Session 93, Individual and Group LTD Experience and Lessons Learned

SOA Antitrust Disclaimer
SOA Presentation Disclaimer

GLTD Credibility Study Stage 2

Paul Correia, FSA
Milliman Inc.

June, 2019

SOCIETY OF ACTUARIES

GLTD Credibility Study
Stage 2 Objectives

- Develop manual rates
- Test different credibility formulas commonly used for pricing LTD
- Test predictive modeling methods
- Identify variables important for predicting future experience
- Generate case rates using predictive modeling methods
- Compare PM method to standard industry approaches

SOCIETY OF ACTUARIES

Data and Analytical Methods

- Policy and claim data submitted by 14 disability insurers
  - 300,020 claims incurred between 2004 – 2011
  - 102,951 policies inforce for at least 5 consecutive years between 2004 – 2011

- Claim Cost Ratio = $\dfrac{\text{PV Expected Future Benefits}}{\text{Covered Payroll}}$

- PV calculated at 3.5% as of the end of the elimination period, based on claim termination rates from the 2012 GLTD Basic Table, and the gross benefit amount payable under the policy with no offsets.

- Relative error = $\dfrac{\text{Abs. value (Predicted Rate – Actual Rate)}}{\text{Predicted Rate}}$

Development of Manual Rates

- Used predictive modeling methods to develop manual rates

Training variable: Preliminary rate that varies by case size, elimination period, voluntary indicator, definition of disability, and industry.

Independent variables:

- Industry
- Region
- Elimination Period
- Benefit Percent
- Benefit Period
- Voluntary Indicator (employer-paid vs. employee-paid)
- COLA
- Definition of Disability
- Integration with STD
- Case Size

Dependent variable: Claim rate from three-year experience period

Variable Importance	
Variable	**Importance**
STD Integration	53.5%
Industry	15.3%
Region	10.1%
COLA	5.7%
Case Size	4.2%
Definition of Disability	3.4%
Voluntary Indicator Group	3.1%
Elimination Period	2.6%
Benefit Percent	1.9%
Benefit Period	0.1%
GRAND TOTAL	**100%**

- Output is a unique manual rate for every case

Credibility Formulas Tested

- Industry Formula 1:

$$Z_1 = \text{Min}\ [\ 100\%\ ,\ \sqrt{\frac{\text{LYE}}{25{,}000}}\ ]$$

- Industry Formula 2:

$$Z_2 = \text{Max}\ [\ 0\%,\ \text{Min}\ [\ 100\%\ ,\ \frac{(\text{Exp.claims per 1000})\times\frac{\text{LYE}}{1000}}{(\text{Exp.claims per 1000})\times\frac{\text{LYE}}{1000}+25-\frac{\text{LYE}}{1000}}\ ]\ ]$$

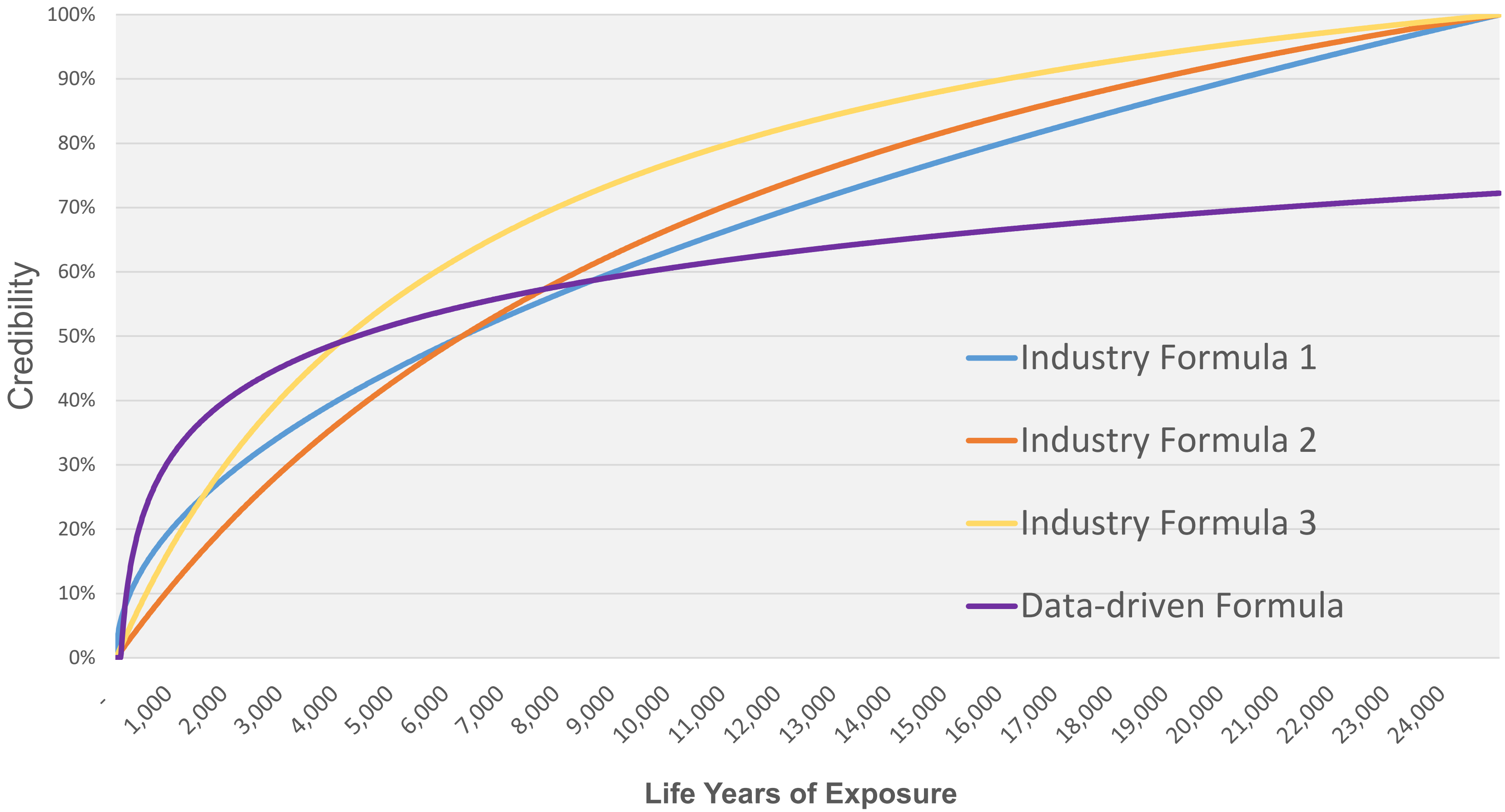
- Industry Formula 3:

$$Z_3 = \text{Max}\ [\ 0\%,\ \text{Min}\ [\ 100\%\ ,\ \frac{(\text{Actual claims})}{(\text{Actual claims})+25-\frac{\text{LYE}}{1000}}\ ]\ ]$$

- Data-driven Formula : based on experience data used for this study, and designed to minimize relative error between predicted and observed claim costs within each LYE group.

$$Z_4 = \text{Max}\ [\ 0\%,\ \text{Min}\ [\ 100\%,\ 0.1272 * \ln(\text{LYE}) - 0.5657\ ]\ ]$$

Comparison of Credibility Formulas



Comparison of Credibility Formulas

- Industry Formula 3 produces lowest relative error in most LYE segments and overall
- Industry formulas, when compared to data-driven approach based solely on LYE, produce reasonable credibility weights

Relative Error Comparison of Credibility Formulas				
LYE Group	**Data-driven**	**Industry 1**	**Industry 2**	**Industry 3**
0-99	180.0%	178.9%	179.4%	175.0%
100-499	137.0%	136.8%	137.4%	132.9%
500-999	90.3%	89.6%	90.0%	86.1%
1,000-1,999	69.1%	69.5%	69.9%	66.4%
2,000-2,999	56.0%	56.7%	56.5%	54.4%
3,000-3,999	47.6%	48.8%	48.4%	46.7%
4,000-4,999	44.3%	44.6%	44.3%	42.8%
5,000-7,499	40.1%	40.4%	40.0%	40.3%
7,500-9,999	36.9%	37.1%	38.0%	35.0%
10,000-19,999	30.7%	31.0%	31.9%	29.1%
20,000-29,999	26.2%	28.9%	29.5%	28.5%
30,000-39,999	24.8%	24.8%	24.8%	24.8%
40,000-49,999	24.7%	28.8%	28.8%	28.8%
50,000+	25.8%	25.9%	25.9%	25.9%
GRAND TOTAL	**63.1%**	**63.5%**	**63.8%**	**61.6%**

Full Credibility Thresholds

- The 25,000 full credibility threshold produces lowest overall relative error and for LYE groups < 7,500
- Higher full credibility thresholds produce lower relative errors for larger LYE groups, implying that optimal credibility formula would approach, but never reach, full credibility

Relative Error Comparison – Alternative Maximum Credibility Thresholds for Industry Formula 3

LYE Group	Maximum Credibility Threshold			
	25,000	35,000	45,000	55,000
0-99	175.0%	176.0%	176.6%	177.1%
100-499	132.9%	134.0%	134.8%	135.3%
500-999	86.1%	86.1%	86.1%	88.2%
1,000-1,999	66.4%	67.4%	68.2%	68.8%
2,000-2,999	54.4%	55.3%	56.1%	56.8%
3,000-3,999	46.7%	46.7%	46.7%	49.9%
4,000-4,999	42.8%	42.8%	42.8%	44.4%
5,000-7,499	40.3%	40.3%	40.3%	41.8%
7,500-9,999	35.0%	35.0%	35.2%	35.5%
10,000-19,999	29.1%	29.1%	28.9%	29.3%
20,000-29,999	28.5%	28.5%	28.5%	25.8%
30,000-39,999	24.8%	24.8%	24.8%	23.9%
40,000-49,999	28.8%	28.8%	28.8%	23.9%
50,000+	25.9%	25.9%	25.9%	25.8%
Weighted Average	**61.6%**	**61.8%**	**62.2%**	**62.5%**

Predictive Modeling

Modeling Methods

- Models developed in R using xgboost package
- SHAP importance to identify key variables for predicting future experience
- Random Forest Model 1: Manual Rates
- Random Forest Model 2: Identify key variables
- Random Forest Model 3: Generate case rates
- Test predicted values from RF model against industry formulas

Evaluation Criteria

- Relative errors
- Percentage of cases closest to actual claim rate
- Buckets of disagreement
- Efficient frontier analysis

SOCIETY OF ACTUARIES

Random Forest Model - Case Rates

- RF Model used to generate case rates
- Based on the following independent variables

➢ Delta_pct – variable representing ratio of the experience rate to the manual rate

➢ BetterOrWorse – indicator for whether experience rate is higher or lower than the manual rate

➢ Claim Count – number of claims incurred in the experience period

➢ Total LYE – exposure within 3-year experience period

SHAP Importance	
Variable	**Importance**
Delta_pct	49.7%
BetterOrWorse	24.0%
Claim Count	16.3%
Total LYE	10.0%
GRAND TOTAL	**100.0%**

Test RF Method Against Industry Approaches

Relative Errors

Weighted Average Relative Error (by LYE)				
LYE Group	**RF3 Model**	**Industry 1**	**Industry 2**	**Industry 3**
0-99	175.4%	178.9%	179.4%	175.0%
100-499	131.7%	136.8%	137.4%	132.9%
500-999	87.2%	89.6%	90.0%	86.1%
1,000-1,999	65.9%	69.5%	69.9%	66.4%
2,000-2,999	52.6%	56.7%	56.5%	54.4%
3,000-3,999	44.2%	48.8%	48.4%	46.7%
4,000-4,999	42.2%	44.6%	44.3%	42.8%
5,000-7,499	38.9%	40.4%	40.0%	40.3%
7,500-9,999	34.9%	37.1%	38.0%	35.0%
10,000-19,999	28.5%	31.0%	31.9%	29.1%
20,000-29,999	26.3%	28.9%	29.5%	28.5%
30,000-39,999	23.6%	24.8%	24.8%	24.8%
40,000-49,999	24.6%	28.8%	28.8%	28.8%
50,000+	22.3%	25.9%	25.9%	25.9%
Weighted Average	**60.3%**	**63.5%**	**63.8%**	**61.6%**

- Predicted values from RF model tend to be closer to actual future claim costs

Test RF Method Against Industry Approaches

Percentage of cases closest to actual claim rate

Percentage of Cases with Closest Predicted Values				
LYE Group	**RF3 Model**	**Industry 1**	**Industry 2**	**Industry 3**
0-99	76%	5%	11%	8%
100-499	62%	5%	18%	15%
500-999	55%	3%	16%	25%
1,000-1,999	56%	3%	15%	26%
2,000-2,999	55%	9%	11%	26%
3,000-3,999	55%	18%	7%	21%
4,000-4,999	40%	21%	15%	24%
5,000-7,499	27%	27%	28%	19%
7,500-9,999	30%	24%	25%	21%
10,000-19,999	36%	22%	22%	20%
20,000-29,999	33%	22%	22%	23%
30,000-39,999	28%	24%	24%	24%
40,000-49,999	24%	25%	25%	25%
50,000+	22%	26%	26%	26%
Total	**68%**	**5%**	**14%**	**13%**

Test RF Method Against Industry Approaches

Buckets of Disagreement

1. Difference between manual rate and the predicted value calculated for every case

2. Cases sorted from smallest to largest difference

3. Cases divided into 10 equal buckets

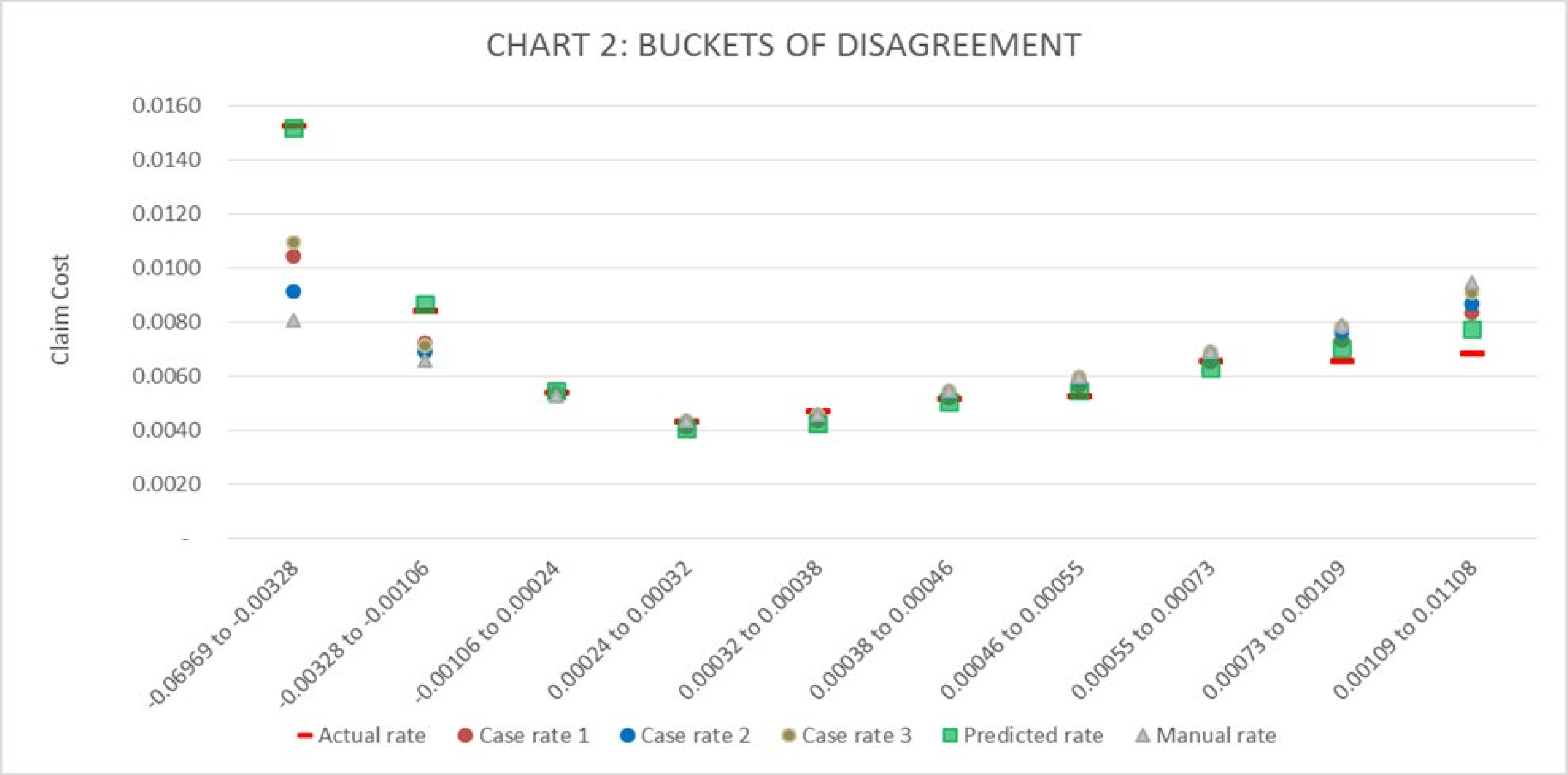
   e.g., if there are 100 observations then the first bucket would contain the 10 observations with the smallest difference between the manual and predicted value

4. Within each bucket, calculate the average manual rate, average predicted values, and average actual claim rate

5. Compare values for each bucket to determine which predicted rates are closest to actual claim rates

Test RF Method Against Industry Approaches

Buckets of Disagreement



- Results support conclusion that RF model predicted values are, on average, closer to actual future claim rate than industry formulas

Test RF Method Against Industry Approaches

Efficient Frontier Analysis

- Evaluate different pricing methods through model that projects future sales and profitability

- For every case, calculate “market rate” to determine likelihood of cases selling

- For every case that sold, estimate earned premium over the next two years:

  Premium = Predicted Rate x Covered Payroll in subsequent period

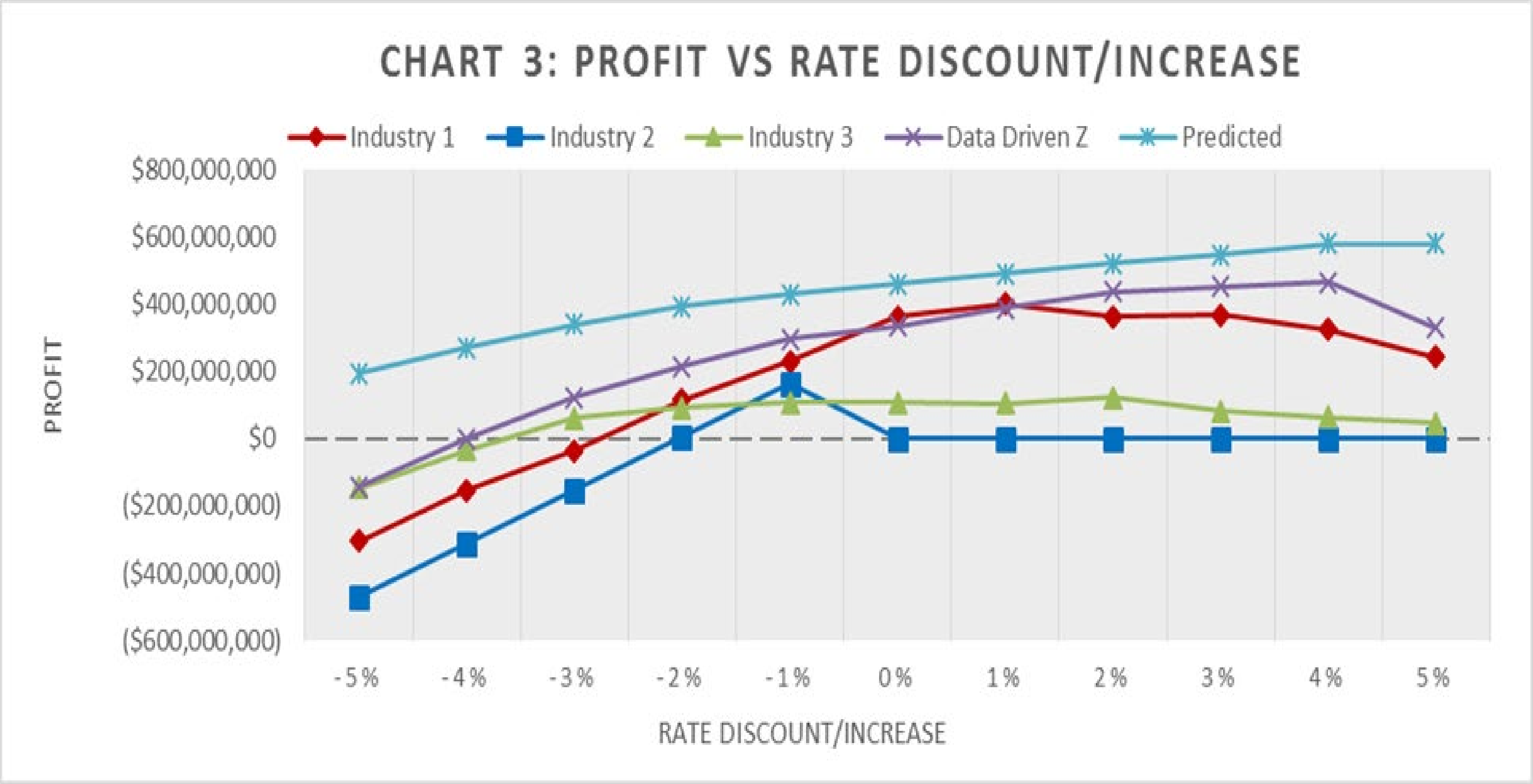
- Determine incurred claims in subsequent two years for all cases that sold

- Calculate gains/losses as difference between the earned premium and incurred claims:

  Gain/Loss = Premium from Step 1 minus Incurred Claims from Step 2

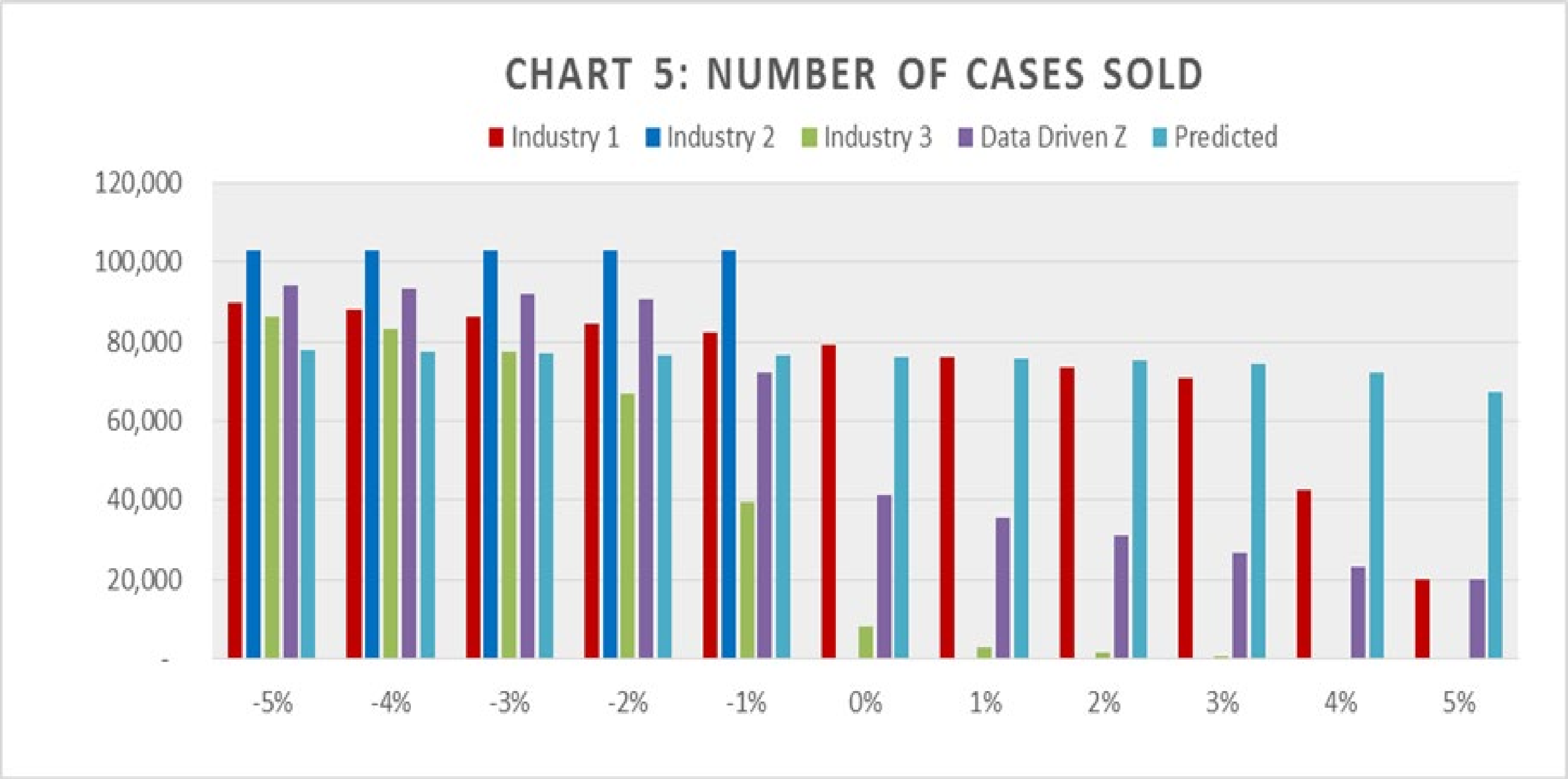
Test RF Method Against Industry Approaches

<u>Efficient Frontier Analysis</u>



Test RF Method Against Industry Approaches

Efficient Frontier Analysis



Conclusions:

- Improving refinement of manual rates leads to better ability to predict claim costs, allowing for reduced credibility

- Increasing full credibility threshold produces better predictions for larger size groups, indicating that that optimal credibility formula would approach, but never reach, 100% credibility

- LTD pricing methods could potentially be improved upon by employing predictive modeling techniques in the development of rates

Thank you

Paul Correia, FSA
Paul.Correia@milliman.com

SOCIETY OF ACTUARIES

June, 2019

Munich RE

Session 093

Individual and Group LTD Experience and Lessons Learned

SOA Health Meeting - June 2019

Mark Costello, Vice President

Group and Living Benefits Reinsurance

In-force Management, Claims and Data Operations

Agenda

1. Lessons learned – IDI vs GLTD
2. Lessons learned -- globally
3. Lessons learned – medical market
4. Lessons learned – individual life

Lessons learned: IDI vs LTD

Section 1

Current approach to LTD morbidity setting

- SOA 2018 Group Long-Term Disability Experience Study Report (Recovery/Death)
- Recent Company / Group Experience
- Manual Rates
- External Benchmark

Current approach to IDI morbidity setting

- Industry Experience (IDEC, CIDA)
- Company experience
- Internal Experience Studies

Section 2 -- shortcomings

LTD

- Dated experience (2018 GLTD is 2004-2012)
- Aligning the basis with the plan/group design
- Adjusting for group movement and aging

IDI

- Dated experience (2013 IDEC is 1990-2007)
- Company experience not granular
- Backward looking and changing benefits/underwriting
- Explosion of other sources

Section 3 -- IDI vs LTD

- Pricing and reserving IDI requires longer projection than LTD.
- For IDI you need to focus on Incidence, Termination, Mortality and Lapse
- In the case of LTD that main focus is termination.

However historically at Munich Re, they were developed from different angle:

- GLTD policyholder data is limited until a claim is filed, so GLTD pricing relies heavily upon termination experience and assumption setting.
- IDI on the other hand has detailed policyholder data for incidence experience, and thus balances incidence and termination in setting pricing assumptions.

Side by side comparison is difficult:

- There are a number of adjustments to make on both sides before rates can be compared apples to apples.

Section 3 -- IDI vs LTD

Examples of potential adjustments needed:

- IDI Accident and Sickness need to be combined
- GLTD Recovery and Death need to be combined
- GLTD EP needs to focus on equivalent EP’s with IDI
- IDI / GLTD Company adjustments should be averaged out
- IDI Disability Definition adjustment needs to be applied

Termination rate slopes by

- EP
- Gender
- Age
- Claim Duration

Adjustments for Gross Monthly Benefit amount

Section 3 -- IDI vs LTD

Potential cross-overs

- Claim Diagnosis
- Medical vs Non-Medical Occupations (or Occ Class)
- Issue State

Lessons learned: Globally

Section 4 -- IDI vs LTD globally

Product Design, distribution, features and riders are similar

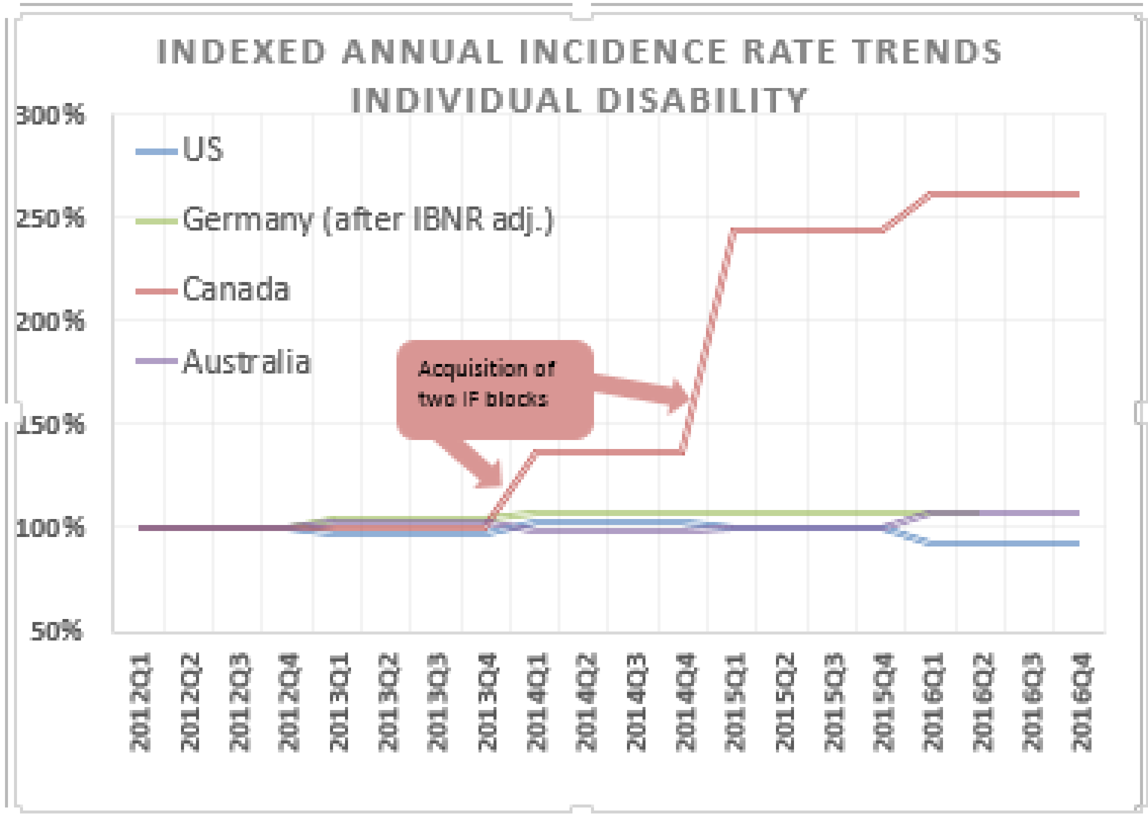
- IDI: CAN, Australia, Israel
- Group: CAN

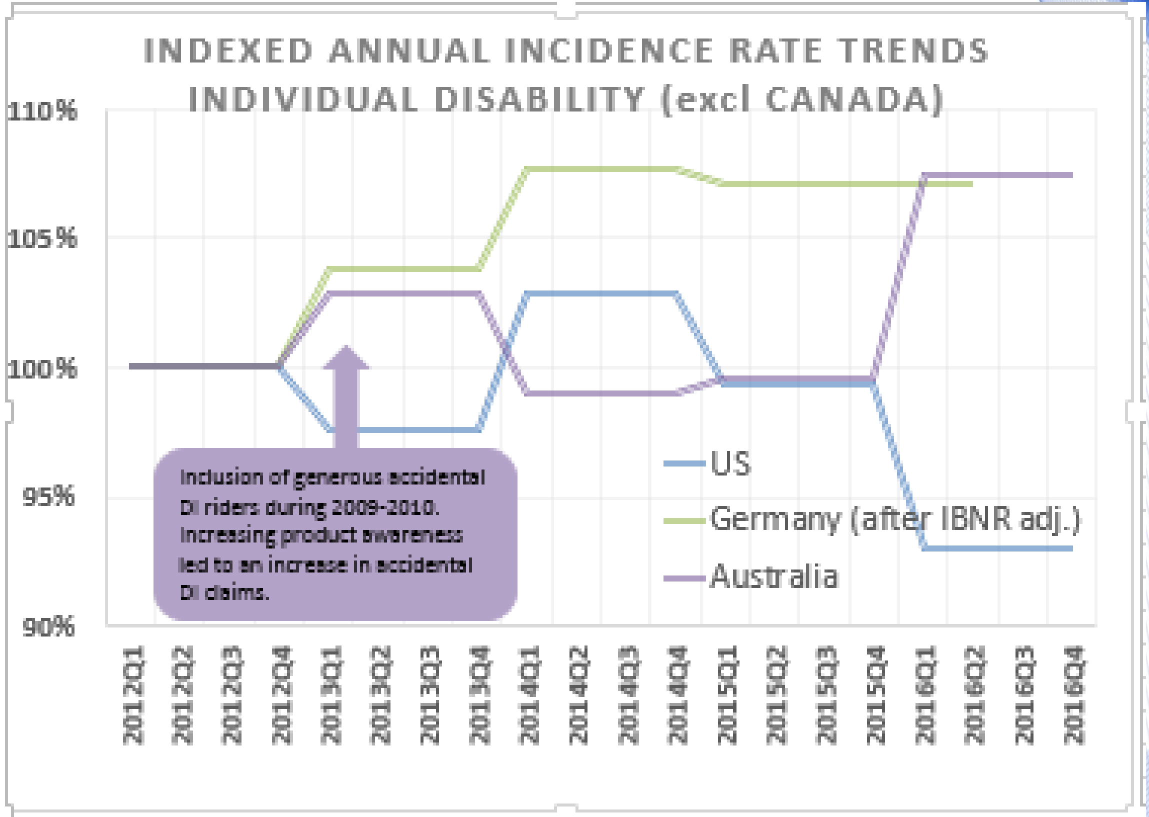
Lessons learned

- USA: let's remember the past…
- Australia : Both individual and Group under spotlight
- Israel : New insurance commissioner directives
- Netherland: Group (WIA) and individual business (AOV) but a segment of WIA had issues (WGA ERD)
- South Africa – change in regulation – Taxable/non taxable

Individual disability incidence rate trends

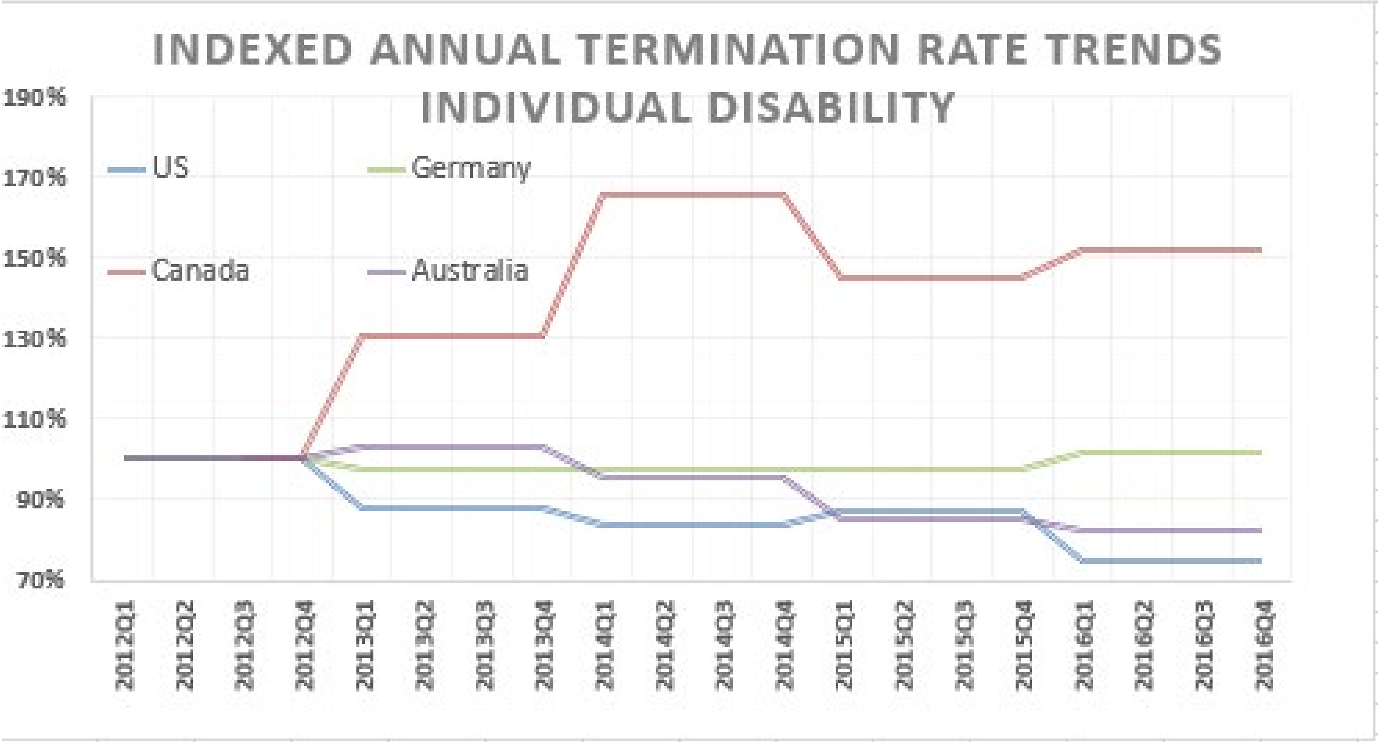
Trends and one-off impacts can be observed when analyzing annualized incidence rates





Individual disability termination rate trends

Trends and one-off impacts can be observed when analyzing annualized termination rates



Miscellaneous US Trends

Overall incidence and termination rates both continue to drop from one year to the next.

• Consistent with the recent Munich Re experience study, medical incidence rates are trending downwards and getting closer to non-medical white collar (occupation class 1). Thus, we see that gap as narrower as compared to what is indicated in the IDEC tables.

• Lately, medical termination rates are about the same as those for occ (occupation) class 1.

• Also consistent with the recent Munich Re experience study, recent incidence rates for physicians are lower than for other medical occupations such as nurses and dentists.

• Incidence is dropping quickly for employee paid GSI. It is still higher than employer paid GSI, but now materially lower than fully underwritten individual. That could be due to tightening of issue criteria for employee paid GSI business.

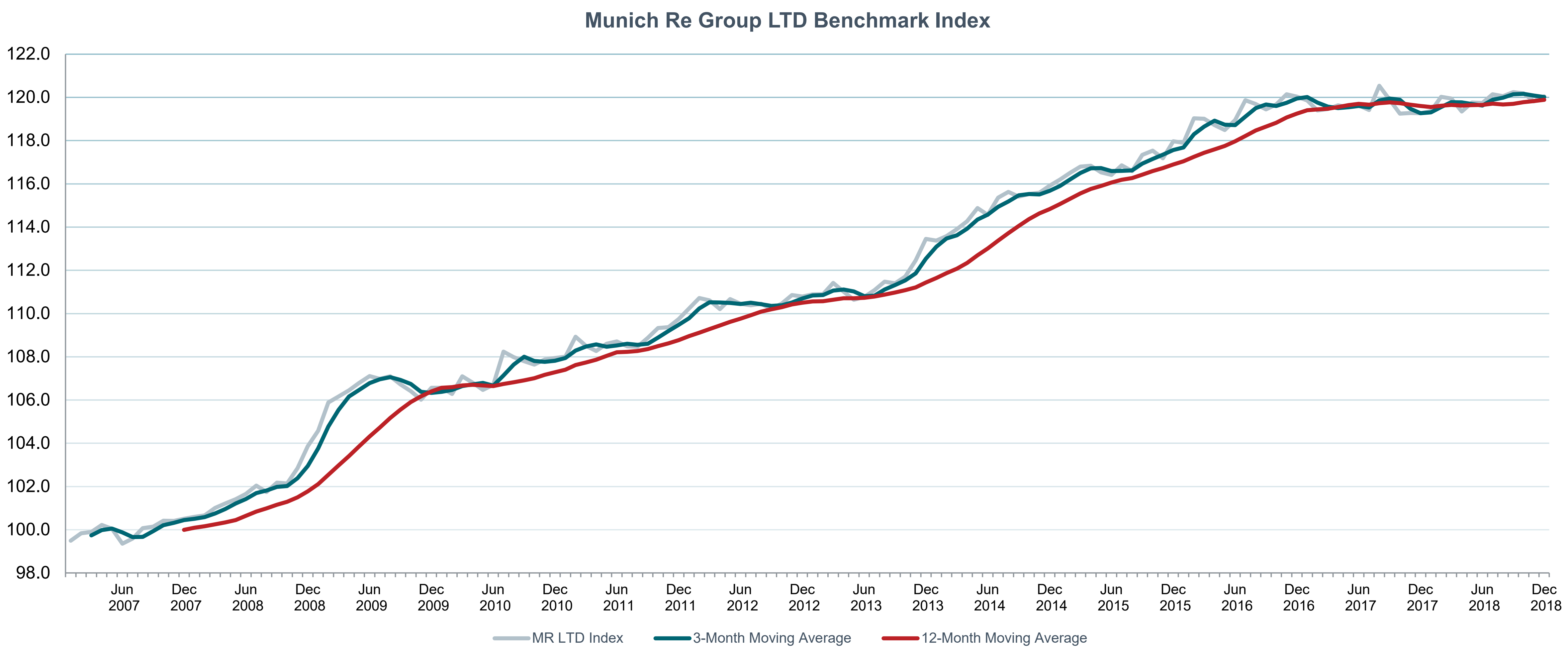
• There is no material difference between male and female termination rates for GSI, Male termination is lower than Female for non-GSI.

• The percentage of smoker policies is gradually declining as a percent of total active policies.

• Large policies (monthly benefit amounts of 10,000 or more) are becoming more common.

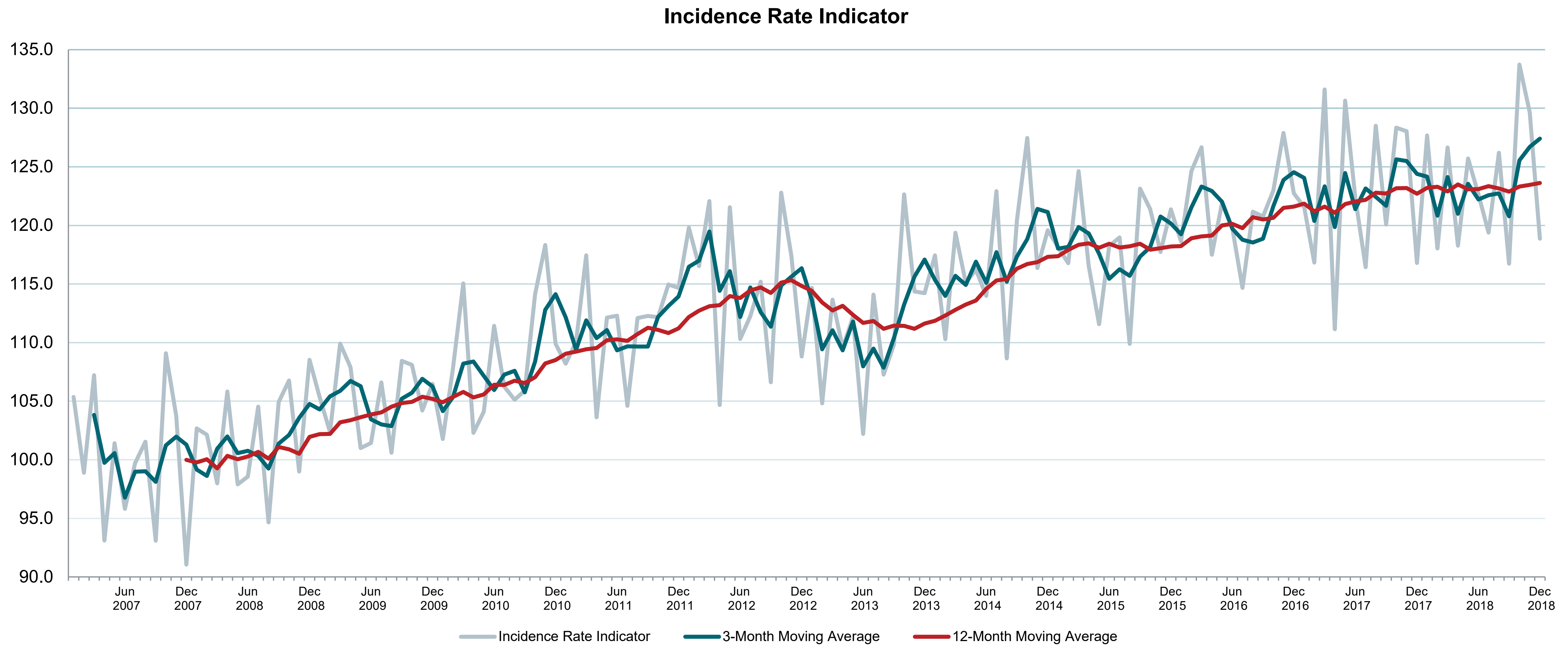
Munich RE

Munich Re Canada LTD Benchmark Index – Q4 2018

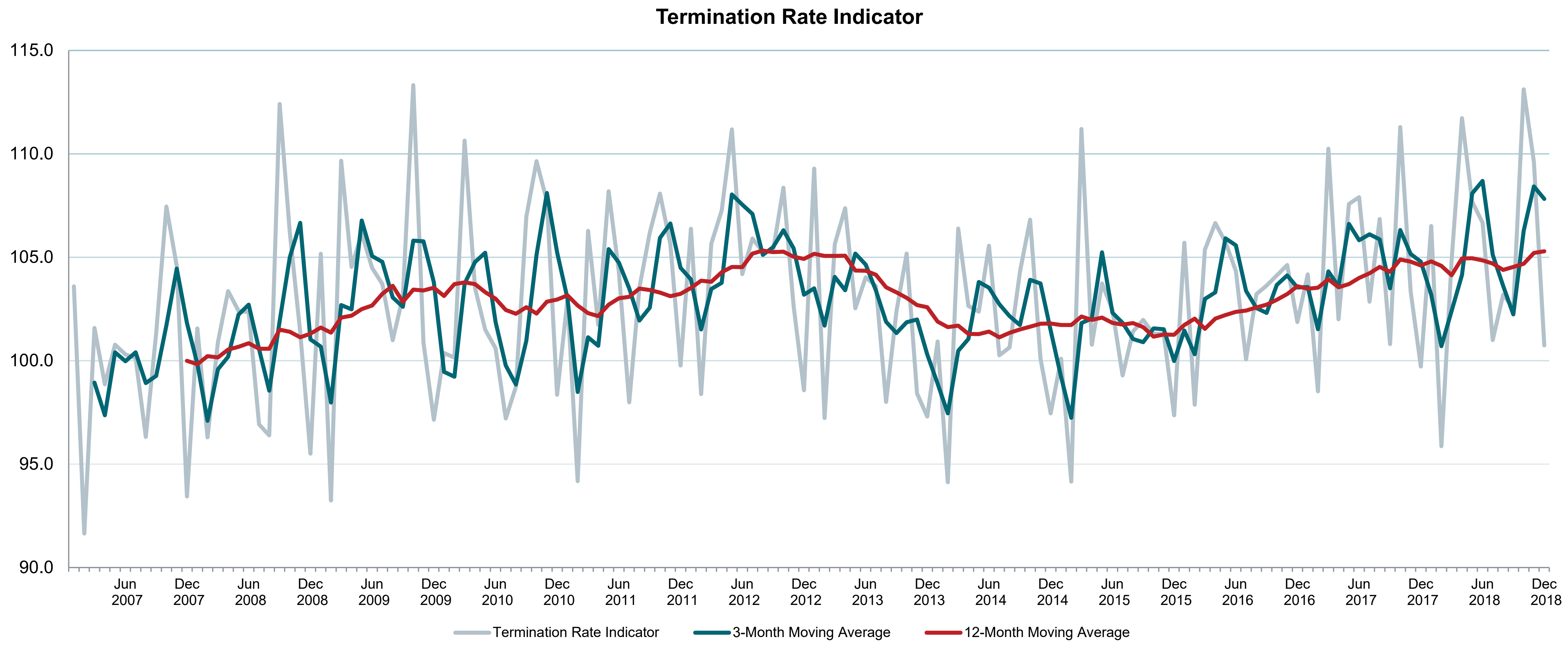


Munich RE

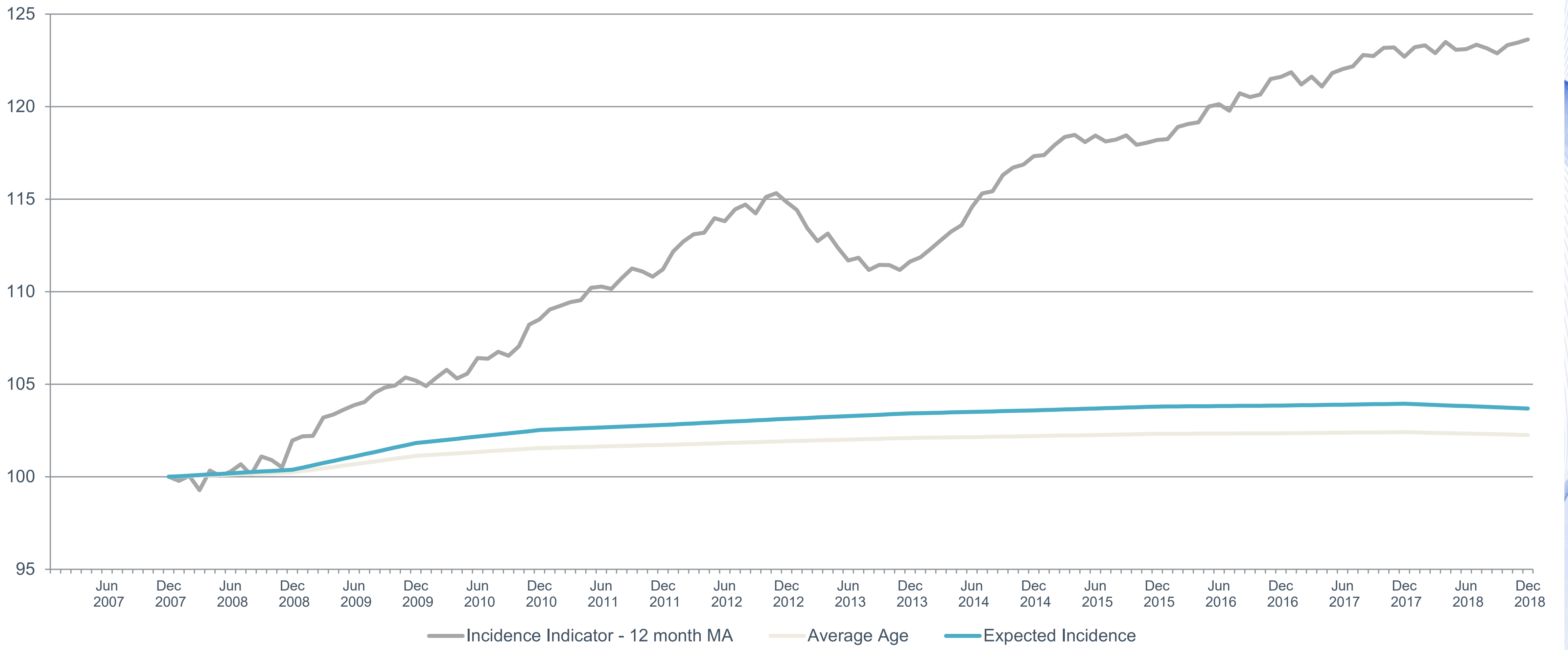
Munich Re LTD Benchmark Index – Q4 2018 Incidence Indicator



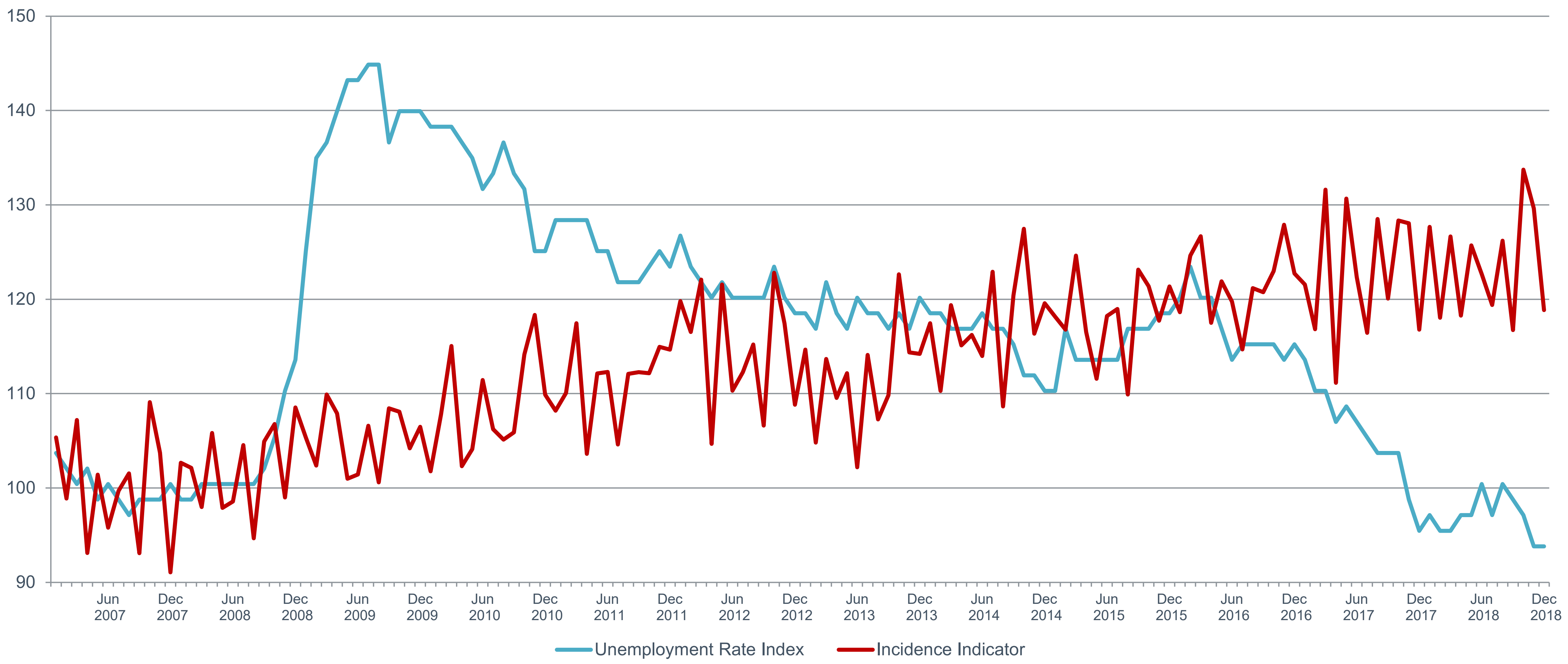
Munich Re LTD Benchmark Index – Q4 2018
Termination Indicator



Does Workforce Aging Provide any Clues?



Link Between Unemployment and Incidence?



Lessons learned: medical market

ACA vs. Managed Care

Compensation
Expectation of shortage -> better negotiating stance?
Less likelihood of payments shocks/reductions (i.e. vs. historical)

Workload
Increase demand –
- Medicaid expansion
- Exchanges and subsidies
- Aging demographics

Increased paperwork requirement

Work Environment
ACA is less negative than managed care
Managed care:
- Loss of Autonomy
- Increase bureaucracy
- Depersonalize physician-patient relationship

2013 – Potential impacts of ACA on Physicians

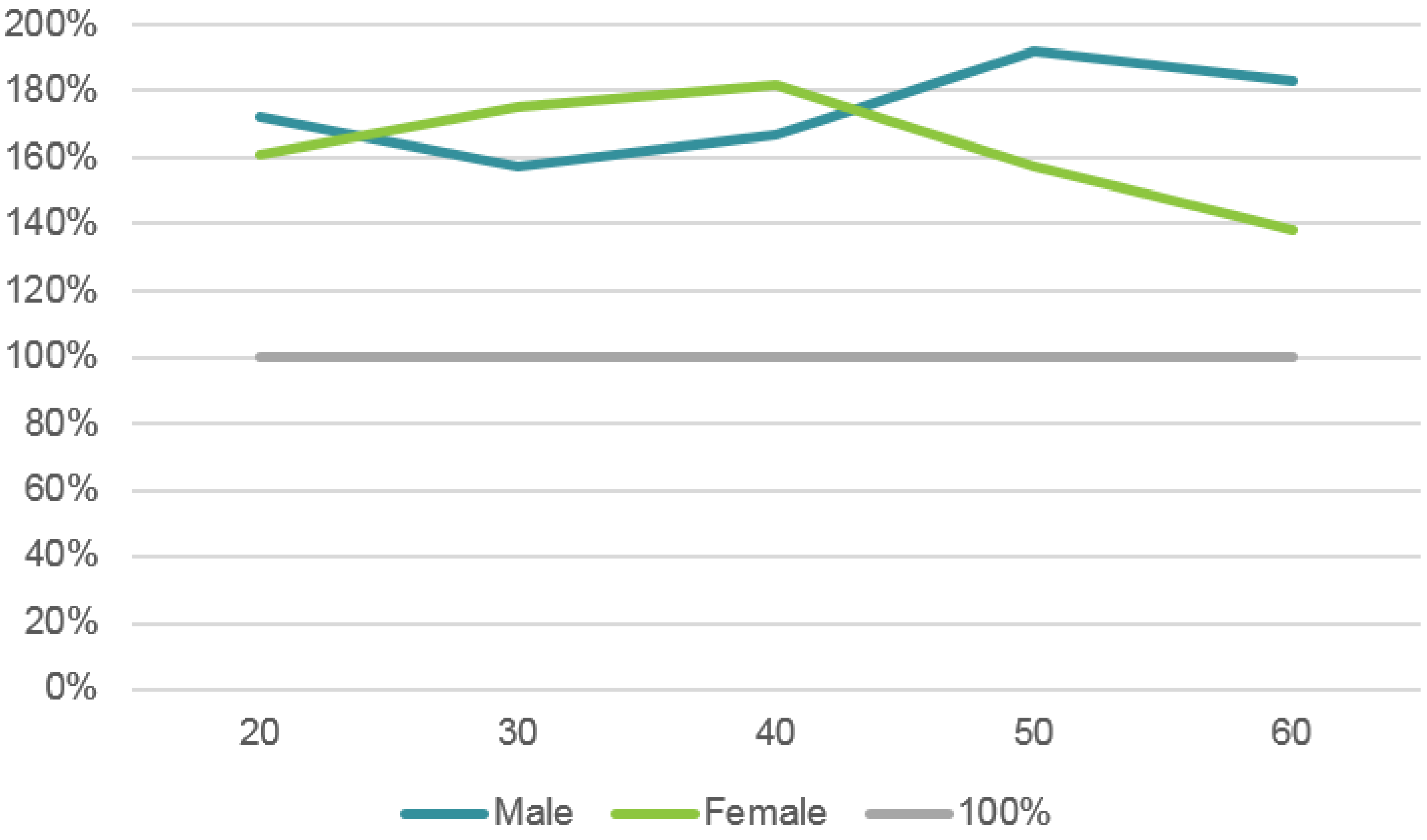
Workload
Relatively unchanged

Compensation
Generally increasing

Work environment
Generally improving but watch for burnout

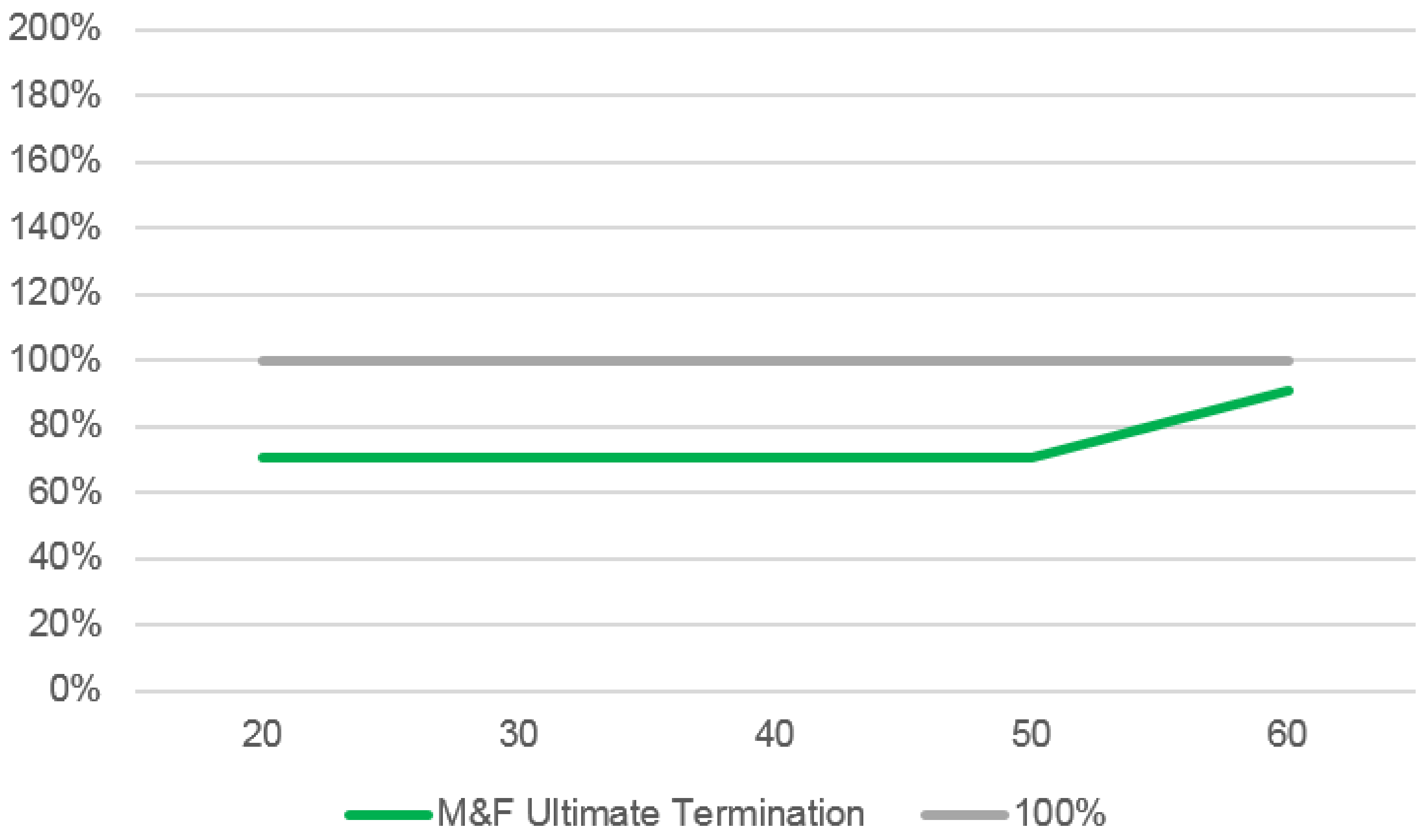
IDEC medical vs non-medical incidence

IDEC Incidence Base Rates - Sickness, 90 WP - Occ Class M / Occ Class 1 - By Age and Gender



IDEC medical vs non-medical termination

IDEC Termination Base Rates - Sickness, 90 WP - Occ Class M / Occ Class 1 - By Age and Gender



Medical vs non-medical incidence experience

Med vs Non-Med	A/E Incidence	Distribution	Credibility
Med	58%	61%	High
Non-Med	74%	39%	High
Total	64%		

- Most non-med is IDEC / CIDA occ class 1; other IDEC occ classes not credible on their own
- MR & IDEC: Outside of base tables, no additional incidence modifiers by occ class
- High incidence A/E for non-med driven by GSI

Incidence by Calendar Year

Year	Med	Non Med
2006	70%	86%
2007	47%	70%
2008	54%	85%
2009	64%	76%
2010	50%	74%
2011	59%	73%
2012	50%	70%
2013	58%	68%
2014	54%	54%
2015	66%	57%

Occupation groups incidence A/E vs IDEC

Occupation Group	<2014	2014+
Anesthesiologist	65%	80%
Cardiologist	46%	29%
Dermatologist	67%	20%
Medical Student	71%	30%
Nurse	92%	126%
Ophthalmologist	41%	51%
Physician	57%	40%
Physician Surgeon	56%	44%
Psychiatrist	18%	175%
Psychologist	21%	54%
Radiologist	37%	80%

Medical vs non-medical termination experience

Med vs Non-Med	A/E Termination	Distribution	Credibility
Med	112%	49%	High
Non-Med	87%	51%	High
Total	100%		

- MR & IDEC: Outside of base tables, no additional termination modifiers by occ class
- Within med only:
  - Female A/E termination < male A/E termination
  - Lower attained age A/E termination > higher attained age A/E termination

Occupation groups termination

Occ Group	A/E Termination	Credibility
Dentist	77%	Medium
Executive	93%	Medium
Lawyer	40%	Low
Physician	124%	High
Physician Surgeon	126%	Medium

Occupation groups termination A/E vs IDEC

Occupation Group	<2014	2014+
Anesthesiologist	141%	171%
Cardiologist	171%	89%
Dermatologist	94%	61%
Medical Student	78%	55%
Nurse	129%	51%
Ophthalmologist	85%	137%
Physician	124%	125%
Physician Surgeon	124%	133%
Psychiatrist	96%	0%
Psychologist	69%	0%
Radiologist	116%	53%

Lessons learned: individual life

Overview

Objective: Incorporate statistical and machine learning techniques to model mortality

Benefits:

- Smoother, more accurate mortality assumptions
- Faster, less manual fitting
- Improve understanding of drivers of mortality experience
- Insight into the relationship between new attributes and mortality

Path 1 : A/E GLM Model

- ✓ Developed base mortality table using glm
- ✓ Profit impact is similar to traditional build
- ✓ q_x output is in existing table structure
- ✓ Multipliers for additional factors

Path 2 : Epsilon Model

- ✓ Matched Epsilon consumer data to IL lives
- ✓ Insights from education, marital status, income and net worth

Path 1: A/E GLM model

Data

databricks
Spark
Azure

MARC IL 2017Q2 Experience Study

- Data prep
- Sampling
- 70% train/30% test

Base table factors

Gender
Smoker
Issue Age
Attained Age
Duration
Face Amount Band

Additional

Preferred Class
Product Type
Issue Year
Fleet Name
Auto/Fac
Standard/Rated
Calendar Year
Level Term Period

Base A/E GLM model

- Focus on core select experience and base factors
- Build glm models on 70% data
- Assess fit
  - Actual / Predicted
  - q_x plots
  - error metrics
- Perform cross validation
- Select top model and refit on entire dataset
- Output q_x in table format

Enhanced models

- Identify relevant additional variables, interactions using LASSO
- Improve understanding of mortality experience and trends
- Output glm coefficients as table multipliers

Base A/E GLM model
Model form

- Core select experience: issue ages 30-79 and durations 1-15
- Poisson GLM with log link, using M2015 as the offset (expected basis)
- Interaction factors selected based on plots and LASSO
- Top model selected based on a combination of fit, smoothness and simplicity

Main factors

- Attained Age
- Duration
- Gender
- Smoker Status
- Face Amount Band

Interactions

- Gender and Face Amount Band
- Gender and Attained Age
- Attained Age and Duration Group
- Issue Age (10yr group) and Attained Age

```
glm (formula = ActualClaimNAAR ~ Gender + SmokerStatus + fa_band + AttainedAge + Duration +
                                 Gender:fa_band + Gender:AttainedAge +
                                 AttainedAge:dur_group + ia_group10:AttainedAge,
     family = poisson, data = data, offset = log(ExpectedNAARM2015BI))
```

Summary and next steps

A/E GLM modeling

Summary

- A glm approach effectively and efficiently produces a base model table with a good fit and smooth rates that are comparable to the traditional build.

Path 2: third party data model

MARC IL 2017Q2 Experience Study

19 M individuals

Azure Search

Name
DOB
State

external 2012 database
250 M individuals
250 variables

Demographics: family composition, education, occupation, …

Affluence: income, net worth, property,…

Credit: active tradelines, mortgage amount, liquid resources, …

Lifestyle: fitness & exercise, nutrition, hobbies,….

6.2 M matches
(34% match rate)

Thank You!