incidence for levels in the data with reasonable levels of exposure, indicating that the model was a good predictor of incidence observed in the hold out data.

The graph below is a lift chart. Expected incidence is grouped into levels of approximately equal exposure, and actual and expected incidence are summarized by this expected incidence grouping. We see that the average actual and expected incidence in each group align well.

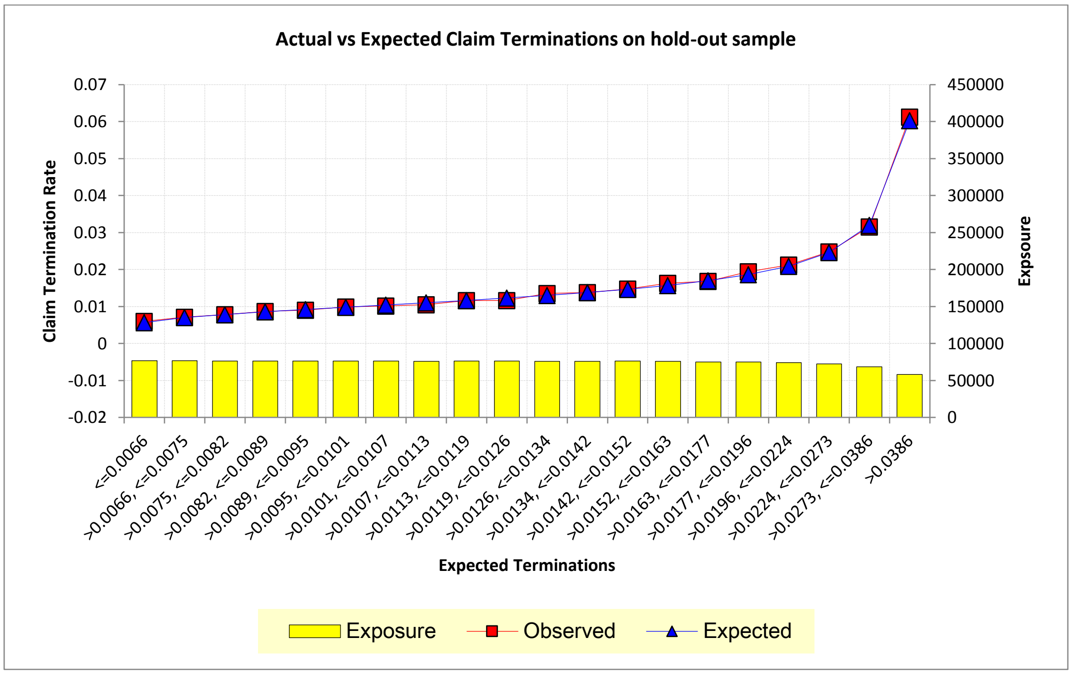



The lift chart, as well as the analysis of actual versus expected incidence summarized by factors in the data, suggested that the model is performing well.

## 4.4 OTHER MODELS

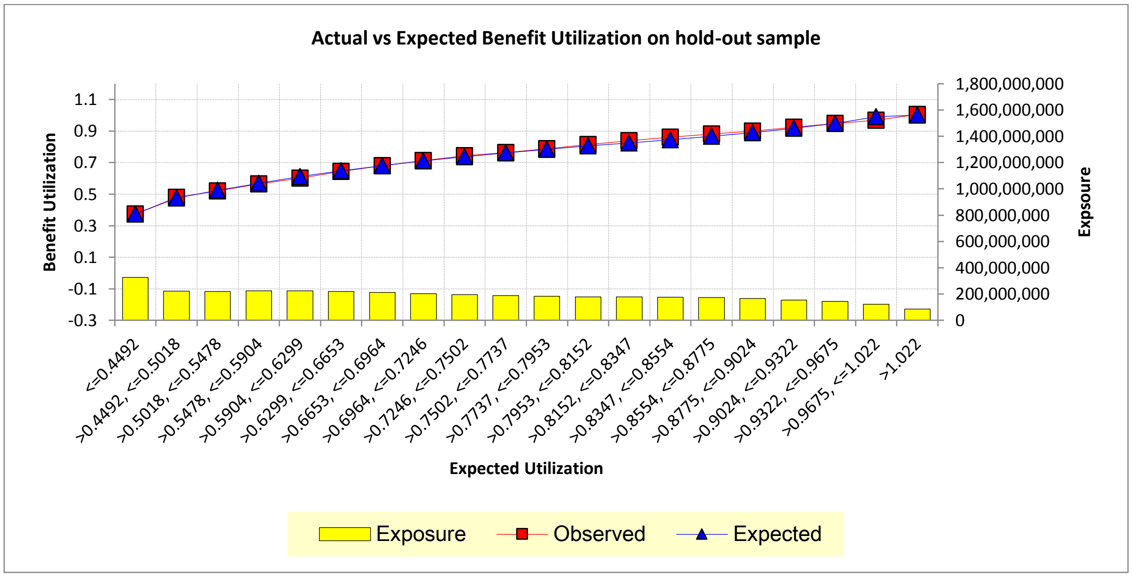
Like the Incidence models, the Claim Termination models were fit using a Log link function and Poisson error structure. Technically, this combination is not the most natural choice, as a proportion is restricted to the interval [0, 1] and the fitted values of a multiplicative model are not. However, the most natural choice (Logit link function and Binomial error structure), for which fitted values are restricted to the interval [0, 1], can be harder to interpret because it is not multiplicative. Further, our analysis (a comparison between Log-Poisson and Logit-Binomial models, as well as validation on a hold-out sample) suggested that, in this case, owing to the low overall level of claim terminations, the performance of the multiplicative model was comparable to that of a Logit-Binomial model.

A lift chart for the claim termination model is below. This, as well as the analysis of actual versus expected incidence summarized by factors in the data, suggested that the model is performing well.



The benefit utilization models were fit using a Log link function and Tweedie error structure. The Tweedie distribution is similar to the Gamma, but with mass at zero. This made it a natural choice for modeling utilization, for which values greater than or equal to zero were possible for each observation.

A lift chart for the benefit utilization model is below. This, as well as the analysis of actual versus expected incidence summarized by factors in the data, suggested that the model is performing well.



# SECTION 5: RELIANCES AND LIMITATIONS

## 5.1 RELIANCES

In developing this report, Towers Watson relied upon data and information supplied by the SOA and the participants, both in writing and in discussions. For each participant, this information includes, but is not limited to:

- Completed questionnaire on long term care incidence / lapse / mortality experience
- Data submission for claim incidence, claim termination and policy termination
- LIMRA report on data checks
- MIB reports on data checks

## 5.2 LIMITATIONS ON USE AND DISTRIBUTION OF REPORT

In accordance with a scope of work dated January 14, 2014, this report is provided for the use of the Society of Actuaries. We accept no responsibility for any consequences arising from any third party relying on the information provided in this report.

## SECTION 6: PARTICIPATING COMPANIES

- Allianz
- Berkshire Life
- CalPERS
- Continental Casualty (CNA)
- Fortis
- Genworth Financial
- John Hancock
- Lincoln Benefit Life
- Mass Mutual
- MetLife
- Mutual of Omaha
- New York Life Insurance Company
- Northwestern Mutual
- Penn Treaty
- Prudential
- Senior Health Insurance Company of Pennsylvania
- State Farm
- Thrivent AAL
- Thrivent LB
- Transamerica-Aegon
- United of Omaha
- UNUM

# APPENDIX A - MODELS

In the associated Microsoft Excel file ("2000-2011 SOA LTC Experience Study - Incidence Models 20150409.xls"), two incidence models with all of the factors are presented.

1. Model 1: presents incidence on a total life basis
2. Model 2: presents incidence on an active life basis

In the associated Microsoft Excel file ("2000-2011 SOA LTC Experience Study - Termination Models 20150717.xls"), four claim termination rate models with all of the factors are presented.

The claim termination rates were developed as the rate of claim termination per 0.5 months of exposure. This was due to the data manipulation required to ensure that individual claim exposure periods did not overlap policy duration, claim durations and calendar years. This allowed for the true effect of these items to be measured within the GLM model development.

As a result, the GLM model output shows the rate of termination for each 0.5 months of claim exposure. The Appendix A Microsoft Excel models illustrate the development of the model itself with two final claim termination rate columns on the summary tabs. The first column shows the output of the claim termination rate per 0.5 months of exposure and the second column shows the claim termination rate per full month of claim exposure.

1. Model 1: presents total claim termination rates with attributes of diagnosis and claim type
2. Model 2: presents total claim termination rates without the attributes of diagnosis and claim type
3. Model 3: presents claim termination rates due to death with attributes of diagnosis and claim type
4. Model 4: presents claim termination rates due to death without attributes of diagnosis and claim type

In the associated Microsoft Excel file ("2000-2011 SOA LTC Experience Study - Utilization Models 20150409.xls"), two benefit utilization rate models with all of the factors are presented.

1. Model 1: presents benefit utilization rates with the attributes of diagnosis and claim type
2. Model 2: presents benefit utilization rates without the attributes of diagnosis and claim type

# APPENDIX B - GENERALIZED LINEAR MODELING TECHNICAL BACKGROUND

The following explanation of GLM methodology is taken directly from the Basic Ratemaking text (Werner, G; Modlin, C; 2010, pages 176-177). This text is on the Casualty Actuarial Society's examination syllabus and is considered a primer in ratemaking theory and practice.

Both linear models (LMs) and GLMs aim to express the relationship between an observed response variable (Y) and a number of explanatory variables, referred to as predictor variables. For example, the response variable may be claim frequency for homeowners insurance, and the predictor variables may include amount of insurance, age of home and deductible. The observations in the data (e.g., claims on individual exposures) are considered a realization of the response variable.

Linear models express the response variable (Y) as the sum of its mean (μ) and a random variable (ε), also known as the error term:

$$Y = \mu + \varepsilon.$$

They assume the mean can be written as a linear combination of the predictor variables. For example,

$$Y = (\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4) + \varepsilon,$$

where X1, X2, X3 and X4 are each predictor variables, and β1, β2, β3 and β4 are the parameter estimates to be derived by the LM.

Linear models also assume that the random variable, ε, is normally distributed with a mean of zero and constant variance, σ2.

The aim of the linear model is to find the parameter estimates, which, when applied to the chosen model form, produce the observed data with the highest probability. The function used to achieve this aim is usually the likelihood function (or, equivalently, the log-likelihood). Maximum likelihood relies on linear algebra to solve a system of equations. In practice, due to the high volume of observations in classification ratemaking datasets, numerical techniques such as multi-dimensional Newton-Raphson algorithms are employed. These techniques find the maximum of a function by finding a zero in the function's first derivative. Also note that in the specific case of linear models, the likelihood function is equivalent to minimizing the sum of squared error between actual and indicated.

GLMs are a generalized version of linear models that remove the restrictions of the normality assumption and constant variance. They also allow a function, called the link function, to define the relationship between the expected response variable (e.g., claim severity) and the linear combination of the predictor variables (e.g., age of home, amount of insurance, etc.). The choice of various link functions means the predictor variables do not have to relate strictly in an additive fashion (as they do with LMs). For example, GLMs fit to insurance claims experience for ratemaking purposes often specify a log link function, which assumes the rating variables relate multiplicatively to one another. There are other components of the GLM formularization (offset terms, prior weights) that are beyond the scope of this document.

In order to solve a GLM, the modeler must:

- Supply a modeling dataset with a suitable number of observations of the response variable and associated predictor variables to be considered for modeling.
- Select a link function to define the relationship between the systematic and random components.
- Specify the distribution of the underlying random process, typically a member of the exponential family of distributions (e.g., normal, Poisson, gamma, Tweedie); this is done by specifying the mean and the variance of the distribution, the latter being a function of the mean.

The maximum likelihood approach then maximizes the logarithm of the likelihood function and computes the resultant values for each variable.

More comprehensive detail on the theory of GLMs is beyond the scope of this text, but may be found in Section 1 of “The Practitioner’s Guide to Generalized Linear Models” (Anderson , D., et al. 2005).

# APPENDIX C - DATA DEFINITIONS

Below is a list of the fields used in the models created and described in Appendix A:

- BenefitPaymentStyle: This field represents whether a policyholder has reimbursement or indemnity style payments associated with their claim benefits. Policies that have cash benefits would be considered in the indemnity category.
- ClaimType: The original site of care of the claim. This includes home health care, assisted living facility and nursing home. This field does not change if a claimant transfers from one type of care provider to another.
- Coverage type: This is policyholder coverage in terms of site of care options. Comprehensive coverage includes care at home, in assisted living or nursing home facilities.
- Daily_Ben_Dollars_Inflated: At the time of claim payment, the inflated amount of maximum daily benefit.
- Diagnosis_Category: The primary diagnosis at the beginning of the claim as reported by the company grouped into major ICD-9 code buckets.
- DurationMonth: Claim duration month.
- DurYear: The duration of the policy. Exposure and claims can be exposed to any duration as long as it falls in 2000-2011. Services rendered before 1/1/2000 or after 12/31/2011 were not included in any of the analysis.
- Gender: The sex of the policyholder or claimant.
- IncurredAge: Attained age at which the claim begins (at the beginning of the elimination period).
- IncurredPolicyYear: The policy duration in which a claim begins at the beginning of the elimination period.
- Infl_Rider_Description: The type of benefit inflation rider selected by the policyholder.
- Marital_Status: This is the marital status at issue. This field provides information as to whether a spousal discount was provided to a policyholder for being married. Married policyholders without a spouse having coverage were considered married in this analysis.
- MaxBP: This is the maximum benefit period for the policyholder. If there are differing benefit periods based on benefit type, the maximum is selected.
- MaxDB: The original maximum daily benefit purchased by the policyholder. These have been grouped into $10 increments.
- MinEP: The minimum elimination period of the policy in calendar days.
- Prem_Class: The class in which the policy was issued relative to the base policy. If an underwriting discount or load was given, then preferred or substandard was provided by the company. No alignment across companies was completed. In other words, all “preferred” policies are the same regardless of which company underwrote them.
- Region: The geographic region in which the policy was issued.
- TQ Status: Tax-qualified status of the policy. “N” represents policies that are not tax qualified. “Q” represents policies that are tax qualified.
- UW_Type: The self-reported type of underwriting provided by the company. They were grouped into “Full Underwriting” and “Other.” The Other includes guaranteed issue, simplified underwriting and “actively-at-work” definitions.

| State | Region | State | Region |
|---|---|---|---|
| AK | West | MT | West |
| AL | South | NC | South |
| AR | South | ND | Mid-West |
| AZ | West | NE | Mid-West |
| CA | West | NH | Northeast |
| CO | West | NJ | Northeast |
| CT | Northeast | NM | West |
| DC | South | NV | West |
| DE | South | NY | Northeast |
| FL | South | OH | Mid-West |
| GA | South | OK | South |
| HI | West | OR | West |
| IA | Mid-West | PA | Northeast |
| ID | West | RI | Northeast |
| IL | Mid-West | SC | South |
| IN | Mid-West | SD | Mid-West |
| KS | Mid-West | TN | South |
| KY | South | TX | South |
| LA | South | UT | West |
| MA | Northeast | VA | South |
| MD | South | VT | Northeast |
| ME | Northeast | WA | West |
| MI | Mid-West | WI | Mid-West |
| MN | Mid-West | WV | South |
| MO | Mid-West | WY | West |
| MS | South | | |

| DIAGNOSIS CATEGORY | ICD-9-CM CODES INCLUDED |
|---|---|
| ALZHEIMER'S | 290-294, 331, 797 |
| ARTHRITIS | 710-739 |
| CANCER | 141-239, V10, V16 |
| CIRCULATORY | 390-398, 410-417, 420-429, 440-459 |
| CONGENITAL | 742-756 |
| DIABETES | 250 |
| DIGESTIVE SYSTEM | 530-579 |
| ENDOCRINE,IMMUNITY SYSTEM | 242-249, 251-289 |
| GENITOURINARY SYSTEM | 580-618 |
| HYPERTENSION | 401-405 |
| INJURY | 800-996 |
| MENTAL | 295-319 |
| ILL-DEFINED AND MISCELLANEOUS CONDITIONS | 780-796, 798-799 |
| NERVOUS SYSTEM AND SENSE ORGANS | 323-330, 332-380 |
| RESPIRATORY | 011, 012, 460-496, 500-519 |
| SKIN AND SUBCUTANEOUS TISSUE | 680-709 |
| PREGNANCY DISORDERS | 620-674, 759-772 |
| STROKE | 430-438 |