LOGISTIC REGRESSION FOR INSURED MORTALITY EXPERIENCE STUDIES

Zhiwei Zhu\(^1\), Zhi Li\(^2\), David Wylde\(^3\), Michael Failor\(^4\), George Hrischenko\(^5\)

ABSTRACT

Properly adapted statistical modeling methodology can be a powerful tool for coping with a broad range of challenges related to life and annuity insurance industries’ experience studies. In this paper, we present a logistic regression model based U.S. insured mortality experience study with a focus on gaining study efficiency and effectiveness by addressing multiple analytical predicaments within one statistical modeling framework. These predicaments include but not limit to: a) test statistical significances or credibility of potential mortality drivers; b) estimate normalized mortality, slopes, and differentials; c) quantify study reliability; and d) extrapolate for under-experienced mortality, smooth between select and ultimate estimations, and assist basic experience table developments.

1. INTRODUCTION

Insured population mortality estimation is essential to (re)insurers’ developing liability expectations, meeting regulatory solvency requirements, and competing for market shares. The following three aspects are equally important in insured mortality studies.

a) Mortality trend: how mortality evolves over time
b) Mortality slope: how mortality increases or decreases by age or duration
c) Mortality differential: how mortality, mortality trend, and mortality slope vary between insured segments such as males vs. females or preferred class vs. standard class.

Insured mortality trends are often approximated based on general population mortality-trend studies. Although the life insurance industry collects an enormous amount of data, the collections usually have limited consistency and sufficiency for credible long term trend analyses. As a sub-group of the general population, insured population carries the mortality risk genes of the general population. It is natural for researchers and the industries to extend certain study methods, findings, and hypotheses of general population to the insured population. For example, Lee-Carter (1992) introduced a twin-model method for studying general population mortality: one model for fitting the population past experiences and the other for extrapolating future expectations. Eilers and Marx (1996) proposed the use of p-spline regression for fitting empirical data and smoothing the fit. Currie, Durban, and Eilers (2004)

---

1 Zhiwei Zhu, Ph.D., is vice president of risk modeling and analytics at SCOR Global Life Americas, zhu@scor.com
2 Zhi Li, Ph.D., FSA, CERA, is associate director of financial risk modeling at SCOR Global Life Americas, zli@scor.com
3 David Wylde, FSA, MAAA, CLU, ChFC, is pricing research actuary at SCOR Global Life Americas, dwylde@scor.com
4 Michael Failor, ASA, MAAA, is modeling actuary at SCOR Global Life Americas, mfailor@scor.com
5 George Hrischenko, FSA, MAAA is vice president, marketing actuary leader at SCOR Global Life Americas, ghrischenko@scor.com
applied the p-spline concept to projecting the U.K. insured lives experience. Both Lee-Carter and Currie’s studies formed the foundation of the insured mortality stochastic estimation tool developed by the U.K.’s Continuous Mortality Investigation Bureau (e.g. CMI 2005). Similarly, Li, Hardy, and Tan (2006) employed both Lee-Carte models and Currie’s p-spline to fit and to project Canadian general population mortality and then calibrated the learning to estimate Canadian insured mortality improvement. Their study provided supporting evidence for changes in the Standards of Practice for the valuation of insurance and annuity business in Canada (Canadian Institute of Actuaries 2010).

Insured mortality slopes by age are also frequently estimated by leveraging general population study methods and findings. There are two schools of thoughts related to aging. One school represented by Vaupel (2014) finds evidence that developed countries’ life expectancies have never stopped expanding and believes that mortality curve by age will continue to be stretched out to older ages. The other school follows Fries (1980) hypothesis that human life span has a limit and mortality will be more compressed into shorter age ranges close to that limit. Therefore, we should expect flattened and then spiking age slopes. In practices, insurers often rely on models that fit empirical experiences to estimate relationships of mortality and age. Since the Gompertz (1825) Law of Mortality was proposed, a variety of models that outline human mortality trajectory by age have been developed. Thatcher (1999) provided good descriptions and comparisons of four force of mortality models (with some simplifications):

(1.1) Gompertz(1825): \[ \mu \approx \alpha \cdot \exp(\beta \cdot x) \]

(1.2) Weibull(1951): \[ \mu = \alpha \cdot x^\beta \]

(1.3) Heligman and Pollard(1980): \[ \mu \approx \alpha - \frac{1}{2} \beta^2 + \beta \cdot x \]

(1.4) Kannisto(1994): \[ \mu = \frac{\alpha \cdot \exp(\beta \cdot x)}{1 + \alpha \cdot \exp(\beta \cdot x)} \]

His conclusion was: “when these four models are fitted to actual (general population) data, they are all relatively close to the data at ages where most of the deaths are concentrated, and hence relatively close to each other.” He also confirmed with various population data (Thatcher et al. 1998, Thatcher 1999) that the Kannisto model, or the logistic \( \mu \) model, that has a finite \( \mu \) asymptote approximates old age force of mortality well.

Insured mortality differential studies have to be more insured data driven and to the granular levels that cannot be dealt with the general population data. Comparing to the general population, the U.S. insured population is very dynamic. Insurers constantly and individually initiate risk selection efforts through improving underwriting, adjusting pricing strategies, expanding markets, and developing new products, which not only attracts various levels of new risks but also causes current insureds’ anti-selection reactions such as policy lapsedion and conversion. Insurers’ risk selection and consumers’ anti-selection activities form and reshape insured cohorts that are characterized not only by time or age but also by characteristics such
as smoker status, underwriting class, policy type, or a mix of these characteristics that usually are neither relevant to nor captured in general population data.

Insured mortality differentials have also been broadly studied. Many of the studies search for new explanatory variables other than age and gender to explain insured mortality or longevity variations. Some examples are Ornelas, Guillen and Alcañiz (2013), Kwon and Jones (2006), Vinsonhaler, Nalini, Jeyaraj, and Guy (2001). In this paper, we modify the Kannisto model (1.4) in two ways for analyzing the U.S. insured mortality differentials. Instead of modeling force of mortality $\mu$, we model probability of death $q$. In addition to age $x$, we test dozens of variables to select nine life insurance specific variables such as product and underwriting class to generate insights that are directly applicable for business decisions. In addition to confirming these insured mortality predictors, our main goal is to gain efficiency and effectiveness in addressing a spectrum of mortality risk assessment issues that are traditionally handled separately with descriptive select-mortality analyses, population data based ultimate mortality estimation, expert improvement & lapse adjustments, Whittaker-Henderson graduation, and so on. A common statistical modeling context is applied for

- Evaluating credibility of potential risk drivers (hypothesis test and variable reduction)
- Deriving normalized mortality slopes and differentials (parameterization and estimation)
- Verifying study reliability (model performance validation)
- Smoothing select and ultimate mortality (model design and link functions)
- Adjusting for short term trend (normalization by time)
- Adjusting for lapse effect (multinomial models, discussion only)
- Estimating advanced-age or under-experienced mortality (model interpolation and extrapolation)
- Constructing multidimensional basic experience tables (model scoring)

The rest of the paper is organized as follows: Section 2 summarizes the data used for this study; section 3 briefs logistic models and how to model mortality differentials; section 4 presents key findings; and section 5 discusses limitations and possible enhancements in using logistic regression models for industry experience studies.

2. THE DATA SOURCES

The Human Mortality Database (HMD) is the source for the U.S. general population mortality experience. At the time of our study, the database covered 1933 to 2010. The data is mainly used for comparison purpose, not for insured mortality approximation.

The insured experience data were collected by a major consulting company and a global reinsurer. The data file consists of experiences from more than 60 insurers with exposure from 2000 to 2009. A total of 174 million policy exposure years and 1.6 million death claims are available for study. In comparison, the data used for the Society of Actuaries (SOA) 2008
Valuation Basic Table (VBT) development were contributed by 35 companies, including about 700,000 death claims (SOA 2008).

To exhibit our study’s potentials for customization, we elected to present study findings based on a subset of the source data, policies issued since 1950 and face amount $\geq$ $50,000$, to better approximate some company’s target population and future experiences. The observed age ranges of the total and the selected data are both 1 to 109. The following table summarizes the distributions and q by attained age of the total and the selected data. The selected data have much lower q for all age groups, which reflect the mortality improvement and less underwriting wear-off.

**Table 2.1 Summary of the insured data**

<table>
<thead>
<tr>
<th>Sex</th>
<th>Attained age</th>
<th>Total data</th>
<th>Selected data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Claim count</td>
<td>Exposed count</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>00–22</td>
<td>3,525</td>
<td>6,303,991</td>
</tr>
<tr>
<td></td>
<td>23–27</td>
<td>3,105</td>
<td>3,304,725</td>
</tr>
<tr>
<td></td>
<td>28–32</td>
<td>4,175</td>
<td>5,964,346</td>
</tr>
<tr>
<td></td>
<td>33–37</td>
<td>7,204</td>
<td>10,166,729</td>
</tr>
<tr>
<td></td>
<td>38–42</td>
<td>13,114</td>
<td>13,884,778</td>
</tr>
<tr>
<td></td>
<td>43–47</td>
<td>22,948</td>
<td>16,060,846</td>
</tr>
<tr>
<td></td>
<td>48–52</td>
<td>36,977</td>
<td>16,212,737</td>
</tr>
<tr>
<td></td>
<td>53–57</td>
<td>54,632</td>
<td>14,819,343</td>
</tr>
<tr>
<td></td>
<td>58–62</td>
<td>73,629</td>
<td>12,072,373</td>
</tr>
<tr>
<td></td>
<td>63–67</td>
<td>89,983</td>
<td>8,450,035</td>
</tr>
<tr>
<td></td>
<td>68–72</td>
<td>109,391</td>
<td>5,993,105</td>
</tr>
<tr>
<td></td>
<td>73–77</td>
<td>146,537</td>
<td>4,640,211</td>
</tr>
<tr>
<td></td>
<td>78–high</td>
<td>490,630</td>
<td>6,539,738</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1,055,850</td>
<td>124,412,956</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>00–22</td>
<td>1,371</td>
<td>5,919,604</td>
</tr>
<tr>
<td></td>
<td>23–27</td>
<td>1,096</td>
<td>3,194,034</td>
</tr>
<tr>
<td></td>
<td>28–32</td>
<td>1,926</td>
<td>5,493,708</td>
</tr>
<tr>
<td></td>
<td>33–37</td>
<td>3,442</td>
<td>8,419,013</td>
</tr>
<tr>
<td></td>
<td>38–42</td>
<td>6,636</td>
<td>10,403,257</td>
</tr>
<tr>
<td></td>
<td>43–47</td>
<td>11,571</td>
<td>11,203,952</td>
</tr>
<tr>
<td></td>
<td>48–52</td>
<td>17,935</td>
<td>10,672,817</td>
</tr>
<tr>
<td></td>
<td>53–57</td>
<td>24,972</td>
<td>9,073,003</td>
</tr>
<tr>
<td></td>
<td>58–62</td>
<td>32,389</td>
<td>6,817,009</td>
</tr>
<tr>
<td></td>
<td>63–67</td>
<td>39,066</td>
<td>4,673,083</td>
</tr>
<tr>
<td></td>
<td>68–72</td>
<td>50,894</td>
<td>3,551,700</td>
</tr>
<tr>
<td></td>
<td>73–77</td>
<td>74,868</td>
<td>3,316,261</td>
</tr>
<tr>
<td></td>
<td>78–high</td>
<td>299,642</td>
<td>4,887,952</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>565,808</td>
<td>87,425,392</td>
</tr>
</tbody>
</table>
The $q$ in the table is defined as the number of deaths divided by number of exposure years. In this paper, mortality, mortality rate, death probability and death rate all refer to the same $q$, unless specified otherwise.

The following four charts compare the 2003–07 five-year mortality experiences based on the total data. These mortality rates are derived without normalizing by variables such as duration and underwriting class. According to Charts 2.3 and 2.4, for a given age group, permanent policyholders have much higher mortality than term policyholders. Later, we will compare these mortality to those obtained after normalizing by nine study variables. Normalization is difficult to do with conventional descriptive analyses.

3. ESTIMATING INSURED MORTALITY WITH LOGISTIC REGRESSION MODELS

For business decisions, the values of an analytical study or method are determined by many factors including: simple to understand, easy to apply, high tolerance of assumption violations, multiple-application and reliable. This governs the designs of our studies. Among the many statistical models, such as the well-known proportional hazard models, the more general generalized linear models (GLM), and the specific ones as displayed in (1.1) – (1.4), we choose to develop logistic models for four reasons:
• It models mortality $q$ that is simple to understand and direct for use in business operations and decisions.

• It produces parameter estimates that can easily be transformed and interpreted as estimates for mortality levels, slopes, and differentials. These metrics are also essential and straight for business decisions.

• It requires relatively lenient assumptions and works well for mortality risk assessments as proved by many researchers (e.g. Thatcher 1999).

• It has been successfully tested and broadly applied for risk analyses in many other fields such as credit, banking, P&C insurances, marketing, and medical researches.

In addition, it can be performed with widely available and well developed commercial software systems. More theoretical and applied explanations of logistic models are broadly available in literatures such as Harrell (2001), Frees (2009), and Hosmer, Lemeshow, & Rodney (2013).

Our logistic regression models have the general form of

$$q = \frac{e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots)}}{1 + e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots)}}$$

or

$$\ln \left( \frac{q}{1-q} \right) = \alpha + \sum_i \beta_i x_i + \cdots$$

(3.1)

where

$q$ is probability of death in an exposure year, given a policyholder survived to the beginning of the year

$x_i$ are explanatory variables (e.g., age, gender, duration, product) by which $q$ may vary

$\alpha$ is the intercept, to be estimated with experience data

$\beta_i$ are coefficients of the explanatory variables, to be estimated with experience data. An exponential transformation of $\beta_i$ may represent mortality slope or mortality ratio estimation, depending on designs such as if the corresponding $x_i$ is continuous or categorical.

The models may be enhanced with various technics such as adding interactions or applying variable transformations. Our tests reveal that only minimal model performance improvement would be achieved for our study data. For simple interpretation, in this paper we focus on main effect model (3.1).

As a reminder, for life/annuity applications, the probability of death $q$ and the odds of death $q/(1-q)$ are approximately equal for nearly all age groups because $1-q \approx 1$. This allows us to treat logistic modeling outputs to be reasonably interpreted in the terms of probability ratios or
mortality differentials, which is a unique and key attraction of using logistic models for mortality experience studies.

4. STUDY DESIGNS AND FINDINGS

As mentioned earlier, the goal of this paper is to present how to deliver multiple mortality experience analyses by following the same statistical modeling framework. In this section, we highlight the modeling designs that deliver four such analyses in four subsections:

1. Mortality driver credibility analysis
2. Mortality slope and differential analysis
3. Study reliability analysis
4. Potential application in basic experience table construction

In addition to the dependent variable of death claim indicator, nine observable explanatory variables listed in the following Table 4.1 are selected as exploratory variables from dozens of variables available in the source data through statistical selection algorithms and expert judgments on the quality of data collection, and the level of relevance to current business operations. Instead of attained age, the pair of issue age and duration is used to enable quantification of underwriting effect on mortality variations, or underwriting wear-off, which results in various levels of insured mortality estimations for policyholders in the same attained age group. This is a good example that general population mortality study methods using attained age alone may not be sufficient to capture insured mortality differentials.

Table 4.1 Model variable description

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observed value and description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male, female</td>
</tr>
<tr>
<td>Smoker status</td>
<td>Smoker, nonsmoker, unknown</td>
</tr>
<tr>
<td>Duration (years)</td>
<td>Time elapsed between policy issue and study year, 1 to 60 as a continuous variable</td>
</tr>
<tr>
<td>Issue age (last birth)</td>
<td>1 through 109 as a continuous variable</td>
</tr>
<tr>
<td>Study year</td>
<td>2000 through 2009 as a continuous variable</td>
</tr>
<tr>
<td>Face amount</td>
<td>Coverage amount, $50k-$99k, $100k-$499k, $500k+, inflation adjusted</td>
</tr>
<tr>
<td>Medical underwriting</td>
<td>With, without full medical underwriting</td>
</tr>
<tr>
<td>Product</td>
<td>Permanent, term</td>
</tr>
<tr>
<td>Underwriting class*</td>
<td>Preferred, standard, other</td>
</tr>
</tbody>
</table>

*today in the US, underwriters classify applicants into low (preferred), medium (standard), or high (substandard) risk classes based on information collected from applications, paramedical examinations, lab tests, medical records, etc. The number and definition of classes may vary by company and product.

The following are some descriptive statistics of the study data.
Table 4.2 Distribution by gender and smoker status

<table>
<thead>
<tr>
<th>Exposed Count</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsmoker</td>
<td>54%</td>
<td>35%</td>
<td>89%</td>
</tr>
<tr>
<td>Smoker</td>
<td>7%</td>
<td>4%</td>
<td>11%</td>
</tr>
<tr>
<td>Total</td>
<td>61%</td>
<td>39%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 4.3 Distribution by exposure year

<table>
<thead>
<tr>
<th>Study Year</th>
<th>Exposed Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>2,775,303</td>
<td>3%</td>
</tr>
<tr>
<td>2001</td>
<td>4,063,564</td>
<td>4%</td>
</tr>
<tr>
<td>2002</td>
<td>4,843,489</td>
<td>5%</td>
</tr>
<tr>
<td>2003</td>
<td>5,074,087</td>
<td>5%</td>
</tr>
<tr>
<td>2004</td>
<td>11,385,754</td>
<td>12%</td>
</tr>
<tr>
<td>2005</td>
<td>12,170,324</td>
<td>13%</td>
</tr>
<tr>
<td>2006</td>
<td>14,249,153</td>
<td>15%</td>
</tr>
<tr>
<td>2007</td>
<td>14,870,643</td>
<td>15%</td>
</tr>
<tr>
<td>2008</td>
<td>13,536,254</td>
<td>14%</td>
</tr>
<tr>
<td>2009</td>
<td>13,770,105</td>
<td>14%</td>
</tr>
</tbody>
</table>

Additionally, about 88% of the total volume went through full medical underwriting, 56% are associated with term products, and 61% are from policies with face amounts between $100,000 and $499,000. For the first 10 years of duration, the volume decreased smoothly from 10% to 5%, which covers 76% of business. Majority issue age (92%) is concentrate on issue age 22-62.

SAS software is used for data preparation and PROC logistic for model development. We retain the titles of the presented analysis outputs so that readers who are familiar with the software can have more insights about how the analyses are performed and interpreted.

For better reflection of common pricing practices in the US, we split the selected study data into four subsets and fit each subset with its own model (3.1). The four subsets are male smoker, male nonsmoker, female smoker and female nonsmoker. This separate modeling design allows each model’s coefficients to be estimated independently from the other three models, which means that each of the four policy groups’ slopes and differentials are estimated without being constrained by the other three groups. Since gender and smoker status are used to split the study data, they are not explicitly used as exploratory variables in this first modeling effort and no significant tests on these two variables are performed. A second modeling effort is taken to perform the tests on gender and smoker status, developing a full model with all nine exploratory variables using the non-split study data.

4.1 “Analysis of effects” for mortality driver identification and credibility quantification

The following table 4.4 summarizes the significance test results from both modeling efforts.
Table 4.4 “Analysis of effects”

<table>
<thead>
<tr>
<th>Pr &gt; ChiSq (probability-value)</th>
<th>Degree of Freedom</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Non-smoker</td>
<td>Smoker</td>
</tr>
<tr>
<td>Duration</td>
<td>1</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Issue age</td>
<td>1</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Study year</td>
<td>1</td>
<td>0.1719</td>
<td>0.0017</td>
</tr>
<tr>
<td>Face amount</td>
<td>2</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Medical underwriting</td>
<td>1</td>
<td>&lt;.0001</td>
<td>0.2496</td>
</tr>
<tr>
<td>Product</td>
<td>1</td>
<td>0.1363</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Underwriting class</td>
<td>2</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Smoker Status</td>
<td>2</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

Several points are worth noting.

1. As expected, insured mortality varies statistically significantly by gender, smoker status, duration, issue age, underwriting class, and face band for all four subgroups. This confirms that the industries are using some of the most reliable mortality predictors and performing reasonable underwriting classifications.

2. Study year, or exposure year, is included as a placeholder for mortality improvement in the 10-year period covered by the study. The corresponding p-values from the four models imply that, after factoring out what have been explained by the other eight explanatory variables, mortality variation explained by exposure year is statistically significant at $\alpha = 0.05$ only for male smokers (the next subsection will show a mortality decrease). This may imply that more male smokers ceased smoking and resulted in more mortality improvement during the studied period.

3. Mortality differentiation by product (between permanent and term policyholders) is only statistically significant for female nonsmokers and male smokers, after controlling the other eight explanatory variables.

4. At 95 percent confidence level, all nine tested variables have statistical significance in explaining mortality variation in at least one of the four policy groups, which is one of the reasons for us to include them in all four models. Finkelstein and Poterba (2004) studied the UK annuity sector with some other exploratory variables. Vinsonhaler et al. (2001) analyzed the US private pension plan experience data with similar logistic models and only found one significant explanatory variable. Comparing to our study, it implies that the US private pension population is either more homogeneous than the private life insurance population or more risk predictors are yet to be confirmed.

4.2 “Odds ratio estimate” for mortality slope and differential estimation

As a part of our study design, we treat three exploratory variables (issue age, duration and study year) as continuous for three reasons: 1) to estimate smoothed relationships between $q$ and these variables, 2) to allow the coefficients $\beta_i$ of these variables to be transformed as
mortality slope estimates, and 3) to enable model-based mortality extrapolation for older ages and later durations where sparse or no experience data are available. The modeled extrapolation can be useful for ultimate mortality estimation and basic experience table construction. It is worth to note that models can be configured differently to address issues from different angles. For example, to be more specifically quantify mortality variations in early select period such as durations 1, 2, and 3, one may categorize duration or include a categorical and a continuous duration variables in the model, if proper interpretation would be performed.

The values of the other six explanatory variables are categorized based on data credibility so that mortality differential estimation can be obtained through estimating the corresponding $\beta_i$.

From the same four subset models and the one full model as described in Section 4.1, “Odds ratio estimates” are generated and summarized in Table 4.5 and Table 4.6. As explained at the end of Section 3, we simply consider the odds ratios in the “Point estimate” columns as estimated mortality ratios while controlling the modeled exploratory variables.

**Table 4.5 Odds ratio estimates (subset models)**

| Effect (variable) | Male nonsmoker | | | Male smoker | | | Female nonsmoker | | | Female smoker |
|------------------|----------------|---|---|----------------|---|---|----------------|---|---|
|                  | Point estimate | 95% Wald confidence limits | Point estimate | 95% Wald confidence limits | Point estimate | 95% Wald confidence limits | Point estimate | 95% Wald confidence limits |
| Duration         | 1.141          | 1.139 | 1.143 | 1.118          | 1.114 | 1.122 | 1.157          | 1.153 | 1.160 | 1.133 | 1.126 | 1.139 |
| Issue age        | 1.101          | 1.100 | 1.102 | 1.093          | 1.092 | 1.094 | 1.105          | 1.104 | 1.105 | 1.098 | 1.096 | 1.099 |
| Study year       | 0.998          | 0.995 | 1.001 | 1.009          | 1.003 | 1.014 | 0.997          | 0.992 | 1.001 | 1.004 | 0.994 | 1.013 |
| Face $100k–$499k vs. $500k+ | 1.115 | 1.096 | 1.135 | 1.203          | 1.143 | 1.265 | 1.000          | 0.971 | 1.030 | 0.926 | 0.855 | 1.002 |
| Face $50k–$99k vs. $500k+ | 1.284 | 1.258 | 1.311 | 1.407          | 1.335 | 1.484 | 1.037          | 1.003 | 1.071 | 0.988 | 0.911 | 1.072 |
| Med vs Non-med   | 0.92           | 0.902 | 0.939 | 1.018          | 0.986 | 1.05  | 0.95           | 0.921 | 0.981 | 1.044 | 0.992 | 1.099 |
| Product perm vs. term | 1.013 | 0.996 | 1.030 | 0.923          | 0.890 | 0.958 | 1.033          | 1.006 | 1.060 | 0.998 | 0.939 | 1.061 |
| Class one-class vs. standard | 1.042 | 1.027 | 1.057 | 0.930          | 0.893 | 0.967 | 1.038          | 1.014 | 1.062 | 0.938 | 0.881 | 0.999 |
| Class preferred vs. standard | 0.730 | 0.719 | 0.741 | 0.748          | 0.717 | 0.781 | 0.740          | 0.722 | 0.758 | 0.767 | 0.715 | 0.823 |

**Table 4.6 Odds ratio estimates (full model)**

<table>
<thead>
<tr>
<th></th>
<th>Odds Ratio Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
<td>Point Estimate</td>
</tr>
<tr>
<td>Gender: Female vs. Male</td>
<td>0.722</td>
</tr>
<tr>
<td>Smoker: Nonsmoker vs. Smoker</td>
<td>0.441</td>
</tr>
<tr>
<td>Smoker: Unknown vs. Smoker</td>
<td>0.561</td>
</tr>
</tbody>
</table>

Let’s take the male nonsmoker model as an example and point out some interesting findings.

1. Duration and age slopes: when everything else were equal, on average, mortality increases about 14 percent per duration and about 10 percent per issue age (odds ratio = 1.14 and 1.10, respectively). The 10 percent per issue age (attained age at issue) increase is known to
be also true for the general population (Thatcher 1999), which indicates that our insured study findings in fundamental variables such as age is in line with that of general population studies.

2. If everything else were equal, there is a statistically insignificant 0.2 percent annual mortality improvement in the 10 year studied period (odds ratio = 0.998, the 95 percent confidence interval including 1). This finding may seem to be inconsistent with the common thought of higher mortality improvement. There are four possible explanations for this. First, in the past decade or so, U.S. population mortality improvement has been leveling off as shown in Chart 4.1 (data are from the Human Mortality Database; the age range reflects the most commonly insured ages). This may also be true for the insured population. Second, due to the short time period and inconsistent data contributions from insurers, the study data may have not captured the entire insured mortality improvement. Third, insurance underwriting has specifically targeted high death rate causes, such as cardiovascular diseases and smoking and excluded or discouraged these risks being insured, which may have resulted in that insured population has benefited less from the advancement in medicine, treatment and public education about these high mortality causes. Fourth, unlike a univariate analysis that attributes all the mortality variation to the single study variable, a large portion of the insured mortality improvement over the studied years can be attributed by the model to other factors such as the introductions of preferred classes, term products, and mortality compression.

3. If everything else were equal, permanent policy mortality would be about 1.3 percent higher than that of term policies (odds ratio = 1.013, insignificant). This may appear inconsistent with what is shown in charts 2.3 and 2.4. Keep in mind that the descriptive measures in charts 2.3 and 2.4 are obtained without controlling any other variables. Most of the differences displayed in charts 2.3 and 2.4 may be caused by unmatched duration, issue year and underwriting class distributions. The logistic mortality model provides an effective means to perform normalization.

4. Unlike univariate analysis finding, normalized inferential can vary by what and how other factors are included. For example, if issue age is replaced by attained age for our modeling, the coefficient of or mortality variation explained by duration is expected to be different from what has been presented in this paper and difference can be estimated by using the relationship since attain age = issue age + duration. In applied world, these multiple variable co-explained analyses findings can be sources for both insights and confusions.
Normalized mortality estimations are essential in identifying underlying causes and avoiding inappropriately setting pricing factors. They can also be useful for validating industry tables split from an aggregated table, like the American Council of Life Insurers’ 2001 Commissioner’s Standard Ordinary (CSO) preferred class structure tables.

4.3 “Model fit” for overall study reliability measurement

Compared to health or property and casualty insurance claims, mortality claims occur at a much lower frequency and with much more stable patterns. Relatively scarce claim counts and more consistent claim patterns led us to use all available data for model building, without setting aside data for over-fit verification.

One commonly used model-fit measuring statistic is c-statistic, or area under the receiver operating characteristic (ROC) curve. It measures combine predictive accuracy for both mortality (sensitivity) and survivals (specificity) as illustrated in Chart 4.2.

Though pricing actuaries may prioritize their attention to mortality estimation that drives the bottom line results, other insurance operations such as sales and underwriting often have a different focus that bias toward top line results. Therefore, performance statistics like c-statistic can be valuable to practitioners in two ways:

- Serve the need of maintaining a balance between prudency and competitiveness;
- Provide a quantitative metric for comparing predictive model based studies, such as the first note for the following table.

Table 4.7 displays the c-statistics for the four subset models.

<table>
<thead>
<tr>
<th>Association of predicted probabilities and observed responses</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonsmoker</td>
<td>smoker</td>
</tr>
<tr>
<td>c</td>
<td>0.679</td>
<td>0.747</td>
</tr>
</tbody>
</table>

Several items worth to note.

- Again, Vinsonhaler et al. (2001) analyzed private pension plan experience data with a similar but simpler logistic model (only one explanatory variable). Their model had c-statistics in the range of 0.51–0.59 for most of the age groups. Though we are not measuring c-statistic by age group and may have different target populations, the comparison still gives a sense that our four models are built on more exploratory
variables that explain a much larger portion of the mortality variation, which implies the possibility for the pension sector to collect more data and to identify more risk predictors.

- An interesting observation is that the c-statistics of the two nonsmoker models are about 10% lower than that of the two smoker models. This implies that the predictability by the same set of explanatory variables for smokers is about 10 percent higher than for nonsmokers. This 10 percent gain in death predictability is likely from knowing smoking status, which is another way to illustrate the predictive value of smoker status.

4.4 Model Scoring for Basic Experience Table Development

The development of industry mortality tables encompasses analytical exercises such as fitting, estimating, projecting, smoothing, and knowledge based adjustments in trend, selection, behavior, asymptotes, etc. Since our logistic model based experience study integrated many of these analytical exercises, it possesses a great potential to be used for supporting industry mortality table development. To verify and to demonstrate this potential, we performed an experiment including these steps:

- Calibrated the male and female non-smoker models so that they have asymptotes of 0.45 rather than the default 1. This is to approximate SOA’s capping population mortality at 0.45 at attained age its 110 for 2001 and 2008 VBT developments. In fact, this can also be viewed as a way to adjust for censored deaths due to causes such as policy lapsation. See the next section for more discussion.
- Generated mortality $q$ by the calibrated models. This $q$ varies by age and duration and is aggraged by other exploratory variables with exposure weighting.
- Organized the $q$ into age by duration two dimensional tables, which resembles the mortality experience tables.
- Compared the modeled tables to SOA 2001 VBTs and 2008 VBTs.

Since standard industry experience tables incorporate many more fine-tuning and adjustments than our experiment, we elect not to publish the details of our modeled tables. Instead, we provide the following six charts as a sneak view how the modeled tables compare to the SOA tables for some young, middle, and old issue age groups. As described earlier, the main reason that the charts are on issue age and duration basis rather than attained age basis is to reflect the US insured underwriting wear-off and the traditional mortality table structure. Readers who prefer attained age views may add the mid-point of the issue age range to the duration axis to approximate attained age for the corresponding issue age cohort.
By the fit to the experience and relativity to the SOA estimations, we consider that predictive models like our logistic models have promising potentials for industry experience table development.

5. CONSTRAINTS AND POSSIBLE ENHANCEMENTS

Three types of biases can occur in an insured mortality experience study: parameter bias, sampling bias, and data bias. A parameter bias is a systemic bias that reflects technical limitations of a study method (e.g., using a linear model to fit U-shaped experience). A sampling
bias happens when a substitute dataset is used to represent a target population but the substitute does not have the same characteristics of the target (e.g., using a small sample to represent a large population, or using past experience to approximate future outcomes). A data bias is the discrepancy between data and actuality (e.g., misreported ages of deaths or unrecorded lapse).

As in any other large databases, insured experience data consist of plenty data biases such as missing values and inconsistent data coding between companies. These data biases may compromise the quality of any studies. In addition, the following are a few more constraints of using logistic regression for insured experience studies.

1. Insurance experience data are repeated measurements that lack independence at specific policy level. This violates some of the theoretical assumptions for many statistical models including logistic models and introduces data or parameter biases. This longitudinal data issue exists in general population mortality databases as well. The unknown here is to what extent this violation tarnishes the models’ robustness for intended uses. Thatcher (1999) and others found that this violation is less of an issue for estimations with significant amount of experience data. Model extrapolation for old age and ultimate mortality, when scarce experience data are available, are most vulnerable to this violation.

2. Another form of assumption violation in using logistic models to fit insured experience data is policy lapsation. Logistic model (3.1) assumes that all study subjects have continuous risk exposures and asymptotically 100% death will occur. This is a more valid assumption for general population but invalid for death claims observed by insurers. There are various studies about what drive the lapsation of insured policies as a single outcome variable, e.g. Kiesenbauer (2012) and Kim (2005). The main issue here is to understand how claim and lapsation as two outcomes influence each other, similar to what Frees (2005) explored to analyze Pension Plan Termination and Retirement jointly. A potential solution to this issue is to apply multinomial logistic regression to model both claim and lapse jointly.

To understand the difference between death rate and claim rate, let’s reserve $q$ for death rate and assume that each insured policy has three observable statuses (and corresponding probabilities) at the end of an exposure year: lapse ($q_l$), claim ($q_c$) and in-force ($q_i$).

$$q_l + q_c + q_i = 100\%.$$

With the same explanatory variables $x$ as used in (3.2), we can use a multinomial logistic model to model the three probabilities as follows (see Hosmer, Lemeshow and Sturdivant 2013, chapter 8, for more descriptions).
Let us call this model a logistic \( q_c \) model to emphasize claim-rate. The added lapse component \( q_l \) in (5.1) plays a role of estimating the to-be-lapsed portion of exposures and excluding them from contributing deaths to claim-rate \( q_c \) estimation.

By comparing models (3.1) and (5.1), we have the relationship between death rate \( q \) modeled with binary models and claim rate \( q_c \) modeled with multinomial models as follows.

\[
\lim_{\text{duration} \to \infty} q = \lim_{\text{duration} \to \infty} (q_c + q_l) = 1
\]

which implies that multinomial models asymptotically splits the total death rate into a claimed portion and a lapsed portion. The asymptote of the claim portion is not necessarily 1.

\[
\lim_{\text{duration} \to \infty} q_c = \lim_{\text{duration} \to \infty} \frac{1}{1 + e^{(a_i - a_c) + (\beta_l - \beta_i) \text{duration}}} = \begin{cases} 
1 & \beta_i < \beta_l \\
1 & \beta_i = \beta_l \\
0 & \beta_i > \beta_l 
\end{cases}
\]

For projection purposes, \( a_i \) and \( a_c \) are usually related to initial lapse and claim levels; \( \beta_l \) and \( \beta_c \) are related to lapse and claim slopes. A highly simplified interpretation of (5.3) is that, depending on if the death rate asymptotically increases faster than, slower than or equal to the lapse rate of a portfolio, the portfolio claim rate will approach 100 percent, 0 percent or something in between.

3. Logistic models may not fit pre-marriage attained-age experience well. Mortality is usually high in these ages due to causes such as accidents and suicides. As the excess causes level off with age, mortality regresses to a more normal pattern that fits better with logistic function. The main strengths of logistic models are in aggregated mortality slope/differential estimation and model extrapolation. To improve fit, a possible solution could be to further customize a logistic \( q \) or \( q_c \) model with some spline or localized regression methods to fit the ages that have less regular mortality patterns.
4. When scarce experience data are available, such as at very old issue ages or later durations, a logistic function will be the primary driver for estimating modeled $q$ or $q_c$. For more accurate estimations, calibrations with expert knowledge are usually necessary.

5. Shock lapse and shock mortality that occur at the end of the level premium period or during rare events like pandemics cannot be fit or reflected well by a continuous function-based model. At a more granular level, modeling issues such as quantifying the end of the level period effect for a specific portfolio will need more than a logistic mortality model. However, at an industry aggregated level, our study shows logistic models deliver reasonable results.

6. The current lack of a consistently collected long-term insured experience database and consistently applied data collection standards (e.g. the definition of preferred classes) limits the optimization of any industry experience studies, including logistic mortality models. The inconsistency may appear as not all companies and not all product information are presented in an ad hoc industry-experience data collection or preferred classification differing among companies. Special cares are necessary in interpreting model outputs that implicitly assume the consistency. As data-processing technology and analytical methodology advance, it creates more instruments for the industries to establish a mechanism and to consistently collect comprehensive experience data for in-depth industry experience studies.

In summary, though with limitations, properly designed logistic regression offers to improve industry mortality experience studies through addressing multiple analytical challenges by following the same statistical principles and modeling framework.

ACKNOWLEDGEMENT

We would like to thank Jun Han, Daria Ossipova-Kachakhidze, Doris Azarcon and Pat Bresina for their insightful inputs and helpful suggestions.

We are also grateful to the reviewers’ constructive and invaluable feedbacks.

REFERENCES


