IMPROVING ACTUARIAL RESERVE ANALYSIS THROUGH CLAIM-LEVEL PREDICTIVE ANALYTICS

presented by:
Philip S. Borba, Ph.D.
Lori Julga, FCAS, MAAA
Milliman, Inc.

September 15, 2014

Casualty Actuarial Society, Casualty Loss Reserve Seminar, San Diego, CA

Milliman

Casualty Actuarial Society -- Antitrust Notice

The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Overview

- Business Objectives and Predictive Analytics

- Predictive Analytics
  - Starting Considerations
  - High Severity Claims
  - Jumper Claims
  - Business Stratifications

- Implications for Reserving
  - Assumptions in predictive modeling
  - Impact (actuarial implications) of initiatives

Business Objectives and Predictive Analytics

- Business objectives
  - Early recognition of potential high-loss claims to direct specialized adjuster resources
  - Early recognition of potential high-loss claims to control claim costs
- Using predictive analytics for early identification
  - High Severity Claims (claim cohorts with high/low claim costs)
  - Jumper Claims (early identification of claims that have high potential to become the most costly claims)

Business Opportunities

- High Severity Claims
  - Claim characteristics are used to define claim cohorts with high/low claim costs
  - Small number of characteristics often capture subset of high-cost claims
  - Analytics often uncover segments with previously unnoticed characteristics
- Jumper Claims
  - Early identification of claims that have high potential to become the most costly claims
  - Claims have modest paid or incurred amounts at 30 days (for example), but high ultimate costs
  - Large number of characteristics available for the Jumper Claim Score

Predictive Analytics – Starting Considerations

Milliman

Predictive Analytics

- Data for the analytics
- Evaluation periods
- Segmentation Analysis
- Claim Score Analysis
- Business stratifications and expected results

Milliman

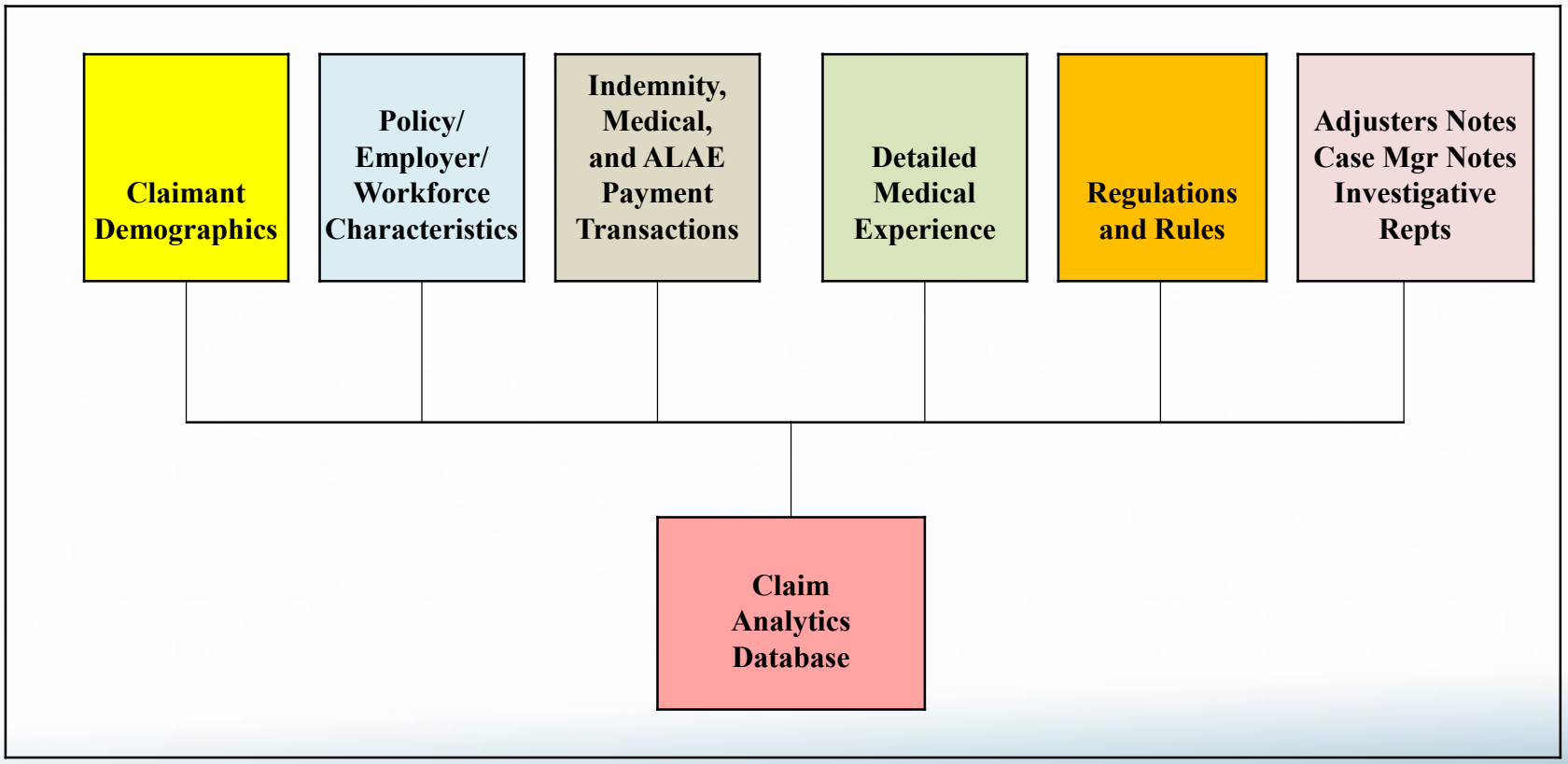
Data and Statistical Techniques

- Typical approach is to use multivariate analyses (eg, multivariate logit, multiple least squares, GLM) that permit identifying relationships between an outcome measure and a set of explanatory variables.
  - These analyses permit identifying and measuring which explanatory variables have the most significant and largest impacts on an outcome measure.

- Recent work with "machine learning" software that extends the principles of multivariate analyses.
  - The new statistical tools enable testing a much larger number of model specifications in a much shorter turnaround time than the conventional multivariate analyses.
  - The tools can include claims where data are not available for all characteristics and characteristics where data are not available for all claims (the pernicious "incomplete" and "missing" data problems).
  - The analytical tools use machine learning, a type of artificial intelligence, to analyze hundreds of characteristics and correlations.

Data for the Analytics – A Multitude of Data Sources

- A multitude of data sources are available.



Milliman

Data for the Analytics

- Claim characteristics that can be included in the analysis:
  - Demographic (age, gender, marital status)
  - Injury (reporting lag, body part, nature of injury)
  - Employment (industry, size of employer, geographic location)
  - Presence of an attorney

- Payment history (assembled from transactions data):
  - Payments by type of benefit (indemnity, medical)
  - Payments by subtype (eg, payments to MD/DO, hospital)

- Medical experience (developed from detailed medical data):
  - Number of visits for office visits, physical therapy, chiropractic services, radiology tests)
  - Surgery, Pharmaceuticals, Durable medical products

Evaluation Periods

- Claim costs as of 30, 60, 180, 360 or some other number of days from injury date
  - Claim demographics
  - Indemnity benefits and medical expenses
  - Detailed (line-item) medical experience can provide the number of services from days from injury (for example, the number of physical therapy services during first 30 days from injury)

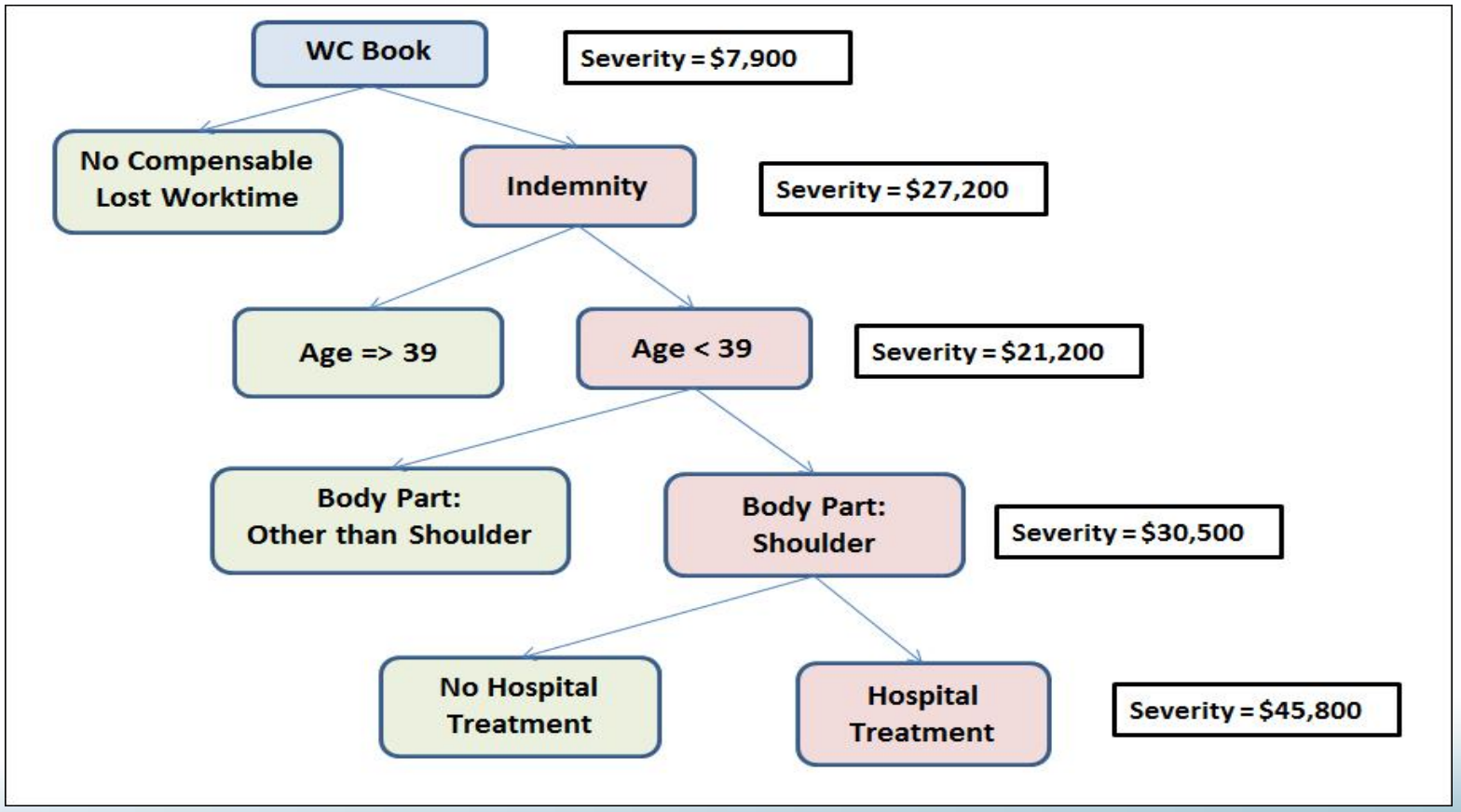
- Analysis captures comparable experience on claims
- Information may be limited due to delays in payment processing
- Text data can be used to supplement transaction data

Predictive Analytics - Segmentation Analysis

Milliman

Segmentation Analysis

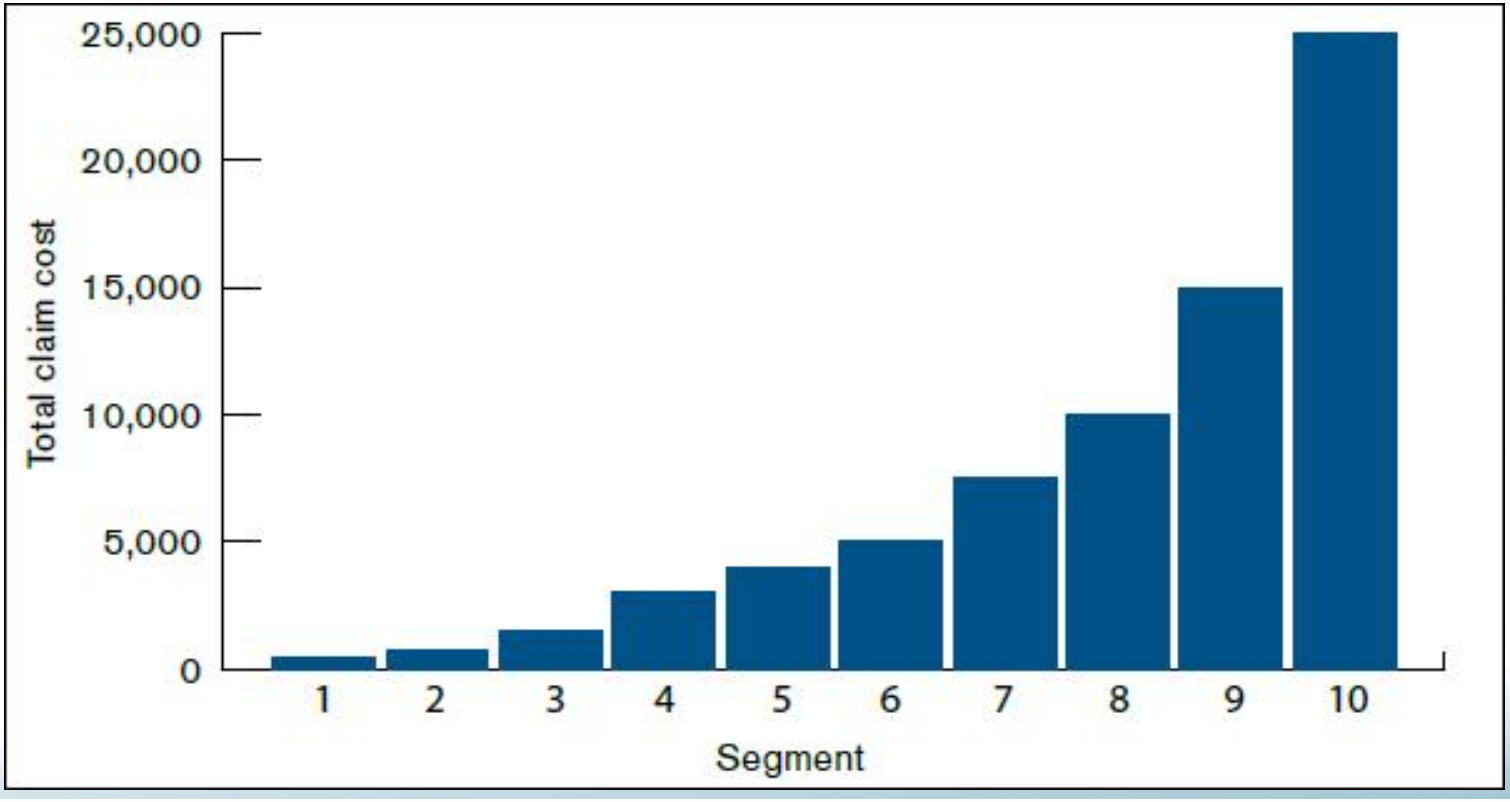
- Segmentation – decision tree method that produces discrete easily understandable segments, but with less lift and precision than Scoring



Milliman

Segmentation Analysis

- Segmenting claims according to total claim cost
- Claim characteristics and medical experience can be used to segment claims into groups with similar total claim costs



Milliman

Segments – Using Claimant, Payment, Medical, et. al., Data

- Claim segments can be defined by the claimant characteristics, payment history, and medical experience
  - Opportunity to include detailed medical experience in claim analytics
  - Opportunity to view claim costs from different timing perspectives

- Claims could be arranged into any manageable number of segments (eg, 4-12)

- Segments may not use the same factors or the same number of factors
  - Worker age or industry group may/may not be a consideration for each segment
  - Number of office visits or physical therapy treatments may/may not be a consideration
  - High-cost segments may include surgery

- Information does not need to be complete for all factors for a claim to be included in the analyses and assigned to a segment

- Information does not need to be complete for all claims for a factor to be included in the analysis and such as a segmenting factor

Milliman

Segments – Factors in an Illustrative Example

- 10 segments
- Not every factor used in each segment
- Claims assigned to segment best fitting segment definition
- Illustration uses claimant characteristics, payment history, and detailed medical experience

	Segment									
Factor	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Body Part	not multiple not back	not multiple			back knee shoulder	back knee shoulder			multiple	multiple
Age	under 40	40+			------	------			------	
Medical	<= 3 med visits	> 3 med visits			13-24 phys ther visits no surgery	> 24 phys ther visits no surgery			> 12 med visits no surgery opiods	> 12 med visits surgery
Industry	not mfg not construct	mfg construction			------	------			------	------
Disability Status	med only	med only			temporary	temporary			permanent	permanent
Region	------	------			------	------			high urban	------
Claim Reporting	------	------			------	------			> 2 wks after injury	------
Claimant attorney	------	------			------	------			------	Yes

Milliman

Variations – Outcome Measures and Claim Groups

- Outcome measures:
  - Starting point: total claim costs
  - Alternatives measures
    - Paid indemnity benefits
    - Paid medical expenses
    - Incurred indemnity benefits
    - Incurred medical expenses

- Claim groups
  - Starting point: all claims
  - Stratifications
    - Lost-time claims: removes large number of low-cost claims
    - Claims limited to a particular type of injury (eg, low back injuries)
    - Claims limited to a particular industry (eg, manufacturing)
    - Claims limited to a particular demographic (eg, workers 40 years and over)

Predictive Analytics - Claim Score Analysis

Milliman

Developing Claim Scores

- Outcome measure: Analyses focus on the difference in claim costs between an early evaluation (such as 30 days) and ultimate

- Claim characteristics are used to develop a "Claim Score"
  - The lower the score the more likely a claim has a large increase in losses (for example, $100,000 increase in total losses between 30 days and ultimate)

- Claim Scores can be arranged into Score Bands

Milliman

Characteristics with a Predictive Influence for Claim Scores

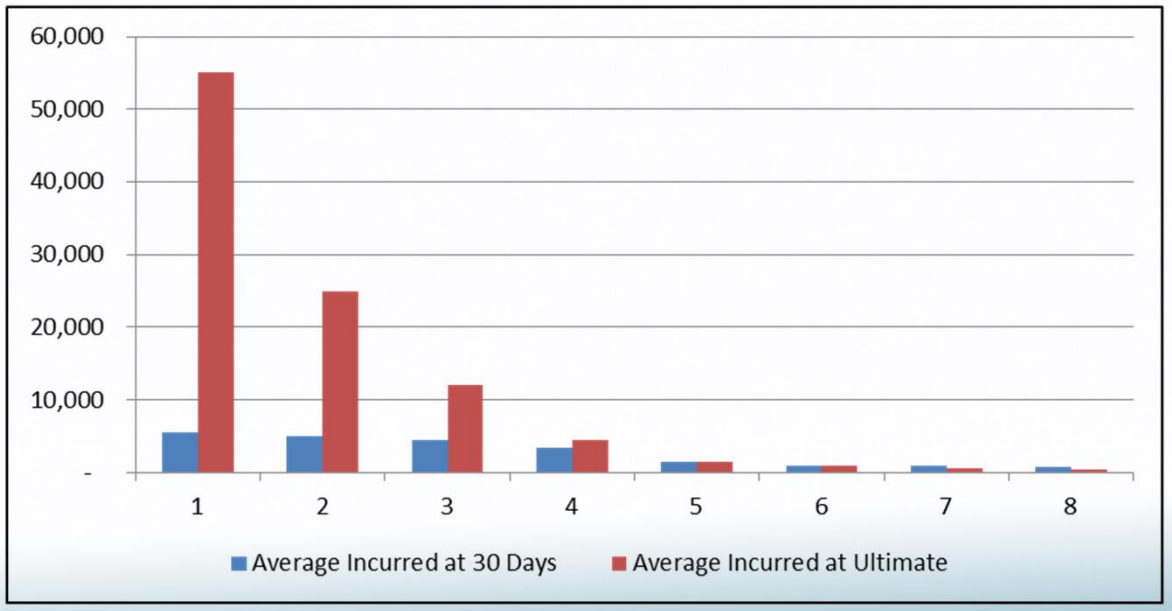
- The table at right lists the variables found to have a predictive influence on the Claim Score.
- The second column in the table presents the relative influence of a predictor on the Claim Score. The injured worker's age at the time of injury, the injured body part, and nature of the injury were the characteristics with the largest influence on the Claim Score.
- Average Claim Scores can be calculated for the selected claim characteristics.

Predictor Variable	Influence on the Jumper Claim Score
AGE_AT_TIME_OF_INJURY	12.0%
PART_OF_BODY	10.7%
NATURE_OF_INJURY	8.6%
AVERAGE_WEEKLY_WAGE	5.5%
ICD9_Diagnosis	4.5%
NUMBER_OF_PRIOR_CLAIMS	4.4%
Number of Office Visits	4.4%
Number of Physical Therapy	3.8%
Surgery (per Adjusters' Notes)	3.7%
Present of attorney (per Adjusters' Notes)	3.0%
Industry of Employment	2.0%

Milliman

Jumper Claims - Sample Results at 30 Days

- Graph presents the average incurred at 30 days and at ultimate.
  - For claims in Score Band 1 (7% of all claims), the average incurred was $5,500 at 30 days and $55,000 at ultimate.
  - For claims in Score Band 2 (5% of all claims), the average incurred was $5,000 at 30 days and $25,000 at ultimate.



Milliman

Scoring Analysis:
10% of Claims = More than 50% of Total Losses

- Illustrative results: Claims with the largest expected losses have the lowest Jumper Claim Scores.
  - Claims in Score Band 1: 5% of all claims and 35% of incurred losses at ultimate.
  - Claims in Score Bands 1 and 2: 10% of all claims and 52% of incurred losses.

Score Band	Score Range	Percent of Claims	Percent of Incurred Losses	Cumulative Percent of Claims	Cumulative Percent of Incurred Losses
1	**001 – 099**	**5%**	**35%**	5%	35%
2	**100 - 199**	5%	17%	**10%**	**52%**
3	200 - 399	20%	23%	30%	75%
4	400 - 599	35%	20%	65%	95%
5	600 - 799	25%	4%	90%	99%
6	800 - 1000	10%	1%	100%	100%

Milliman

Business Stratifications and Expected Results

- Business Stratifications: Segmentation and Scoring Analyses can be designed to focus on certain stratifications of a book of business or claims.
  - Underwriting considerations (size of account, loss-sensitive business).
  - Geographic considerations (analyses by region or state).
  - Claim characteristics (separate analyses for most serious claims).

- Expected Results: influenced by many factors.
  - Current state of claim-cost experience.
  - Ability to introduce new processes into the claims operations.
  - Support from senior management.
  - Regulatory environment.

Implications for Actuarial Analyses

Two Areas of Actuarial Focus

- The underlying assumptions
- The impact (actuarial implications)

Starting Actuarial Issues

- Analyses will use historical data, most likely from multiple years
- Trend
  - Should the data be trended?
  - If so, should the trend be to a period within the evaluation or a future period (e.g., rating period)?
- Loss Development
  - Should the losses be developed to a specific (common) evaluation?
  - Should the losses be developed to an ultimate basis?
  - What development factors should be used?
  - How should development factors be applied?
- Large Losses
  - Should losses be limited?
  - Should large losses be included in the analysis?

Answers Depend on

- What data are you using?
  - Individual claims in one accident year / policy year or many years
  - Closed, open claims or both
  - Single state or multiple states included
  - Long-tail or short-tail business
- What is the purpose of the predictive modeling project?

Data Used in Claims Example

- Predictive Modeling Purpose:
  - Early identification of claims with the potential to become high-loss claims

- Scope of Predictive Modeling analysis:
  - One State
  - Open and closed claims
  - Claims evaluated as of 6/30/2014 for accident year 2011
  - Claim characteristics at 30 days

- Actuarial Assumptions
  - No trending – one accident year all claims evaluated at same time
  - No development – claims evaluated at evaluation point
  - Treatment of large losses (per 30-day information) – removed: fatalities, perm totals, and claims with incurred losses over $500,000 (e.g., severe burns)

Potential Data Used in Alternative Example

- Multiple States
- Open and closed claims
- Three accident years of claims evaluated as of 18, 30 and 42 months of development
- Claim characteristics at 45 days

- Assumption
  - In order to obtain sufficient data – results of multiple states and multiple accident years need to be combined

Potential Data Used in Alternative Example

- Actuarial Adjustments
  - Trend individual claims to common accident year –
    - Does trend vary by coverage (medical / indemnity losses)?
    - Does trend vary by state (states with medical fee schedules verses no fee schedule)?
    - Does trend vary by type of claim, type of injury?
  - Develop individual claims to common accident year
    - Develop open claims only? (What about re-opens?)
    - How are development factors derived?
      - Aggregate development applies to both open and closed claims
      - Aggregate development includes IBNR claims
    - Do you develop incurred losses, paid losses or case reserves
    - Does development vary by state, type of claims, type of injury
  - Do your assumptions create a bias?
    - Should other methods be considered besides development?

Sample Questions to be asked in Predictive Modeling Projects

- Is Premium being used
  - Should premium be trended or on-leveled for rate changes?
  - How is premium on-leveled?
  - What biases are introduced for on-leveling?
- Are Exposures being used?
  - Should exposures be trended?
- Are Claim Counts being used?
  - Should counts be developed?
  - Are IBNR claims allocated to policy?

Sample Questions to be asked in Predictive Modeling Projects

- Impact of large losses
  - What are appropriate methods to handle these?
  - Cap to limit volatility?
  - How to decide on a cap?
  - Remove completely from the modeling?
  - How do you reflect the ultimate value of large losses?
- Removal of immature accident years
  - Can this data be used at all (i.e. by adjusting exposure)? Might be difficult to accurately develop very immature claims, which may skew/bias results
  - Removing too many years of data makes modeling less responsive to changes or shifts in book of business

Sample Questions to be asked in Predictive Modeling Projects

- Creation of Holdout Dataset
  - Use odd/even years as holdout or a percentage of the records?
  - Use older years as training data and more recent years as validation?
  - Potential for cyclical bias? (Effects of recessions on WC claims, etc…)
- Other Considerations
  - Has book of business changed over time?
    - Are the risks written 5 years ago the same as the risks written today? If not, need to adjust data accordingly or remove the years where the risks are different.
    - Do any reforms in a state need to be accounted for and adjusted in the model?

Impacts/ Outcomes of Predictive Modeling Projects

- Objective
  - Early recognition of potential high-loss claims to direct specialized adjuster resources and therefore to control / reduce claim costs

- Potential Implications
  - Speed up reporting pattern
  - Reduction in ultimate losses

Potential Implications

- Baseline – Year before claims initiative implemented

	Premium (000's)	Expected Ultimate Loss (000's)	Reported Losses at 12 Months (000's)	Percent Reported at 12 Months
BaseLine	10,000	6,500	3,575	55.0%

- Assumptions
  - Claims initiative will reduce overall losses by 5%
  - Claims will be reserved quicker

Milliman

Scenario 1 – No change to reported losses

- First Year after claims initiative implemented

	Premium (000's)	Expected Ultimate Loss (000's)	Reported Losses at 12 Months (000's)	Percent Reported at 12 Months
BaseLine	10,000	6,500	3,575	55.0%
Scenario 1	10,000	6,175	3,575	57.9%

- Is there a "speed up" in reporting of the losses
  - Cannot tell with aggregated data
  - Impact of large losses
  - Change in distribution of claims
  - Using last year's pattern – projected ultimate losses $6,500

Milliman

Scenario 2 – Increase in Reported Losses

- First Year after claims initiative implemented

	Premium (000's)	Expected Ultimate Loss (000's)	Reported Losses at 12 Months (000's)	Percent Reported at 12 Months
BaseLine	10,000	6,500	3,575	55.0%
Scenario 2	10,000	6,175	3,933	63.7%

- Is there a "speed up" in reporting of the losses
  - Cannot tell with aggregated data
  - Impact of large losses
  - Change in distribution of claims
  - Using last year's pattern – projected ultimate losses $7,150

Milliman

Diagnostics to Monitor Objectives

- How many claims have been reassigned to specialized adjusters
  - before and after initiative
- Average severities
  - Before and after initiative
  - By type of claim
- Closing rate
- Impact of large claims
- Communication with claim staff
  - Changes to case reserving practices (case reserve and development pattern implications)

Presentation Summary

Milliman

Presentation Summary

- Business Objectives and Predictive Analytics

- Predictive Analytics
  - Starting Considerations
  - Segmentation Analyses
  - Claim Score Analyses
  - Business Stratifications

- Implications for Reserving
  - Assumptions in predictive modeling
  - Impact (actuarial implications) of initiatives