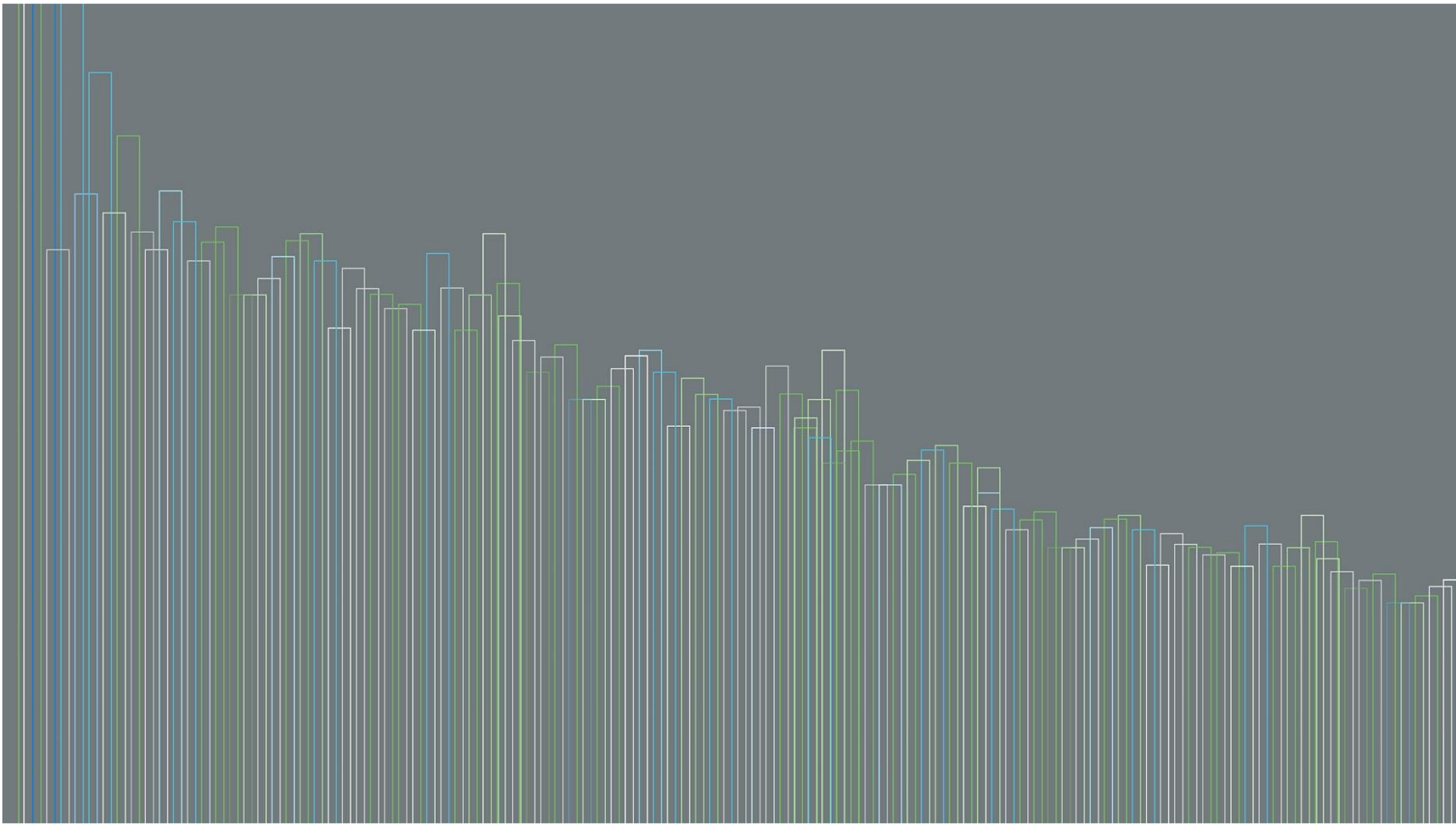
MILLIMAN RESEARCH REPORT

MIMSA III 2020

Study of mortality and lapse rates in level term life insurance

April 2020

Ken Qian, FSA, MAAA
Ben Johnson, MS
Jenny Jin, FSA, MAAA



Milliman

Table of Contents

EXECUTIVE SUMMARY **2**
Main takeaways for the lapse model 2
Main takeaway for the mortality model 3
1. INTRODUCTION **4**
DESCRIPTION OF DATA 4
STUDY OBJECTIVES 4
REPORT OUTLINE 5
2. SUMMARY OF 10-YEAR AND 15-YEAR LEVEL PREMIUM TERM DATA **5**
Number of annual records by policy duration 5
Premium increase 6
Attained age 8
Face amount 8
Premium collection type 9
Risk class group 9
3. OUR APPROACH AND METHODS **10**
PREDICTIVE MODELING 10
LOGISTIC REGRESSION 10
VARIABLE SELECTION PROCESS 11
MODEL VALIDATION 11
4. LAPSE MODEL RESULTS **12**
RESULTS BY POLICY DURATION 12
RESULTS BY CUMULATIVE PREMIUM INCREASE 13
RESULTS BY ATTAINED AGE 14
RESULTS BY FACE AMOUNT 16
RESULTS BY PREMIUM COLLECTION TYPE 17
RESULTS BY RISK CLASS GROUP 18
5. MORTALITY MODEL RESULTS **18**
6. CONCLUSION AND FUTURE ANALYSIS **19**
APPENDIX **20**
PREDICTIVE MODEL FITTING: GENERALIZED LINEAR MODEL 20
LAPSE MODEL: MODEL FORM AND COEFFICIENTS 21
ADDITIONAL LAPSE MODEL VALIDATION CHARTS 23
Lapse model: Validation by cumulative premium increase 23
Lapse model: Validation by attained age 24
Lapse model: Validation by face amount 25

Lapse model: Validation by premium collection type 25
Lapse model: Validation by risk group 26
MORTALITY MODEL: MODEL FORM AND COEFFICIENTS 27
MORTALITY MODEL VALIDATION CHARTS 27

Acknowledgments

This research would not be possible without the support and contributions from our participating companies, and for that we are very grateful. We also wish to thank our Milliman colleagues Al Klein and Melinda Willson for providing valuable guidance and peer review. Hyunsu Kim, Matthias Kullowatz, and many others around Milliman have helped us generate ideas and given feedback along the way, and we are grateful for their contributions.

Authors' notes

In the course of writing this report, we have tried to keep in mind readers who may have varying degrees of familiarity with the technical aspects of predictive analytics. Our goal is to make the report easy to follow while still covering important modeling details. We hope the readers will enjoy reading the report and also derive meaningful takeaways from our research.

Limitations and data reliance

This report is intended solely for educational purposes and presents information of a general nature. It is not intended to guide or determine any specific individual situation and persons should consult qualified professionals before taking specific actions.

In performing this analysis, we relied on data and other information provided by the companies participating in the third Milliman Industry Mortality Study and Analysis (MIMSA III). We have not audited or verified this data and other information. We performed a limited review of the data used directly in our analysis for reasonableness and consistency. In cases where small records in the data were found to have inconsistencies, these records were removed prior to modeling. A full audit of the data is beyond the scope of this assignment. If the underlying data or information is inaccurate or incomplete, the results of our analysis may change as well.

Limits on distribution

This report is intended for the internal use of the recipient, its affiliates, and their advisers. This report may not be distributed, disclosed, copied, or otherwise furnished to any other outside party without our prior consent. Any distribution of this report must be in its entirety. Nothing included in this report is to be used in any filings with any public body, such as but not limited to the U.S. Securities and Exchange Commission, without prior written consent from Milliman.

Executive Summary

This report discusses findings from our research on predicting lapse and mortality rates for 10-year and 15-year level premium term insurance products from the MIMSA III data set.

For this report, we analyzed over 18 million policy year records provided by over 25 companies, covering issue years from 1993 to 2014. The policy years in the data set were from 2007 to 2015. In this executive summary, we provide insights into the main predictors of lapse and mortality rates for post-level premium term products.

Main takeaways for the lapse model

Lapse rates generally decrease by duration during the level premium and post-level premium periods, but spike significantly during the last duration of the level premium period and first duration of the post-level premium period. Lapse rates typically average around 7% at the start of the duration, and decrease steadily as duration increases. Lapse rates spike drastically during the last year of the level premium period as policyholders look to avoid paying higher premiums. On average, our model predicts that lapse rates increase to about 70% during the last year of the level premium period. This lapse rate shock will gradually fall during the next couple years of the post-level premium period before stabilizing again. However, average lapse rates in the post-level premium period are still predicted to be higher than in the guarantee period.

Higher cumulative premium increases result in higher lapse rates during the post-level premium period. However, the sensitivity of lapse rates to premium jumps decreases for later durations in the post-level premium period. The majority of policyholders who decide to lapse are typically healthy individuals who are likely able to obtain coverage at lower premium rates than are offered in the post-level premium period. The increase in lapse rates from higher cumulative premium increases can be quite significant. For example, a 65-year-old male nonsmoker whose premium increased by a cumulative amount of seven times during the first year of the post-level premium period is more than twice as likely to lapse than the same policyholder whose premium only increased by two times.

For most durations in the level premium period and post-level premium period, lapse rates are lower for older attained ages. However, during the last year of the level premium period and first year of the post-level premium period, which is when the premium increase is most significant, lapse rates are *higher* for older attained ages. For instance, during the last duration of the level premium period for a male nonsmoker with a 10-year level premium term product that faces a two-times jump in annual premium the next year, our model predicts that an 80-year-old policyholder is more than 1.5 times more likely to lapse than a 30-year-old policyholder. The relationship between lapse rates and age persists in our model during the first year of the post-level premium period, but will not be as strong.

Lapse rates are predicted to decrease with increasing face amount during the T-1 durations in the level premium period for a T-year level premium term product. However, during the last duration of the level premium period and in the post-level premium period, lapse rates are predicted to *increase* as a function of face amount. This relationship between lapse rates and face amount during the post-level premium period indicates that policyholders are sensitive to the *dollar increase* in the premium payments as well. For a given percentage premium jump, policyholders with higher face amounts will face an overall higher dollar amount of premium increases.

Lapse rates are lower for policies that pay premium through preauthorized checking accounts or credit cards than for policies that pay premium through direct billing. This relationship holds, regardless of which duration the policy is in. The reason for this is policyholders who pay with preauthorized checking or credit card have to take manual action to stop the payments. This higher effort to cancel the policy leads to better retention and persistency.

Lapse rates are projected to be higher for worse (less healthy) risk class groups. These disparities in lapse predictions across risk class groups are consistent with the data across different policy durations.

Main takeaway for the mortality model

Mortality is progressively worse over time for policyholders who experience higher cumulative premium increases during the post-level premium period. The presence of policies that do not lapse after experiencing higher premium increases materially changes the risk profile of the business. Policyholders who decide to persist in the midst of more extreme premium increases implicitly pass along a signal that they have a greater need for the insurance policy, and thus have potential health conditions that will lead to higher mortality rates. Controlling for the impact of all other variables, our model predicts that a policy whose post-level term premium increases relative to their guaranteed premium will have mortality rates that are progressively higher.

1. Introduction

DESCRIPTION OF DATA

We performed analysis on 10-year and 15-year level premium term insurance products from the MIMSA III data set, which consists of annual policy data provided by over 25 companies for policy years from 2007 to 2015. The total data set consisted of about 4 million unique policies, with a total of about 18 million annual policy-year records. The MIMSA III data set includes all standard ordinary product types.

The objective of our study was to build a predictive model that analyzes the drivers of lapse and mortality rates for level premium term insurance products. Each record used in this analysis contains information about the policyholder (i.e., term length, premium amount, risk class), as well as an indicator for whether or not the policyholder lapsed or died in a given year.

STUDY OBJECTIVES

Level premium term products have a *level premium period* where the annual premium is fixed for a specified period (e.g., 10 years). After the level premium period, the product enters a period called the *post-level premium period* where the annual premiums are generally higher than the aforementioned level premium. There is a predetermined guaranteed premium increase schedule, but companies may or may not choose to follow this schedule. In fact, some companies find it more profitable to charge premiums lower than this schedule to retain more and healthier policyholders. Still, the initial premium increases can be very steep, such as a 500% increase from the level premium period.

For our study, we were interested in quantifying the impact higher premium increases will have on lapse and mortality rates. In addition, we were also interested in quantifying the impact of common drivers of lapse and mortality rates, and how these impacts differ between the level premium period and post-level premium period. In particular, we wanted to answer the following questions:

1. What is the impact from higher premium increases after the level premium period ends on the lapse and mortality rates during the post-level premium period?
2. What is the relationship between lapse rates and mortality rates with other potential predictors (e.g., attained age, face amount), and how does this relationship change between the level premium period and post-level premium periods?

We focused on analyzing 10-year level premium term and 15-year level premium term products. These were the most common level premium periods with sufficient post-level premium exposure to study. The 20-year level premium term is the most commonly sold level premium term product in the industry, but these products were just reaching the end of the level premium period shortly before the end of our study window and there was not sufficient data to study.

Our analysis of the impact of premium increases on lapse and mortality rates in the post-level premium period is restricted to level-term business that was issued from the mid-1990s to early 2000s. This is because the mid-1990s was when level premium term plans began to be written and there is not any post-level premium exposure for policies issued after 2005, because the latest study year in the MIMSA III data set is 2015. The post-level premium schedule for policies issued from the mid-1990s to early 2000s is characterized by very sharp initial premium increases (i.e., 600% to 700%), followed by smaller annual premium increases. More recently, companies have implemented smaller increases in premiums after the level premium period is over, which then grade up to higher premium amounts. As a result, our models may be less predictive for this newer issued block of business, because no policies from this newer block have reached the post-level premium period yet.

REPORT OUTLINE

The outline for the rest of the report will be as follows:

- **Section 2** focuses on data exploration and gives summaries of the data across some of the key predictors in our model.
- **Section 3** discusses the methods we used to build a predictive model for both lapses and mortality. In particular, we focus on the model form, as well as how we approach variable selection and engineering the different features for the predictive model.
- **Section 4** discusses details of the lapse model that were used to fit the data, key insights and interpretations from the model, and validation of the model results.
- **Section 5** provides a similar discussion for the mortality model.
- **Section 6** concludes with key takeaways from the analysis, and any additional analysis that did not fall within the scope of this project.

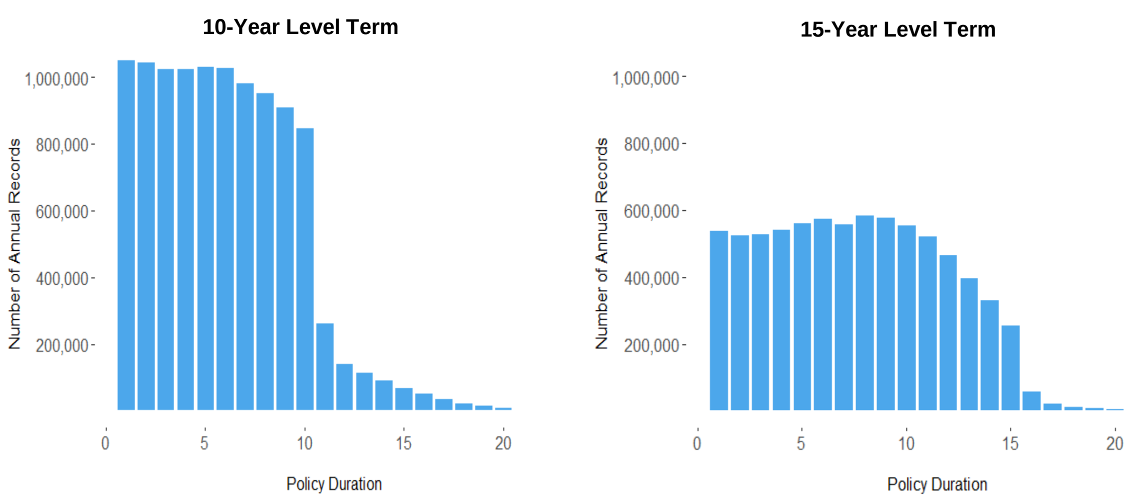
2. Summary of 10-year and 15-year level premium term data

This section summarizes the 10-year and 15-year level premium term data from the MIMSA III data set across key variables that are used in our analysis. For both lapse and mortality models, the target variable is an indicator for whether or not a policy lapsed (or the insured died) in a given policy year, *conditional* upon the policy being active at the start of the policy year. In our data, we checked the distribution of the data by company and no company was found to dominate the exposure.

Number of annual records by policy duration

Figure 1 shows the number of annual records by duration for the 10-year and 15-year level premium term products in two parts.

FIGURE 1A: NUMBER OF ANNUAL RECORDS BY POLICY DURATION



The vast majority of the records occur during the level premium period, as many policies lapse before the post-level premium period begins. Most of these lapses occur during the last year of the level premium period as policyholders try to avoid paying the higher premiums during the post-level premium period. From Figure 1b, we see that about 5% of the total annual number of records are in the post-level premium period. However, this is still a little over 1 million observations, providing enough credibility in our analysis of lapse and mortality rates in the post-level premium period.

FIGURE 1B: PERCENTAGE OF ANNUAL RECORDS BY POLICY PERIOD

	NUMBER OF ANNUAL RECORDS	PERCENTAGE OF ANNUAL RECORDS
First T-1 Durations of Level Premium Period (T = Length of Level Premium Period)	16,425,701	88.78%
Last Duration of Level Premium Period	1,106,872	5.98%
Post-Level Premium Period	969,960	5.24%

Premium increase

Our data set included over 730,000 annual records in the last duration of the level premium period and in the post-level premium period with information on their premium schedules. In our analysis, we define a “Premium Jump” to be the increase in premiums compared to the level term premium amount. Thus, if a premium jump factor is 1x, then the premiums stayed the same from the prior year. If a premium jump factor is 2x, then the premium doubled from the prior year.

Figure 2a plots the distribution of the *initial* premium jump for policies upon reaching the last year of their level premium periods. The distribution is quite wide, and 50% of the data has premium jumps between 4x and 10x. This also indicates that many policyholders face very large premium increases in the first year of the post-level premium period. For instance, Figure 2b shows that the median initial premium jump was 7.21, indicating that over half the policies in our data set (where reasonable premium information was available) had a 621% increase in premium in the first year of the post-level premium period. The average initial premium jump was even higher at 8.22.

FIGURE 2A: NUMBER OF ANNUAL RECORDS BY INITIAL PREMIUM JUMP

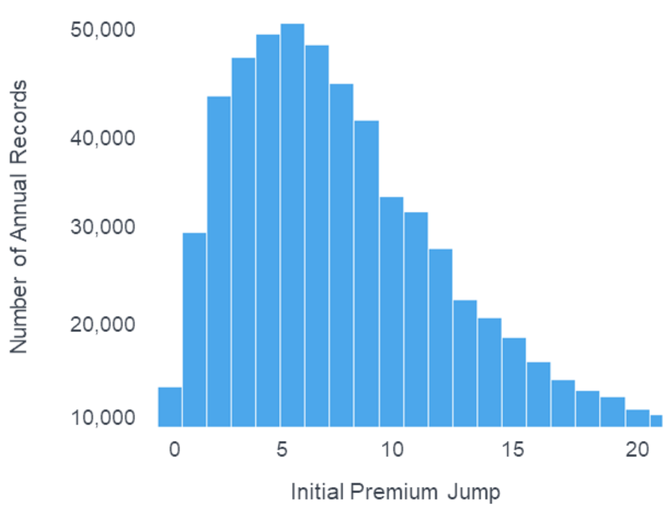


FIGURE 2B: INITIAL PREMIUM JUMPS

	INITIAL PREMIUM JUMP
Average	8.22
Median	7.21

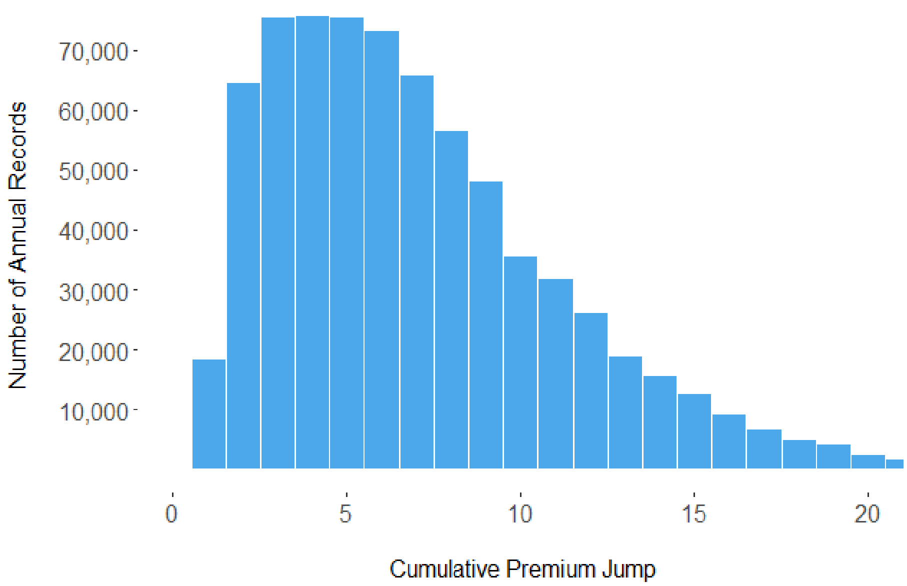
There are a few different ways of engineering a feature in a predictive model to quantify the increase in premium jumps a policyholder experiences after the level premium period ends. Some previous studies on this subject have focused solely on the impact the initial premium increase has on lapse and mortality rates. However, this does not take into account the premium increase information in later years, as the premiums can change year by year after the post-level premium period ends. In order to appropriately reflect premium increases in later durations of the post-level premium period, while still preserving the predictive signal from the first year's premium jump, we parametrized premium increase by looking at the *cumulative* premium increase relative to the level premium.

To illustrate this, suppose a policy's level premium was 100. Now suppose the premium increased to 450 in the first year of the post-level premium period, and then increased to 480 in the second year of the post-level premium period. The cumulative premium increase would be 4.5 in year 1 of the post-level premium period, and then 4.8 in year 2 of the post-level premium period (i.e., 4.5x in year 1 and 1.067x incrementally in year 2).

Parametrizing premium increase as a cumulative product of annual increases reflects that how policyholders respond to increases in their premium is not based solely on the increase experienced in the prior year. If policyholders have experienced consistent, small increases in their premium year-over-year, then they may behave differently from those who have seen their first increases in premium.

Figure 3 displays the number of annual records by cumulative premium jump for all observations in the post-level premium period. The shape of this distribution is similar to the shape in the histogram in Figure 2A, which focused only on initial premium jump. This reflects the fact that, although there are premium increases in the data each year during the post-level premium period, typically the initial premium jump will be the largest.

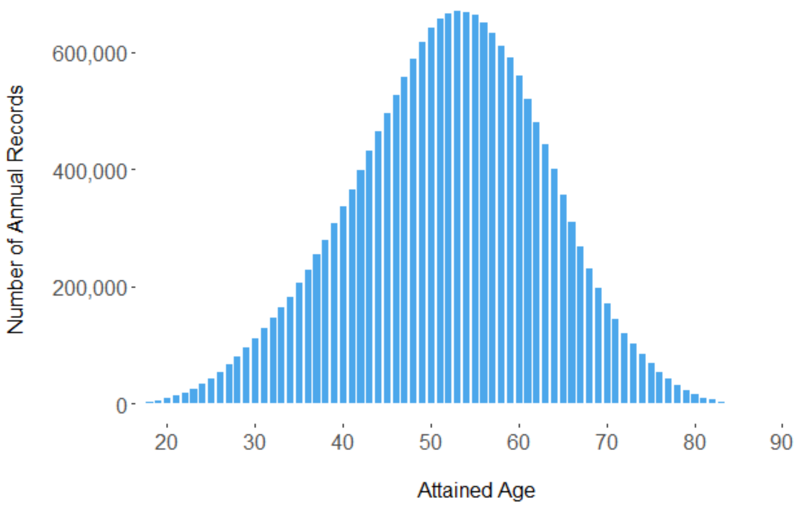
FIGURE 3: NUMBER OF ANNUAL RECORDS IN POST-LEVEL PREMIUM PERIOD -BY CUMULATIVE PREMIUM JUMP



Attained age

Figure 4 displays the number of annual records by attained age for the level premium term data set—50% of the data had attained ages between 45 and 60.

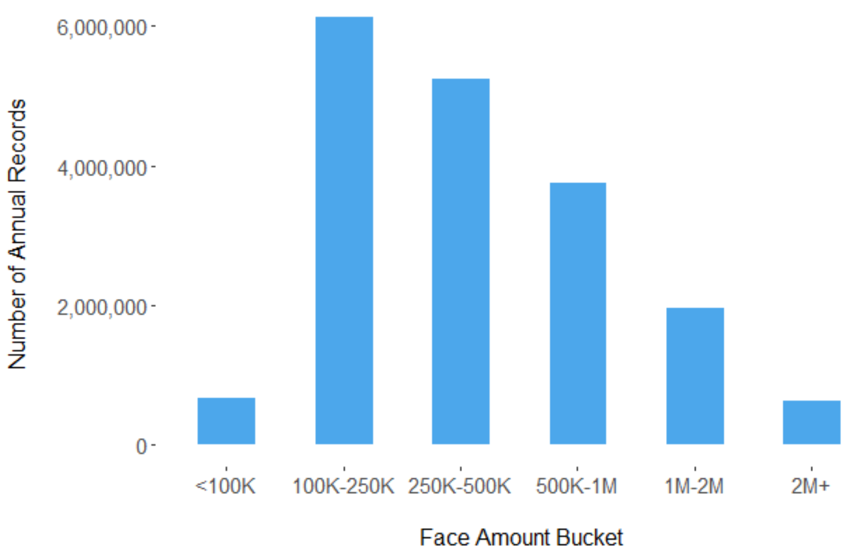
FIGURE 4: NUMBER OF ANNUAL RECORDS BY ATTAINED AGE



Face amount

Figure 5 displays the distribution by face amount at issue for the level premium term data set used in our analysis, grouped into different buckets. The largest exposure groups come from policies with face amounts of $100,000 to $250,000 and $250,000 to $500,000.

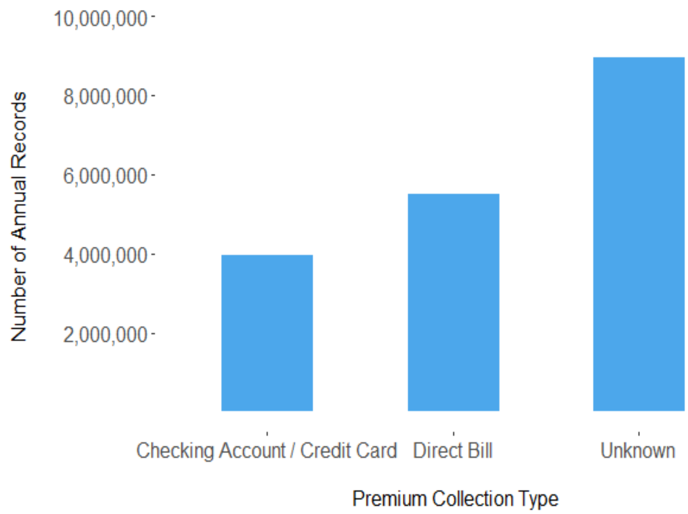
FIGURE 5: NUMBER OF ANNUAL RECORDS BY FACE AMOUNT



Premium collection type

Figure 6 displays the distribution by premium collection type. The premium collection type refers to how payments are made in a policy. Of policies that provided data for premium collection type, most belong to direct billing, followed by automatic payments from a preauthorized checking account or credit card. The number of observations by premium collection type is shown in Figure 6. Note that there were quite a few records in the data set where premium collection type information was not provided.

FIGURE 6: NUMBER OF ANNUAL RECORDS BY PREMIUM COLLECTION TYPE



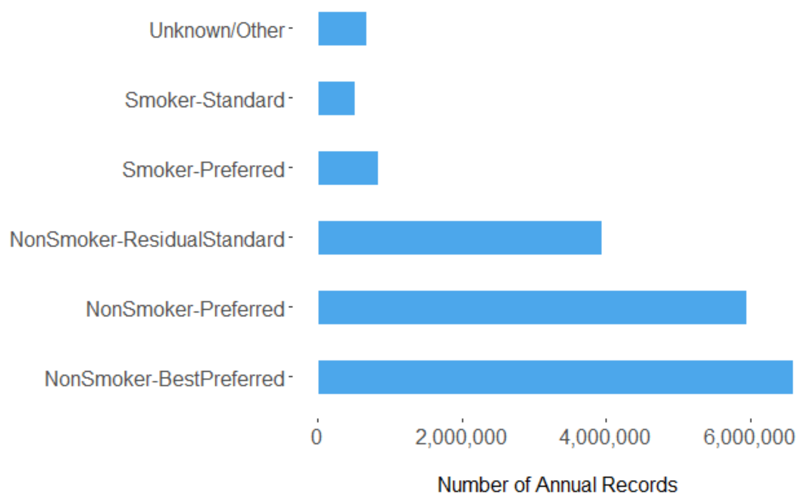
Risk class group

The risk class group of a policy combines information regarding the policy's underwriting and smoker status. Policies are assigned to one of the following categories:

1. Nonsmoker - Best Preferred
2. Nonsmoker – Preferred
3. Nonsmoker – Residual Standard
4. Smoker – Preferred
5. Smoker – Standard
6. Unknown

Figure 7 shows the number of annual records by this risk-class group variable. Most of the records are nonsmokers with the largest exposures in risk classes Best Preferred, Preferred, and Residual Standard, respectively.

FIGURE 7: NUMBER OF ANNUAL RECORDS BY RISK CLASS GROUP



3. Our approach and methods

PREDICTIVE MODELING

For our analysis, we chose to build a predictive model to forecast lapse and mortality rates. In the simplest terms, predictive modeling is the use of data and mathematical models to improve predictions of future events. In this paper, we use a predictive model to estimate lapse and mortality rates for a given set of policyholder features. The following is a list of benefits from using predictive models:

1. **Predictive modeling helps separate signal from noise**. In particular, it can help isolate the impact of a specific predictor on lapse and mortality rates, after controlling for the other predictors in the model. For example, a predictive model can tell us what the impact of attained age is on lapse rates after controlling for other features of a policy and policyholder, such as face amount, premium increase, sex, etc.

2. **A predictive model can be constructed with common variables in a typical in-force file to allow ease of implementation**. The models presented in this paper use drivers that are readily available in a typical in-force data file, making them suitable for implementation in existing pricing and valuation platforms.

3. **Predictive modeling of lapse and mortality behavior offers a statistical framework for demonstrating assumption effectiveness to internal and external stakeholders**. Rating agencies and regulators are placing higher scrutiny on how companies set assumptions around policyholder behavior. A predictive model built on statistical principles can provide sound validation metrics for measuring the effects of explanatory variables and the accuracy of the predictions.

LOGISTIC REGRESSION

For this analysis, we fit a logistic regression model to the MIMSA III level premium term data to predict lapse and mortality rates. Logistic regression was a convenient choice to model lapse behavior because the lapse response variable on a data set with annual records is a binary event, where each policyholder will either lapse in the subsequent year or they will not. The mortality response variable is also binary, so the same reasoning applied to modeling lapse with logistic regression also applies to using logistic regression to model mortality. Unlike with lapse, there already exist industry standard tables used to predict mortality. Instead of building a predictive model completely from scratch, we've built our mortality model to further improve predictions from using the Society of Actuaries (SOA) 2015 Valuation Basic Table, RR 100 (2015 VBT) tables, by considering the effects of additional variables available in the MIMSA III data set. We have fit a logistic regression model that uses predictions from the 2015 VBT as an offset to fine-tune the life tables to the experience of the MIMSA III level premium term data. More technical details on logistic regression can be found in the appendix

VARIABLE SELECTION PROCESS

We explored the data set and used business judgment to select the set of predictors and potential interaction terms to include in our model. Based on this analysis, some of the key predictors we included in our lapse and mortality models were:

1. Attained age
2. Sex
3. Premium collection type
4. Risk class group
5. Face amount
6. Cumulative premium increase
7. The percentage of time remaining in the level premium period of a policy
8. Duration in the post-level premium period
   - This variable represents how many years a policyholder has been in the post-level premium period. For example, in year 17 for a 10-year level premium product, the value of this variable is set to 7
9. Term phase
   - This is a categorical variable that classifies the durational period that a policyholder is in. For a product with a level premium period of T years, there will be four different term phases:
     1. Durations <= T - 1
     2. Duration T
     3. Duration T + 1
     4. Durations >= T + 2
10. The length of the level premium period of the product

In addition to the main effect variables described above, we also modeled a few interaction terms between these main predictors. We then fit the specific model form to our data set. More detail about the model form and coefficients can be found in the appendix.

MODEL VALIDATION

In order to facilitate validation of our predictive models, we randomly partitioned our data into two subsets. We fit our predictive model to a training data set consisting of 80% of the entire data set while the remaining 20%, the holdout data set, was reserved for testing the model. Holdout data does not influence the model coefficients so using it to validate our model will highlight whether our model is over-fit to our training data set.

Comparing model predictions on the holdout against observed rates will assess how well our predictive model explains the relationships between the models' independent and dependent variables, in addition to evaluating the usefulness of our predictions on blocks of business that are not included in the MIMSA III data set. Generally, the ratio of actual to predicted lapse (or mortality) rates should be close to 1 in order to assure the reasonability of the model's predictive capabilities.

In the rest of this paper, we will discuss key insights from our lapse and mortality models. Model validation results that compare the predictions from our final lapse and mortality models with the observed lapse and mortality rates from the holdout data set can be found in the Appendix.

4. Lapse model results

To graphically illustrate predictions from our models, we used a sample policyholder with the characteristics shown in Figure 8 to make predictions.

FIGURE 8: SAMPLE POLICYHOLDER

FEATURE	VALUE
Sex	Male
Age	65
Risk class group	Nonsmoker (Best Preferred)
Face amount	$500,000
Premium collection type	Direct billing
Term length	10 years

For the rest of this paper, graphs that illustrate model predictions using a "sample policyholder" will refer to a policyholder with the characteristics shown in Figure 8, unless specified otherwise.

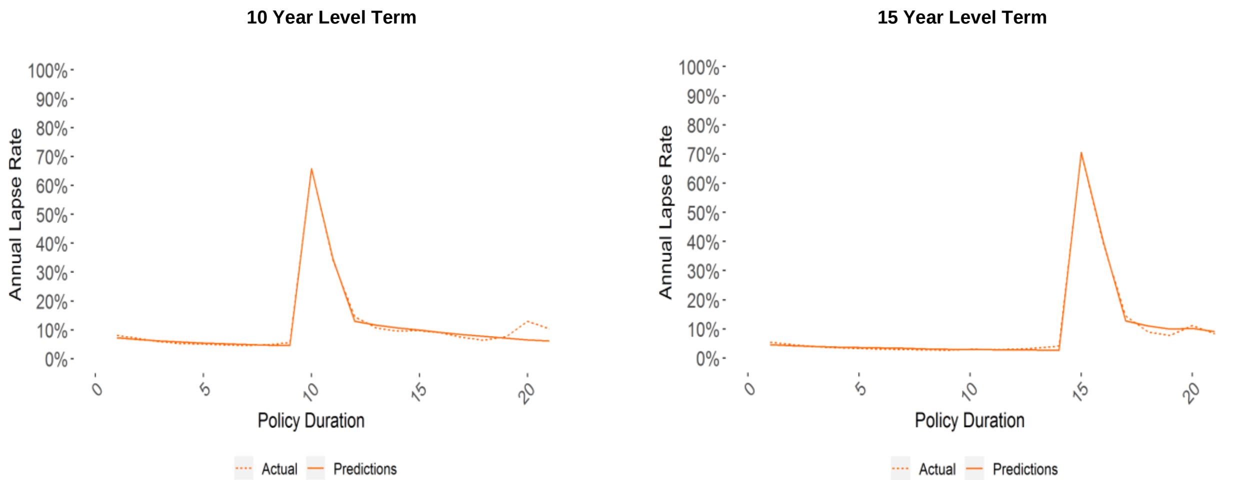
RESULTS BY POLICY DURATION

Lapse rates generally decrease by duration during the level premium, spike significantly during the last duration of the level premium period and first duration of the post-level premium period, and then decrease by duration again.

Figure 9 compares the predicted and actual lapse rates for the 10-year and 15-year level premium term products by duration in our holdout data set.

- As expected, the model predicts that lapse rates will follow the pattern just described.
- Our model predicts a large spike in lapse rates in the last duration of the level premium period because of the higher premium increases during the post-level premium period.
- Our model also predicts that lapse rates during the *first year* of the post-level premium period (duration 11 for the 10-year level premium term, and duration 16 for the 15-year level premium term) are significantly higher than lapse rates during the level premium period, also because of the higher premium increases..
- For the rest of the durations in the post-level premium period, lapse rates stabilize to the same trend observed during the level premium period. However, our model still predicts that lapse rates during the post-level premium period are slightly higher than lapse rates in the guarantee period.

FIGURE 9: RESULTS BY POLICY DURATION



RESULTS BY CUMULATIVE PREMIUM INCREASE

Higher cumulative premium increases result in higher lapse rates during the post-level premium period. However, the sensitivity of lapse rates to premium increases flattens out for later durations in the post-level premium period.

Figure 10 plots predicted lapse rates from our model for the sample policyholder as a function of the cumulative premium increase during the first four durations of the post-level premium product for a 10-year level premium term product. A table is also provided in Figure 11 with data supporting the chart in Figure 10. Our model predicts that higher cumulative premium increases will result in higher lapses. For instance, our model predicts that in duration 11 our sample policy would have a 66.68% chance of lapsing under a 10x cumulative premium increase. In contrast, our sample policy would only have a 15.32% chance of lapsing if there were no cumulative premium increase (i.e., 1x). This is more than a fourfold increase in the lapse rate due to premium increases.

- Our model also suggests that the slope of the predicted lapse rates as a function of premium increase becomes less steep for higher durations in the post-level premium period. One explanation is that most of the healthy policyholders who do not value the insurance policy enough to pay the higher premiums will lapse in the earlier durations of the post-level premium period. During the later durations of the post-level premium period, the composition of the policyholders will be heavily weighted toward unhealthy individuals who need the life insurance policy because they cannot get coverage elsewhere or the premiums would be even higher than retaining the current policy. There may also be a limited number of policyholders who do not have the desire to shop for other coverage.
- Our model also predicts that lapse rates will be more sensitive to premium increases for lower premium increase buckets. This is illustrated by the fact that the slope of the predicted lapse rates as a function of premium jump is steeper for lower cumulative premium increases (below 4x) than for higher cumulative premium increases (above 4x). This suggests that although policyholders are more likely to lapse under extreme premium increases, they are more sensitive to premium increases in more moderate ranges (i.e., 1x to 4x).

FIGURE 10: PREDICTIONS FOR DIFFERENT DURATIONS IN THE POST-LEVEL PREMIUM PERIOD

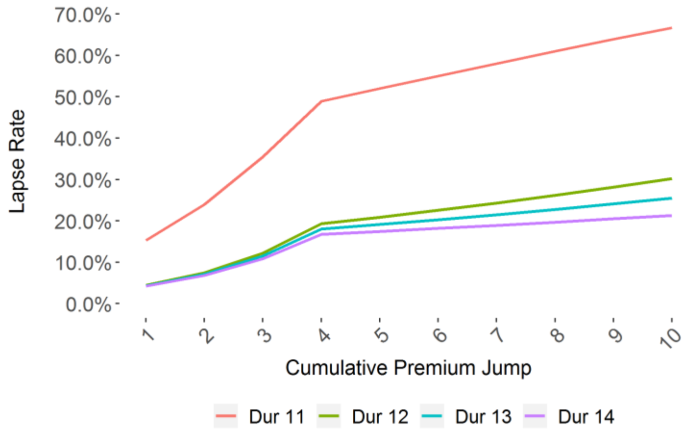


FIGURE 11: PREDICTED LAPSE RATES FOR AGE 65 MALE NONSMOKER (BEST PREFERRED) WITH $500,000 FACE AMOUNT FOR A 10-YEAR LEVEL PREMIUM TERM PRODUCT

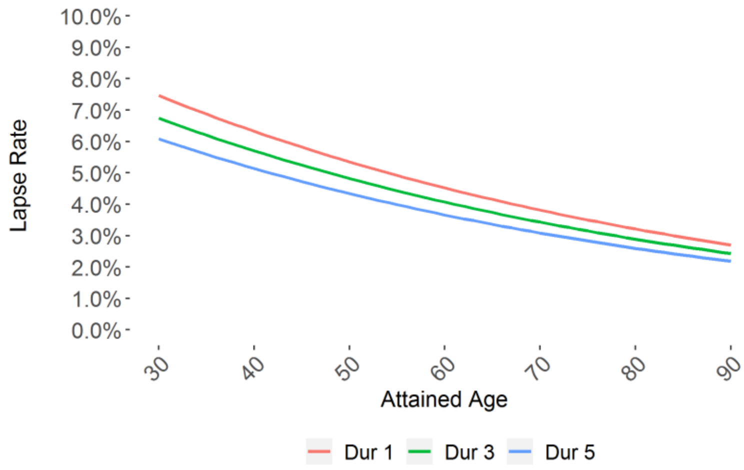
CUMULATIVE PREMIUM JUMP	DUR 11 PREDICTED LAPSE RATE	DUR 12 PREDICTED LAPSE RATE	DUR 13 PREDICTED LAPSE RATE	DUR 14 PREDICTED LAPSE RATE
1	15.32%	4.53%	4.36%	4.20%
2	23.97%	7.53%	7.15%	6.79%
3	35.45%	12.25%	11.51%	10.81%
4	48.89%	19.33%	18.01%	16.77%
5	51.97%	20.91%	19.13%	17.47%
6	55.02%	22.58%	20.30%	18.19%
7	58.05%	24.35%	21.52%	18.93%
8	61.01%	26.20%	22.79%	19.70%
9	63.89%	28.15%	24.11%	20.49%
10	66.68%	30.18%	25.49%	21.30%

RESULTS BY ATTAINED AGE

During the level premium period, lapse rates are generally lower for older attained ages. However, during the last year of the level premium period and first year of the post-level premium period, lapse rates are *higher* for older attained ages.

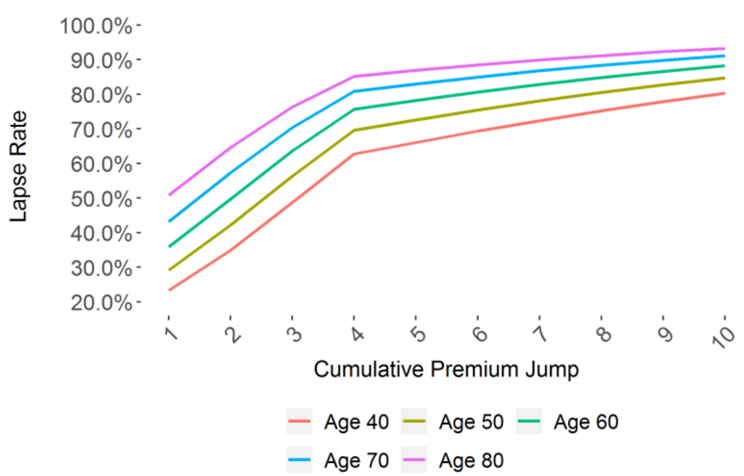
Figure 12 illustrates our model's trend of predicting lower lapse rates for older attained ages for select durations inside the level premium period for the sample policyholder. Older policyholders are more likely to retain their policies within the level premium period because these policyholders are more likely to be less healthy and it will be more expensive for them to get a new policy with another insurance company.

FIGURE 12: PREDICTIONS FOR DIFFERENT AGES IN THE LEVEL PREMIUM PERIOD



When the level premium period is about to end, there is a shift in the relationship between lapse rates and attained age. During the last year of the level premium period (duration 10 for the 10-year level premium term, and duration 15 for the 15-year level premium term), lapse rates increase as a function of attained age. This relationship is also present during the first year of the post-level premium period, although the relationship is not as strong. Figure 13 illustrates this from our lapse model for the last duration of the level premium period. This plot shows that, even after controlling for premium increase and face amount differences, older policyholders have higher lapse rates in this duration.

FIGURE 13: PREDICTIONS FOR LAST DURATION OF LEVEL PREMIUM PERIOD



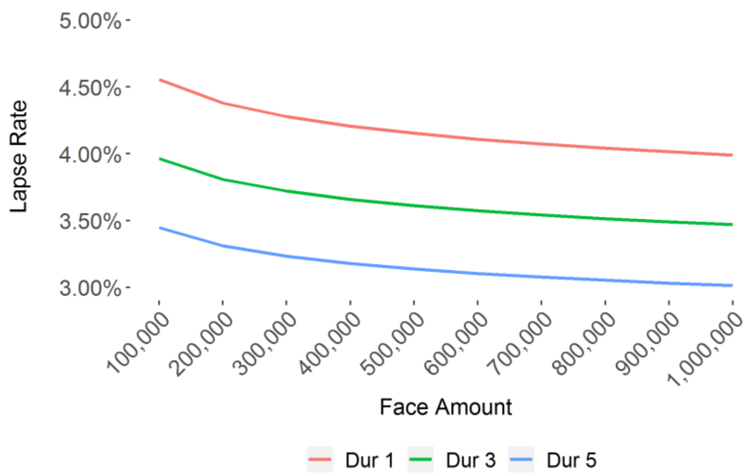
One explanation for why lapse rates increase for older policyholders at the end of the level premium period is that older policyholders have higher premiums. Thus, they likely experience higher premium increases in absolute terms when the level premium period ends, leading to higher lapse rates than for the younger ages.

RESULTS BY FACE AMOUNT

Lapse rates are predicted to decrease with increasing face amount during the first T-1 durations in the level premium period for a T-year level premium term product. However, this trend reverses during the last duration of the level premium period and in the post-level premium period, as lapse rates are predicted to *increase* as a function of face amount.

Figure 14 plots our model's predicted lapse rates as a function of face amount for our sample policyholder inside the level-premium period. In this chart, we see that lapse rates slightly decrease for higher face amounts for the different policy durations displayed.

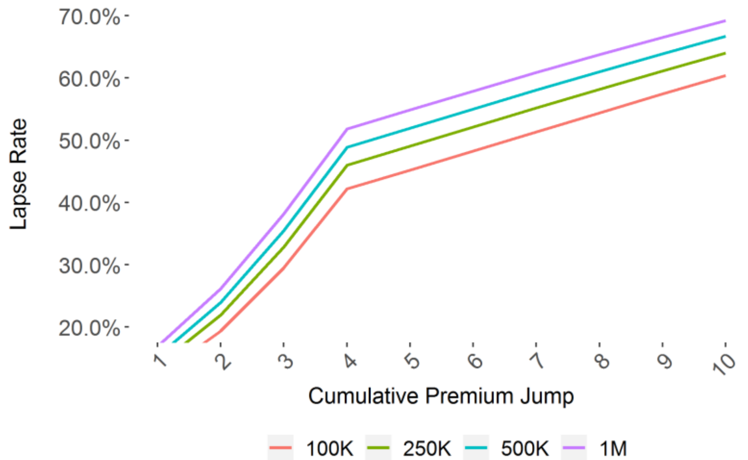
FIGURE 14: PREDICTIONS FOR DIFFERENT FACE AMOUNTS IN LEVEL PREMIUM PERIOD



However, our model's predicted relationship between lapse rates and face amount is drastically different for durations that cover the last year of the level-premium period and the post-level premium period. For these policy durations, our model predicts that higher face amounts result in higher lapse rates. It appears that policyholders are sensitive to the *dollar increase* in the premium payments. For a given premium increase (i.e., 200% premium increase), policyholders with higher face amounts will experience higher dollar amounts of premium increase than policyholders with lower face amounts. Thus, this positive relationship between lapse rates and face amount can be interpreted as a way of quantifying a policyholder's propensity to lapse from having to make a higher dollar amount of premium payments.

Figure 15 plots our model's predicted lapse rates for different face amounts for our sample policyholder in the first duration of the post-level premium period. This chart illustrates that, for a given a cumulative premium increase, our model will predict higher lapse rates for policyholders with higher face amounts.

FIGURE 15: PREDICTIONS FOR DIFFERENT FACE AMOUNTS IN FIRST DURATION OF POST-LEVEL PREMIUM PERIOD



RESULTS BY PREMIUM COLLECTION TYPE

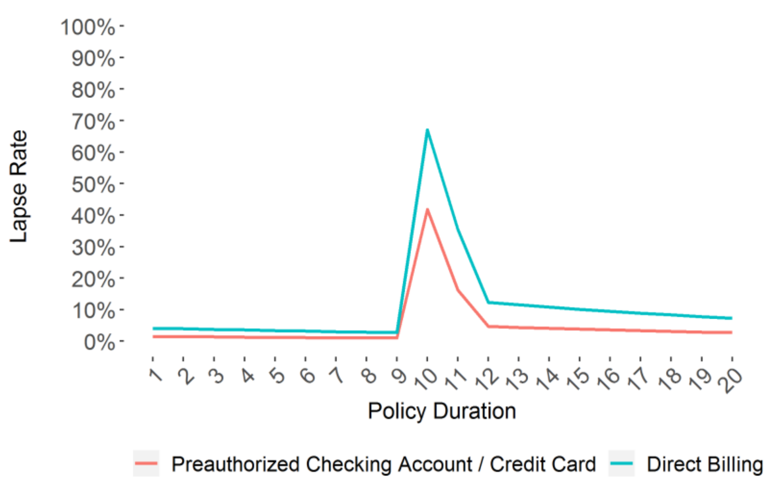
Our model predicts that lapse rates for premium collection types using an automatic preauthorized checking account or credit card will be lower than lapse rates for policies that pay using direct billing.

This model result is consistent across all durations of the policy. Lower lapse rates for credit card or preauthorized checking account methods is intuitive, because these payment types are automatic and require the policyholder to take an action to stop the payments and cancel the policy. In contrast, direct billing requires the policyholder to make an action each time a payment is made, so these policyholders are likely more aware of their premium payments and would have more opportunities to cancel an insurance policy.

Also, credit card and preauthorized checking account payments are typically monthly while direct bill payments are usually annual. This also leads to lower premium dollar jumps for the former and subsequently lower lapse rates.

Figure 16 plots predicted lapse rates as a function of policy duration for the two premium collection types for our sample policyholder. For most durations of this sample policy, the predicted lapse rate under direct billing payment can be about 2x higher than the lapse rate under credit card or preauthorized checking account arrangements.

FIGURE 16: PREDICTIONS FOR DIFFERENT PREMIUM COLLECTION TYPES

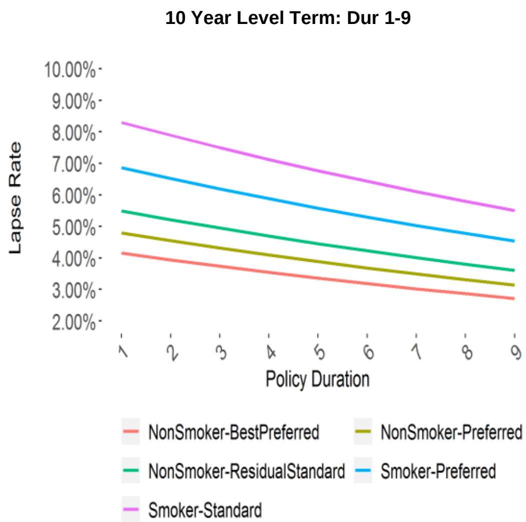


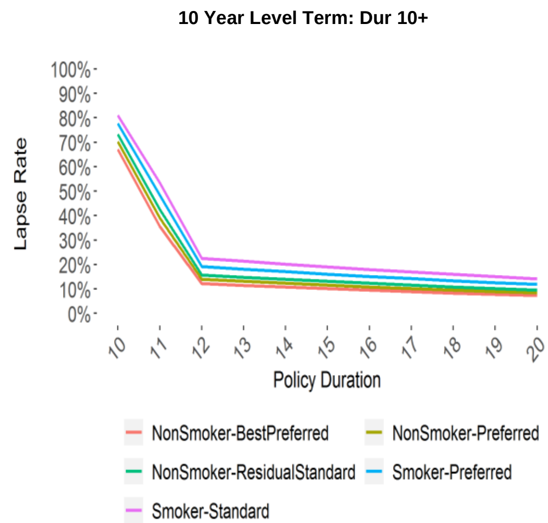
RESULTS BY RISK CLASS GROUP

Lapse rates are projected to be higher for the risk class groups with less healthy individuals.

In particular, the five risk class groups in our data set with a hierarchal order for riskiness are, in increasing order of riskiness: 1) Best Preferred Nonsmoker, 2) Preferred Nonsmoker, 3) Residual Standard Nonsmoker, 4) Preferred Smoker, and 5) Standard Smoker. Figure 17 plots the predicted lapse rates in our model for these different risk class groups across the different policy durations for our sample policy. These charts illustrate that the predicted lapse rate increases for groups with higher risk (i.e., smokers) in both the level-premium period and post-level premium period, while holding other policyholder features constant (i.e., premium increase, sex, attained age).

FIGURE 17: PREDICTIONS FOR DIFFERENT RISK CLASS GROUPS





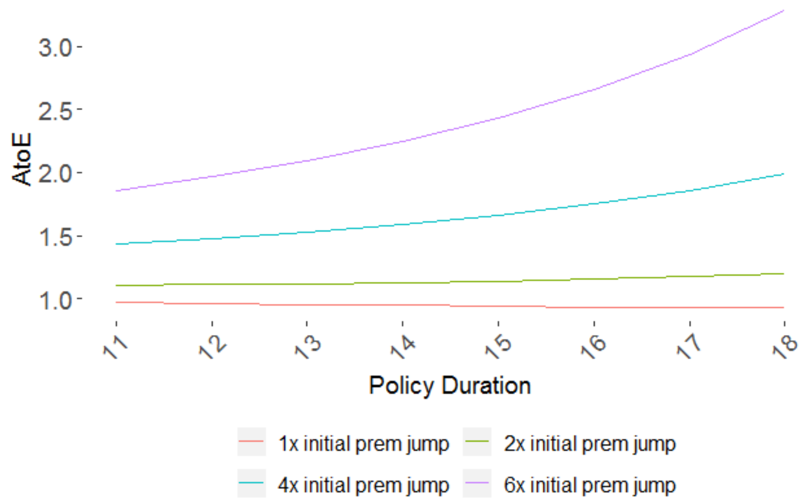
5. Mortality model results

Mortality is progressively worse over time for the group of policyholders that experience higher cumulative premium increases during the post-level premium period. This is because higher premiums cause healthier lives to lapse, leaving a less healthy pool of policyholders remaining.

Figure 18 illustrates this by comparing our model's predicted mortality rates relative to the 2015 VBT mortality table for different levels of premium increases in the post-level premium period.

- Figure 18 looks at mortality rates for our sample policyholder. In particular, we focus on the cases when the initial premium increase for the first year of the post-level premium period is flat (i.e., 1x), doubles (2x), increases by four times (4x), and increases by six times (6x). For each initial premium jump, we assume premiums increase 10% annually thereafter.
- Figure 18 shows that, for low cumulative premium increases, the model's predicted mortality in the post-level premium periods is similar to the mortality from the 2015 VBT table. However, our model's predicted mortality can be significantly higher than the 2015 VBT for higher premium increases.
- For instance, if our sample policy faces a 4x premium increase in the first year of the post-level premium period, our model predicts that mortality will be almost 1.5 times worse than mortality implied by the 2015 VBT table.

FIGURE 18: MODEL VS. 2015 VBT MORTALITY RATES FOR SAMPLE POLICYHOLDER IN POST-LEVEL PREMIUM PERIOD



Policyholders who do not lapse the insurance policy after experiencing extreme premium increases are more likely to need it due to potential health conditions. Thus, the group of policyholders in the post-level premium period will likely have higher mortality rates.

6. Conclusion and future analysis

Lapse and mortality experience for post-level premium term products and other types of insurance products continue to have an important impact on in-force management and new product design. With a robust data set and the right modeling techniques, we can analyze nuanced relationships and understand complex interactions. Our modeling on post-level premium term products demonstrated that post-level premium term lapse behavior can be predicted with a high degree of confidence using variables such as duration, cumulative premium increase, attained age, face amount, premium collection type, etc. Furthermore, mortality experience is also similarly influenced by the lapse behavior resulting from premium increase.

As we obtain more data going forward in the MIMSA III data set, there will be additional analysis we can perform. One potential area for future analysis is comparing the impact of using year-over-year premium increases instead of cumulative premium increases on lapse and mortality rates in the post-level premium period. For this analysis, we used cumulative premium increase as a predictor because most of the policyholders in our data set experienced the largest premium increase in the first duration of the post-level premium period. However, it is possible that, for a later duration in the post-level premium period, year-by-year premium jumps may differ, even if the cumulative premium jump is the same. This will be particularly relevant as more recently issued level premium policies with more gradual premium increases enter their post-level premium periods. We hope to continue to track emerging experience over time to see how these models perform in the future and share our insights with our readers. Again, we thank all the MIMSA III participating companies for contributing valuable data, which made this study possible.

Appendix

PREDICTIVE MODEL FITTING: GENERALIZED LINEAR MODEL

For our lapse and mortality models, we use logistic regression to model the behavior. Logistic regression is a type of generalized linear model, which is a class of linear regression models that employ transformations of the dependent variable through link functions. Linear regression is a very popular predictive modeling approach—the response variable is modeled as a linear combination of a finite set of predictor variables. Model simplicity provides results that are easy to interpret and with relatively low computational requirements for implementation. However, there are drawbacks to using linear regression. One such shortcoming is that the presence of multicollinearity between predictors will lead to unreliable estimates of regression coefficients. Overcoming this pitfall requires additional analysis to ensure that covariates are not highly correlated with each other.

Using linear regression is only appropriate if the response variable has an approximately linear relationship with the selected predictors. In ordinary linear regression, the algorithm maximizes the sample likelihood by assuming that the error of the response variable is normally distributed. When this assumption of normality is invalid, transforming the response variable may lead to an error distribution assumption that is more suitably modeled as a linear function of the selected covariates. Generalized linear models (GLMs) are a class of linear regression models that employ such transformations through the *link* function.

Each link function is unique to a specific error distribution and corresponds to different types of model output. For example, if the response variable is binomially distributed then it can be modeled with logistic regression by transforming the output using the logit link function. A logistic regression model predicts log-odds of an event, which in turn yields probabilities for the event of interest by using the inverse of the logit link: the logistic function.

To be more precise with the notation, suppose we have *N* observations in our data set, and build a logistic regression model with p predictors.

- We will let x_i represent a (p + 1)-dimensional feature vector for the i^{th} observation in the data set, where the first element of x_i will just be the bias term 1.
- We will let β=($\beta_0, \beta_1, \beta_2, \ldots., \beta_p$) be a (p + 1)-dimensional vector that has the model coefficients corresponding with each component of x_i. Note that β_0 is the y-intercept of the model.
- We let y_i be the binary response variable for the i^{th} observation, where y_i equals 1 if the policyholder lapses (dies) in the lapse (mortality) model, and equals 0 otherwise. We will let π represent the probability that y_i equals 1.

Under this notation, the form of the logit link function being modeled in our regression, as well as the probability π is given below:

Logit link: $\ln\left(\frac{\pi}{1-\pi}\right) = {x_i}^T\beta = \sum_{j=0}^{p} x_{ij}\beta_j$

Logistic function: $\pi = \frac{e^{{x_i}^T\beta}}{1+e^{{x_i}^T\beta}} = \frac{e^{\sum_{j=0}^{p} x_{ij}\beta_j}}{1+e^{\sum_{j=0}^{p} x_{ij}\beta_j}}$

Under a generalized linear model using the logit link, one can show that, given the set of coefficients β=($\beta_0, \beta_1, \beta_2, \ldots., \beta_p$), the log-likelihood function of the data is given by:

Logistic log-likelihood function: $l(\beta) = \frac{1}{N}\sum_{i=1}^{N} y_i({x_i}^T\beta) - \log\left(1+e^{{x_i}^T\beta}\right)$

For this analysis, we fit a logistic regression to solve for the set of model coefficients, β, which maximizes the log-likelihood function of the data given by $l(\beta)$ above.

LAPSE MODEL: MODEL FORM AND COEFFICIENTS

We fit the following main effect variables in our lapse model

1. Attained age
2. Sex
3. Premium collection type
   - This is a categorical variable with one of three values: 1) automatic credit card or preauthorized checking account, 2) direct billing, or 3) unknown.
4. Risk class group
5. Log face amount
   - This is a continuous variable that is equal to the natural log of the face amount of the policy.
6. Indicator variable for whether or not premium increase information during the post-level premium period was available for a policyholder. A value of 1 means that no premium information was available for the policyholder, while a value of 0 means premium information was available.
7. Cumulative premium increase
8. Cumulative premium jump capped
   - This is a variable defined as: max(Cumulative premium increase – 4, 0). By including this variable as a predictor, our model was able to fit a piecewise function to lapse rates as a function of premium increase (one for premium jumps between 1x and 4x, and another for premium jumps greater than 4x)
9. Term phase
   - This is a categorical variable that classifies the durational period that a policyholder is in. For a level premium term product with a level-premium period of T years, there will be four different term phases:
     1. Durations <= T - 1
     2. Duration T
     3. Duration T + 1
     4. Durations >= T + 2
10. Term length
    - A categorical variable that specifies whether or not a policyholder owns a 10-year or 15-year level premium term product.
11. The percentage of time remaining in the level-premium period of a policy
    - For example, if a policyholder of a 10-year level premium term product is in duration 8, then the percentage of time remaining in the level-premium period for this observation is 20%. For this same observation, in duration 9, the percentage of time remaining will decrease to 10%.
12. Policy duration in the post-level premium period
    - For example, if a policyholder of a 10-year level premium term product is in duration 15, the value of this variable in this observation is 5.

In addition to the main effect variables described above, we also modeled the following interaction terms:

13. Interaction between the cumulative premium increase and the number of years in the post-term period
14. Interaction between the cumulative premium jump capped and the number of years in the post-term period
15. Interaction between the attained age and the term phase
16. Interaction between the log face amount and the term phase
17. Interaction between the cumulative premium increase null indicator and term length
18. Interaction between the term length and term phase

We present two Lapse Model Coefficients Tables below showing the coefficients of the lapse model. The first table displays the coefficients of the main effects of the predictors in the lapse model. The second table displays the coefficients of the interaction terms in the lapse model. We have displayed the coefficients of every category for categorical variables in the main effects table, but excluded the baseline category in the interaction coefficients table for brevity.

LAPSE MODEL COEFFICIENTS TABLE 1: MAIN EFFECTS

VARIABLE	COEFFICIENT VALUE
Attained Age	3.1%
Sex (Female)	0.0%
Sex (Male)	3.3%
Sex (Unknown)	-425.4%
Premium Collection (Credit Card/Checking Account)	0.0%
Premium Collection (Direct Bill)	104.2%
Premium Collection (Unknown)	110.5%
Risk Class (Nonsmoker - Best Preferred)	0.0%
Risk Class (Nonsmoker - Preferred)	15.0%
Risk Class (Nonsmoker - Residual Standard)	29.4%
Risk Class (Smoker - Preferred)	53.1%
Risk Class (Smoker - Standard)	73.6%
Risk Class (Unknown/Other)	35.2%
Log Face Amount	10.8%
Cumulative Premium Increase Null Indicator	83.3%
Cumulative Premium Increase	57.1%
Cumulative Premium Increase - Jump Capped	-42.3%
Term Phase (Duration T)	0.0%
Term Phase (Duration T+1)	-17.7%
Term Phase (Durations <= T-1)	214.8%
Term Phase (Durations >= T+2)	-18.2%
Term Length (10 Yr Level Term)	0.0%
Term Length (15 Yr Level Term)	-3.1%
Percentage of time remaining in level premium period	54.9%
Policy duration in the post-level premium period	-2.4%

LAPSE MODEL COEFFICIENTS TABLE 2: INTERACTION EFFECTS

INTERACTION	COEFFICIENT VALUE
Cumulative Premium Increase : Policy duration in the post-level premium period	-1.6%
Cumulative Premium Increase - Jump Capped: Policy duration in the post-level premium period	-0.9%
Attained Age : Term Phase (Duration T+1)	-2.8%
Attained Age : Term Phase (Duration <= T-1)	-4.8%
Attained Age : Term Phase (Duration >= T+2)	-3.4%
Log Face Amount : Term Phase (Duration T+1)	6.0%
Log Face Amount : Term Phase (Duration <= T-1)	-16.9%
Log Face Amount : Term Phase (Duration >= T+2)	-1.1%
Cumulative Premium Increase Null Indicator : Term Length (15 Yr Level Term)	28.2%
Term Length (15 Yr Level Term) : Term Phase (Duration T+1)	10.9%
Term Length (15 Yr Level Term) : Term Phase (Duration <= T-1)	-37.7%
Term Length (15 Yr Level Term) : Term Phase (Duration >= T+2)	-14.0%

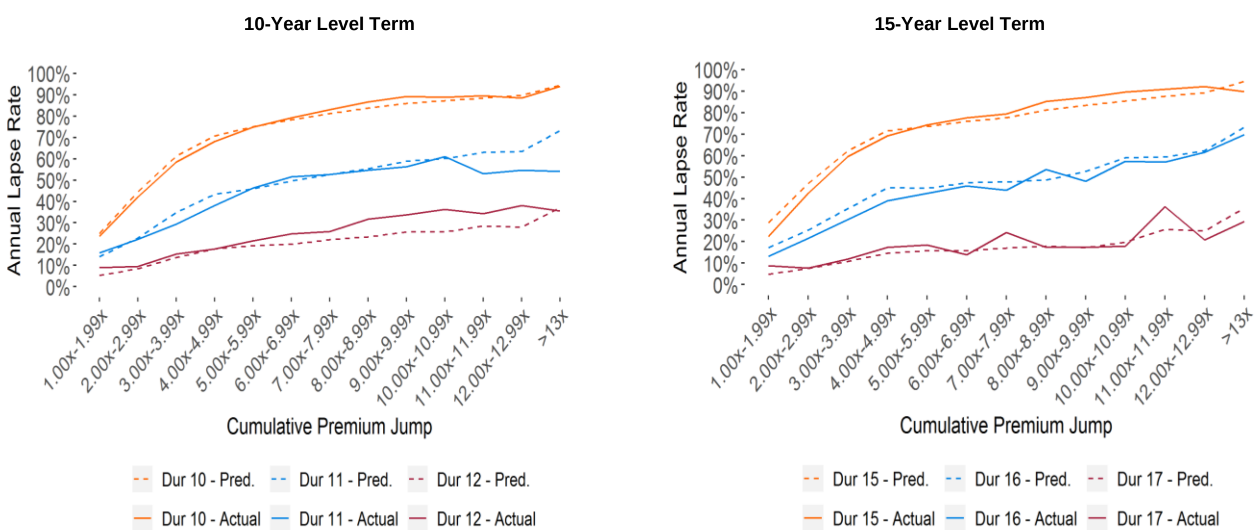
ADDITIONAL LAPSE MODEL VALIDATION CHARTS

The charts below provide additional model validation plots along other dimensions of the holdout data set for the lapse model.

Lapse model: Validation by cumulative premium increase

Figure 19 plots the predicted and actual lapse rates on our holdout data set by different cumulative premium jump buckets for the last duration of the level premium period and the first two durations of the post-level premium period. The predicted lapse rates line up fairly closely with the actual lapse rates for both the 10-year and 15-year level premium term products, although the actual data has more noise due to lower exposure for certain premium jump buckets.

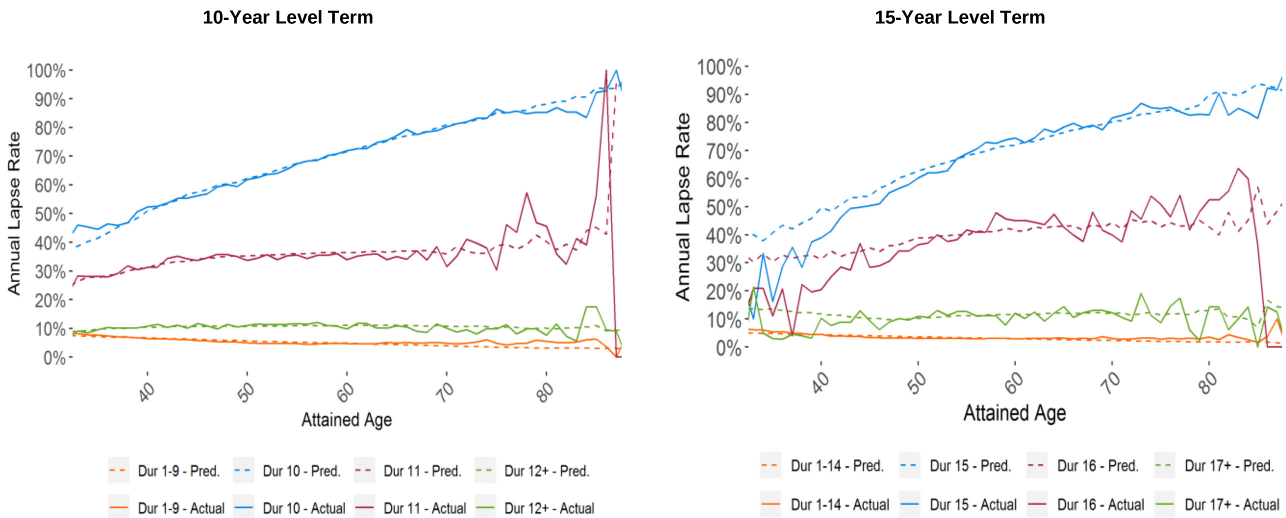
FIGURE 19: RESULTS BY CUMULATIVE PREMIUM JUMP FOR DIFFERENT DURATIONS



Lapse model: Validation by attained age

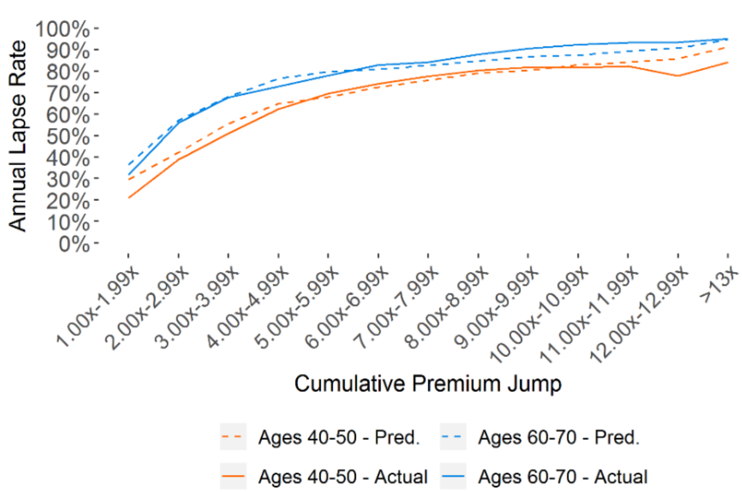
Figure 20 plots predicted lapse rates on the holdout data set as a function of attained age for the four different term phases that are considered in the lapse model. Overall, the predicted lapse rates from the model follow closely with the actual lapse rates across all term phases. However, the data is more noisy for the 15-year level premium term for younger and older attained ages where there is less exposure. The orange lines in the charts emphasize again that, for a level premium term product with duration T, lapse rates actually decrease as a function of age over the first T-1 durations. However, the blue and red lines illustrate that, during the last duration of the level premium period and the first duration of the post-level premium period, lapse rates actually increase for older policyholders in our holdout data set.

FIGURE 20: RESULTS BY ATTAINED AGE FOR DIFFERENT TERM PHASES



In the MIMSA data set, part of the positive relationship between lapse rates and age can be explained by the fact that older policyholders tend to receive higher premium increases. However, this relationship still persists even after controlling for cumulative premium increase. Figure 21 compares actual and predicted lapse rates as a function of cumulative premium increase for the 40-50 and 60-70 age groups during the last duration of the level premium period. Regardless of the size of the initial premium increase, policyholders from the 60-70 age group had higher lapse rates than those from the 40-50 age group, as would be expected because of the higher dollar premiums, as described above.

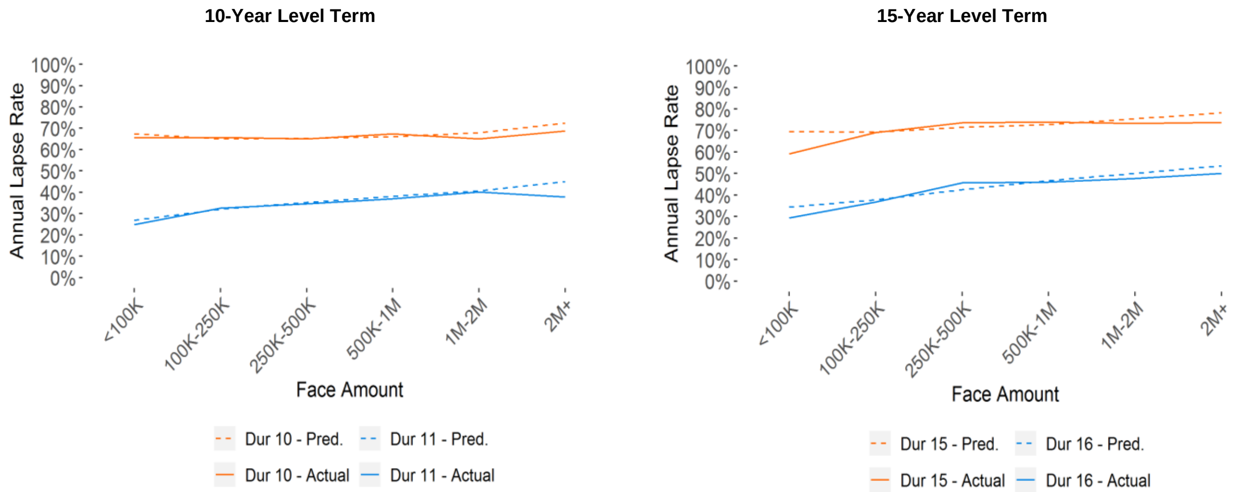
FIGURE 21: RESULTS BY CUMULATIVE PREMIUM JUMP FOR DIFFERENT ATTAINED AGES, LAST DURATION OF LEVEL PREMIUM PERIOD



Lapse model: Validation by face amount

Figure 22 plots actual and predicted lapse rates on the holdout data set as a function of different face amount buckets for the last duration of the level-premium period and the first duration of the post-level premium period. Although the model fit is not perfect, the predicted trends are consistent with the actual lapse rates from the holdout data set.

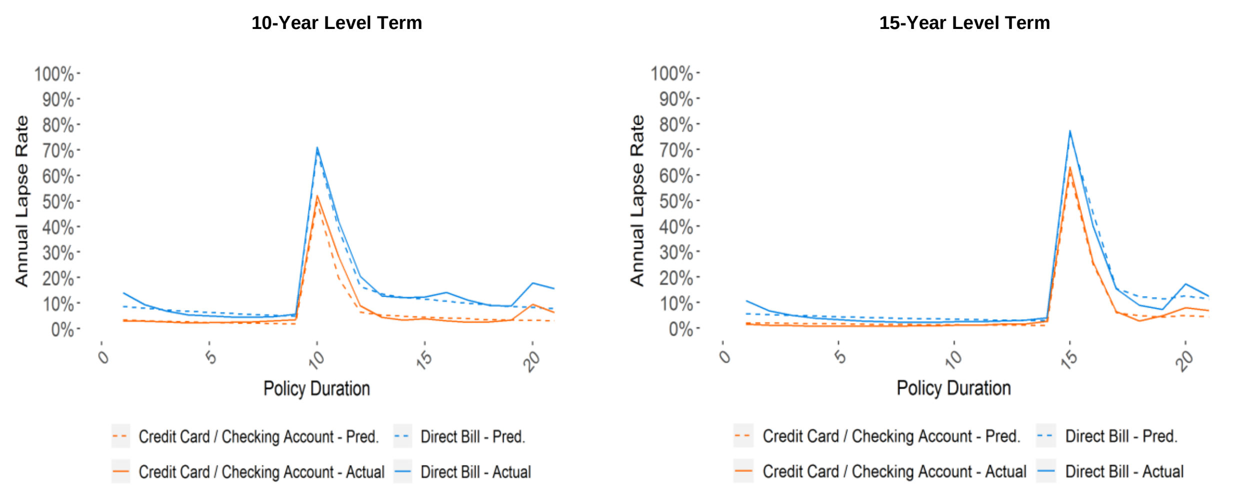
FIGURE 22: RESULTS BY FACE AMOUNT FOR DIFFERENT TERM PHASES



Lapse model: Validation by premium collection type

Figure 23 plots the predicted and actual lapse rates by duration in the holdout data set for the two different premium collection types. This plot illustrates that the gap between lapse rates for direct billing and credit card or checking account premium collection types is present in our data, and that our model predictions are fairly close to the actual lapse rates from our data.

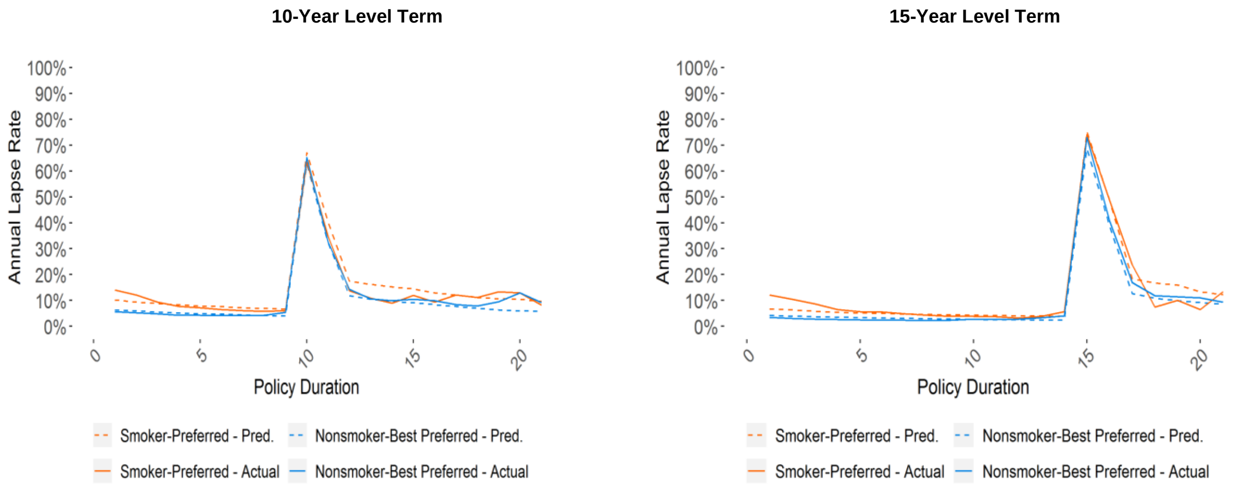
FIGURE 23: RESULTS BY POLICY DURATION FOR DIFFERENT PREMIUM COLLECTION TYPES



Lapse model: Validation by risk group

Figure 24 plots predicted and actual lapse rates on the holdout data set across duration for Standard Smokers and Preferred Nonsmokers. We can see that, for both the 10-year and 15-year level premium term products, our model predictions and the data show that the Standard Smoker risk class group has higher lapse rates during the level-premium period. However, note that the model fit to the holdout data during the post-level premium period for smokers is not as accurate, with our model overestimating lapse rates across most of the post-level premium durations. Nevertheless, because there was low exposure for smokers during the post-level premium period, we decided not to potentially overfit our model and add additional interaction terms that could capture different lapse rate behavior for smokers during the post-level premium period.

FIGURE 24: RESULTS BY POLICY DURATION FOR DIFFERENT RISK CLASS GROUPS



The table in Figure 25 summarizes the actual and predicted lapse rates (averaged across all durations) for the different risk class groups. The actual-to-expected (A/E) ratios are very close to 100%, and we also can see the trend of higher average lapse rates for riskier risk groups in the data.

FIGURE 25: HOLDOUT DATA SET, MODEL VALIDATION METRICS FOR RISK CLASS GROUPS

	ACTUAL LAPSE RATE	PREDICTED LAPSE RATE
Nonsmoker - Best Preferred	8.81%	8.82%
Nonsmoker - Preferred	8.22%	8.21%
Nonsmoker - Residual Standard	9.43%	9.45%
Smoker - Preferred	12.20%	12.20%
Smoker - Standard	12.90%	12.70%

MORTALITY MODEL: MODEL FORM AND COEFFICIENTS

The table in Figure 26 displays the fitted coefficients of the mortality model.

FIGURE 26: MORTALITY MODEL COEFFICIENTS TABLE, - MAIN EFFECTS

VARIABLE	COEFFICIENT VALUE
Risk Class (Nonsmoker - Best Preferred)	0.0%
Risk Class (Nonsmoker - Preferred)	17.6%
Risk Class (Nonsmoker - Residual Standard)	45.6%
Risk Class (Smoker - Preferred)	11.5%
Risk Class (Smoker - Standard)	38.1%
Risk Class (Unknown/Other)	46.3%
Log Face Amount	-6.5%
Cumulative Premium Increase Null Indicator	5.0%
Cumulative Premium Increase	13.0%
Term Phase (Duration T+1)	0.0%
Term Phase (Durations <= T)	-26.6%
Term Phase (Durations >= T+2)	0.7%
Term Length (10 Yr Level Term)	0.0%
Term Length (15 Yr Level Term)	0.7%
Percentage of time remaining in level premium period	5.7%
Policy duration in the post-level premium period	-2.4%

MORTALITY MODEL VALIDATION CHARTS

Figure 27 plots predicted and actual mortality rates on the holdout data set across duration for the 10-year level premium term product. This chart shows that our model predictions are relatively consistent with observed mortality experience.

FIGURE 27: 10-YEAR LEVEL TERM, RESULTS BY POLICY DURATION

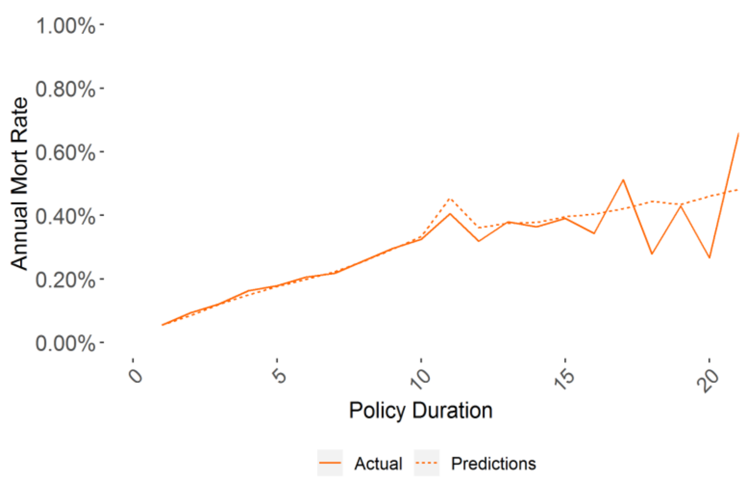
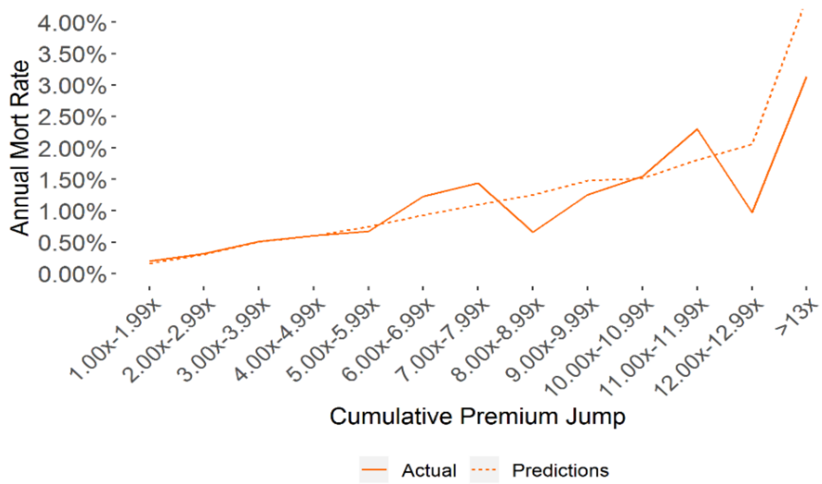


Figure 28 plots predicted and actual mortality rates on the holdout data set across cumulative premium increase for all durations in the post-level premium period. The fit is good for lower premium increases, but is not as good for higher premium increases where there is less exposure and more noise in the data set.

FIGURE 28: 10-YEAR LEVEL TERM, RESULTS BY CUMULATIVE PREMIUM JUMP IN POST-LEVEL PREMIUM PERIOD



CONTACT

Ken Qian
ken.qian@milliman.com

Ben Johnson
ben.johnson@milliman.com

Jenny Jin
jenny.jin@milliman.com