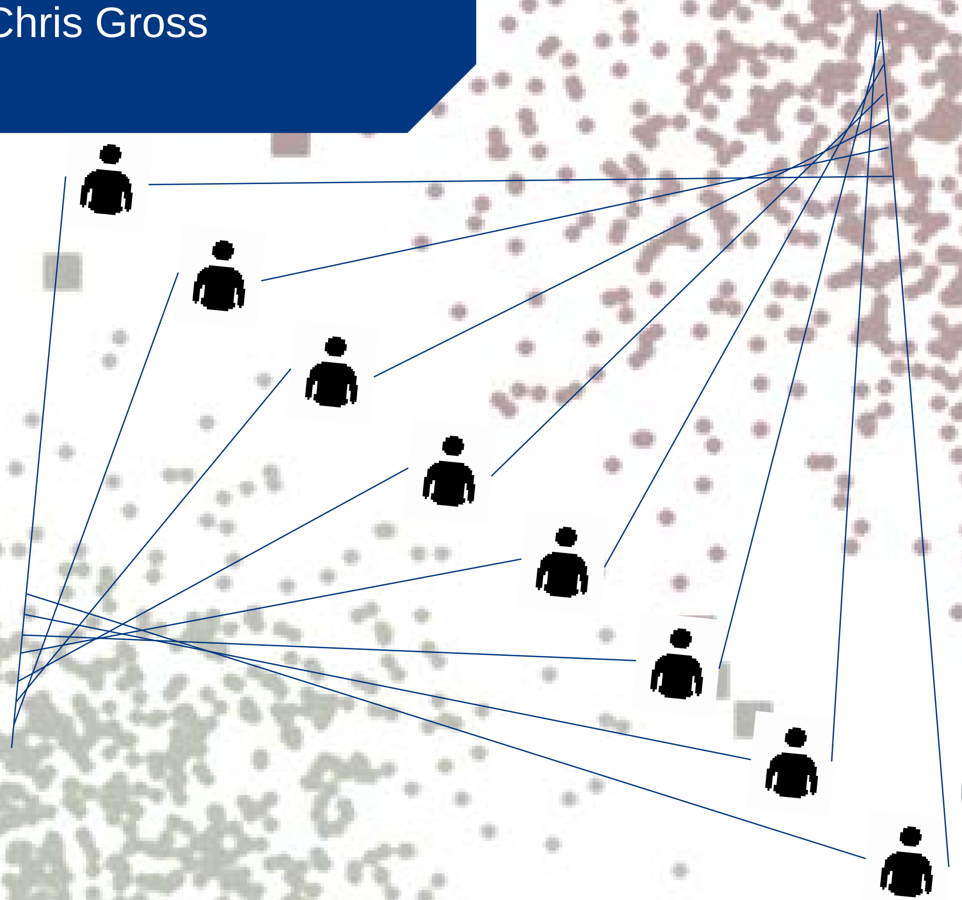
Claim Level Analytics

Ron Lettofsky and Chris Gross
May 23, 2017



Claim Level Analytics Working Party

- Granular data available (policies, claims, exposures, environment).
- Little is known about how to analyze such data and integrate the results into insurers' processes. P&C actuaries are insufficiently prepared to take advantage of this opportunity.
- CAS formed the Claims Level Analytics Working Party February 2016
- 22 members

Claim Level Analytics Working Party

- Purpose of the working party:
  - Determine what tools and methods are needed and what is currently available.
  - Investigate having a unified, sophisticated and continuously updated model that covers not only the claim generation but also the entire claim “lives” that could underlie both pricing and reserving and provide useful inputs for claim departments.
- Final report expected June 2017

Literature search

A few key papers are:

- "Loss reserving with GLMs: a case study",

  by Greg Taylor and Gráinne McGuire.
- "Individual claim modelling of CTP data",

  by Gráinne McGuire.
- "Loss Reserving Using Claim-Level Data“,

  by James Guszcza and Jan Lommele.
- "Using a Claim Simulation Model For Reserving & Loss Forecasting For Medical Professional Liability",
  by Rajesh V. Sahasrabuddhe.

Areas of Potential Benefit Identified by the Working Party

- Claims Management
- Pricing
- Actuarial Reserving
- Enterprise Risk Management
- Reinsurance

Areas of Potential Benefit Identified by the Working Party

- **Claims Management**
- Pricing
- Actuarial Reserving
- Enterprise Risk Management
- Reinsurance

— Data —

Data for Pricing Models

1. Policy Characteristics
    a. Exposures
    b. Limits and Deductibles
    c. Coverages and Perils
2. Rates, ILFs, Modification factors, etc.
3. Insured Characteristics
    a. Credit Information
    b. Prior loss experience
    c. Payment history
4. Risk Inspection data
5. Geography based on insured locations
    a. Real Estate Values
    b. Auto Repair Costs
    c. Jurisdictional Orientation
    d. Demographics
    e. Crime
6. Agency Characteristics

Data for **Claims** Models

1. Policy Characteristics
   a. Exposures
   b. Limits and Deductibles
   c. Coverages and Perils
2. Rates, ILFs, Modification factors, etc.
3. Insured Characteristics
   a. Credit Information
   b. Prior loss experience
   c. Payment history
4. Risk Inspection data
5. Geography based on insured locations
   a. Real Estate Values
   b. Auto Repair Costs
   c. Jurisdictional Orientation
   d. Demographics
   e. Crime
6. Agency Characteristics

1. Claim information
   a. FNOL
   b. Claimant data (Credit info, geography, social data, etc.)
   c. Other participants (insured, doctors, lawyers, witnesses, etc.)
   d. Cause, type of Injury/Damage
   e. Injury or damaged object
   f. Coverage
   g. Loss Description
   h. Loss Location
   i. Date and time of Loss and Report
   j. Weather at time & location of loss
   k. Claim adjuster notes
   l. Prior payment amounts & counts
   m. Timing of prior payments
   n. Case reserve amounts
2. Details from Prior Claims
   a. from same insured
   b. from same claimant
   c. from same location

Claim Models

— Binomial Regression Models —

Subrogation Identification

Fraud Detection

Claim Triage

Severity Models

Example of Binomial Regression

X1: Hours studied	X2: Attempt # on this exam	X3: Used flash cards	Y: Event
425	2nd	yes	fail
450	3rd+	unknown	PASS
175	1st	yes	PASS
200	1st	no	fail
325	2nd	yes	fail
350	1st	yes	fail
150	1st	yes	fail
550	2nd	unknown	PASS
475	1st	yes	PASS
275	2nd	unknown	PASS
175	1st	yes	fail
100	1st	unknown	fail
50	2nd	yes	fail
500	1st	unknown	fail
225	1st	unknown	PASS
400	1st	no	PASS
125	2nd	no	fail
250	1st	yes	fail
300	1st	yes	fail
400	1st	yes	fail

Binomial Regressions are GLMs.

- Binomial Distribution
- Logit, Probit, or CLogLog link functions

Response variable, Y, is specified as the event you want to predict, like Event='PASS'

The predicted values, $\hat{Y}$, represent the probability that the event occurs based on the input variables.

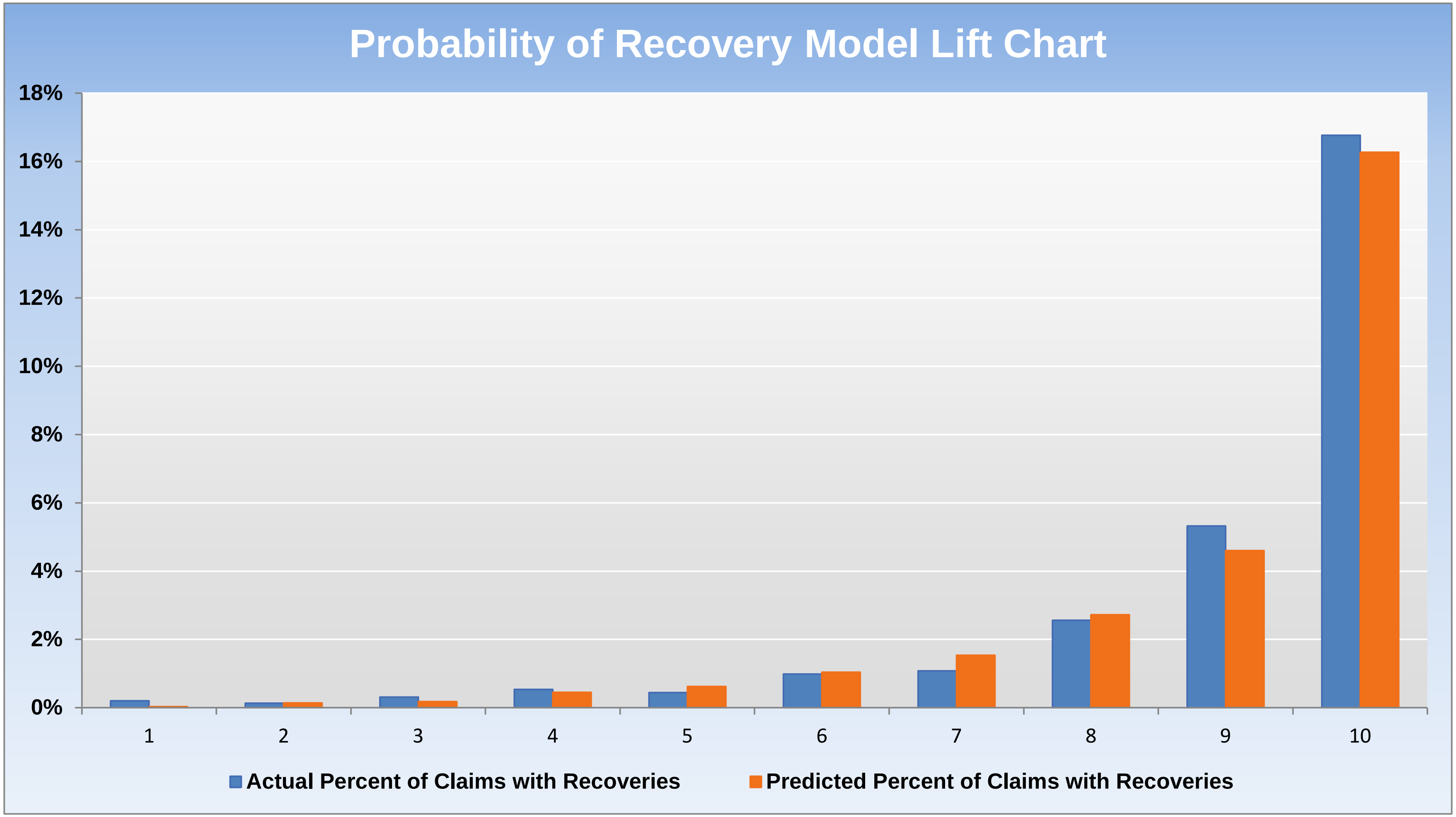
X1: Hours studied	X2: Attempt # on this exam	X3: Used flash cards	$\hat{Y}$: Probability of Passing
300	2nd	No	56.5%
300	2nd	yes	80.1%

Some Milestones in the Life of a Claim

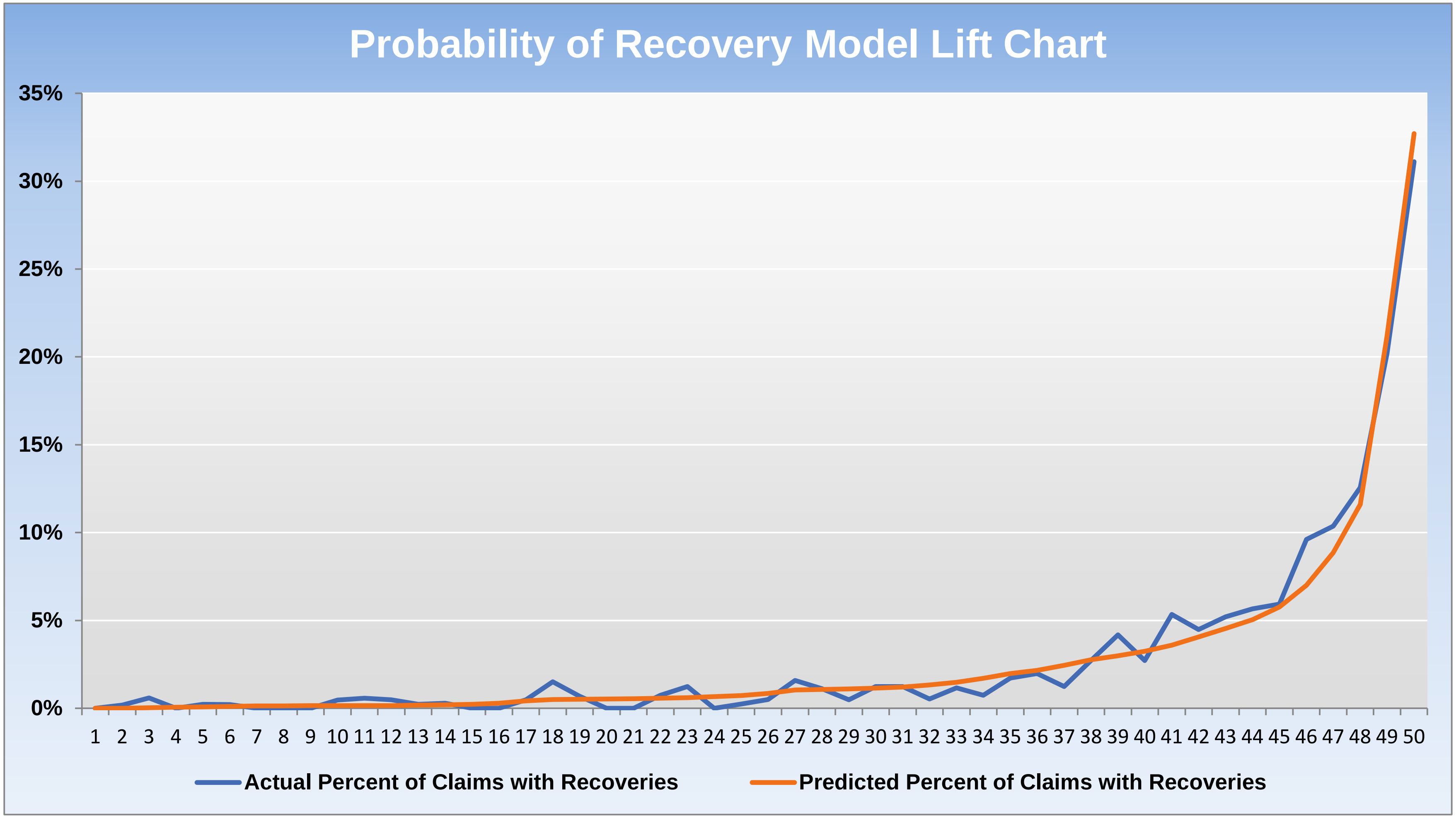
Life of a claim could include:		These could be modeled -
▪ Coverage verification	→	Likelihood to close without payment
▪ Loss exceeds $ threshold	→	Large (or small) loss detection
▪ Claimant sues insurer	→	Litigation propensity
▪ Litigation outcome	→	Likelihood of winning litigation
▪ Claim reopens	→	Probability a claim will reopen
▪ Major adjustment to case reserves	→	Case reserve +\- 20% of ultimate
▪ Suspected to be fraudulent	→	Fraud detection
▪ Suspected to have a recovery	→	Salvage or Subrogation potential
▪ Claim closes	→	Probability claim will close next month

All of the above can be modeled using Binomial Regression.

Probability of Subrogation Recovery Model Performance



Probability of Subrogation Recovery Model Performance



Claim Models

Binomial Regression Models

— **Subrogation Identification** —

Fraud Detection

Claim Triage

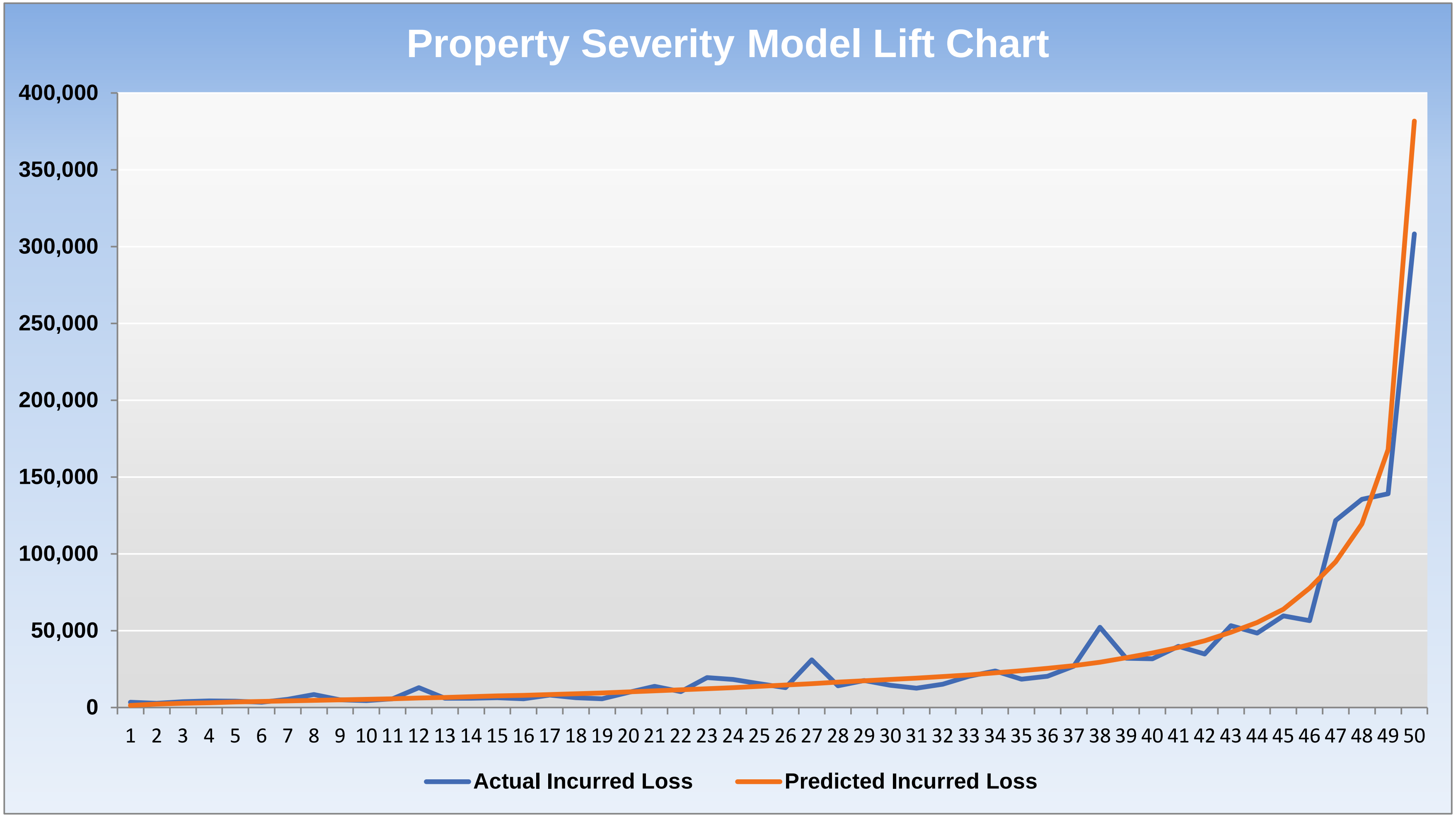
Severity Models

What to model?

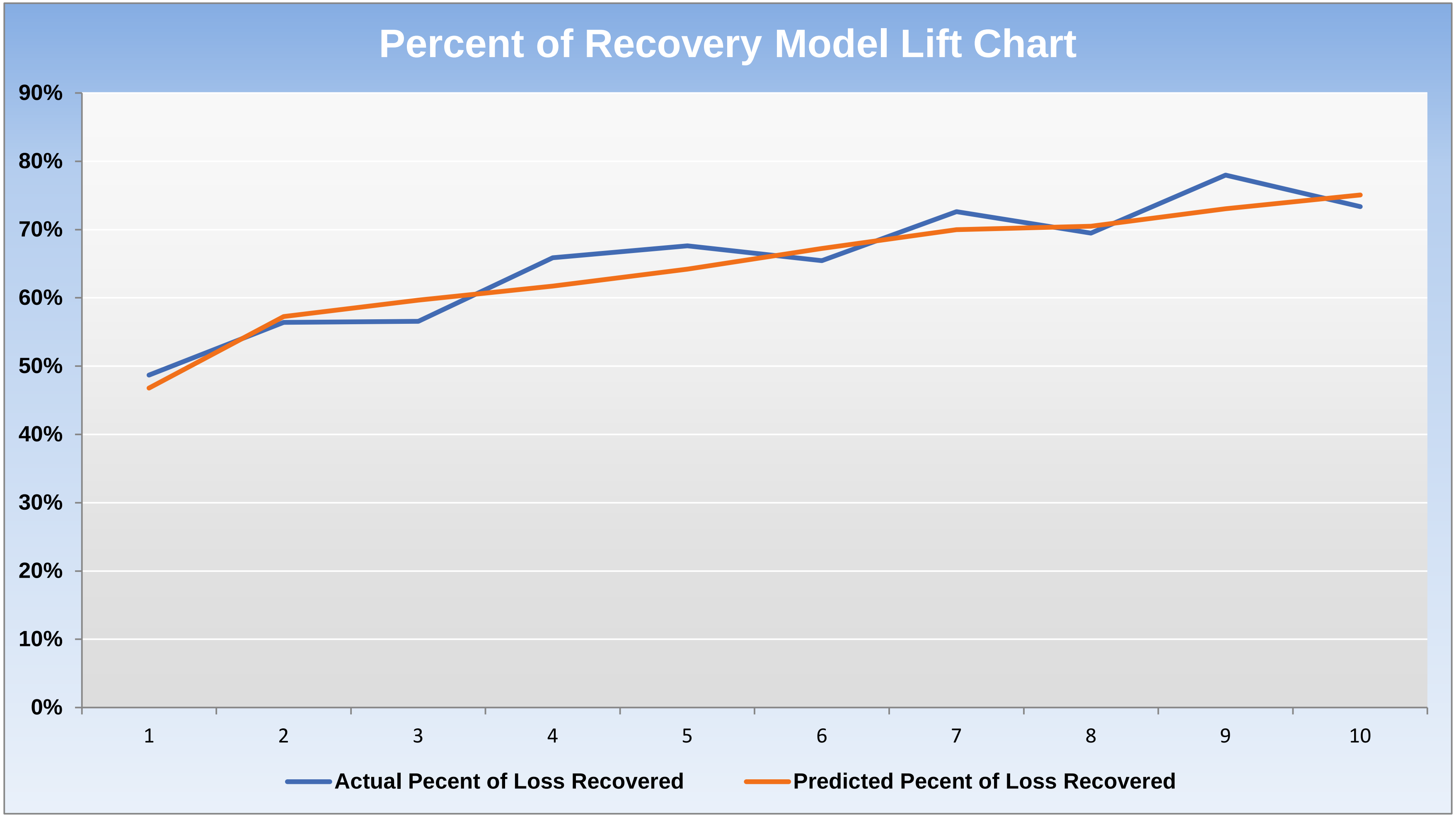
$$E[Recovery] = \Pr[Recovery] * E[Loss] \; * E[\frac{Recovery}{Loss} | (Recovery > 0)]$$

Model	Dependent Variable	Model	Data
Probability of Recovery	$\begin{cases} 1, if\ recovery > 0 \\ 0, otherwise \end{cases}$	Logistic	All
Loss Severity	Ult. Ground-up Loss	Tweedie GLM	All
Percent of Loss Recovered	Gross recovery ÷ Ground-up loss	Fractional Logit	Gross Recovery>0

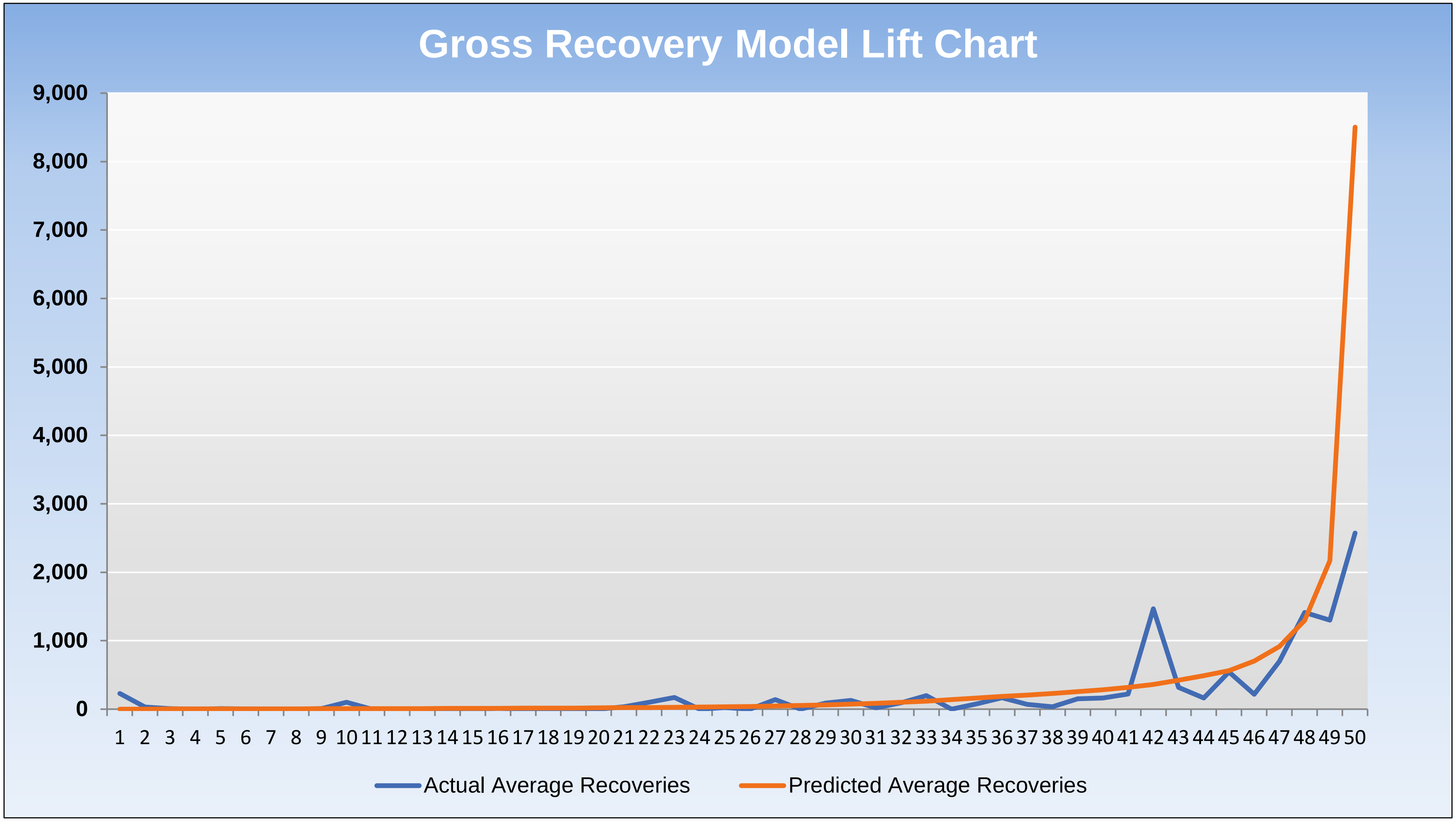
Model Performance – Severity Model



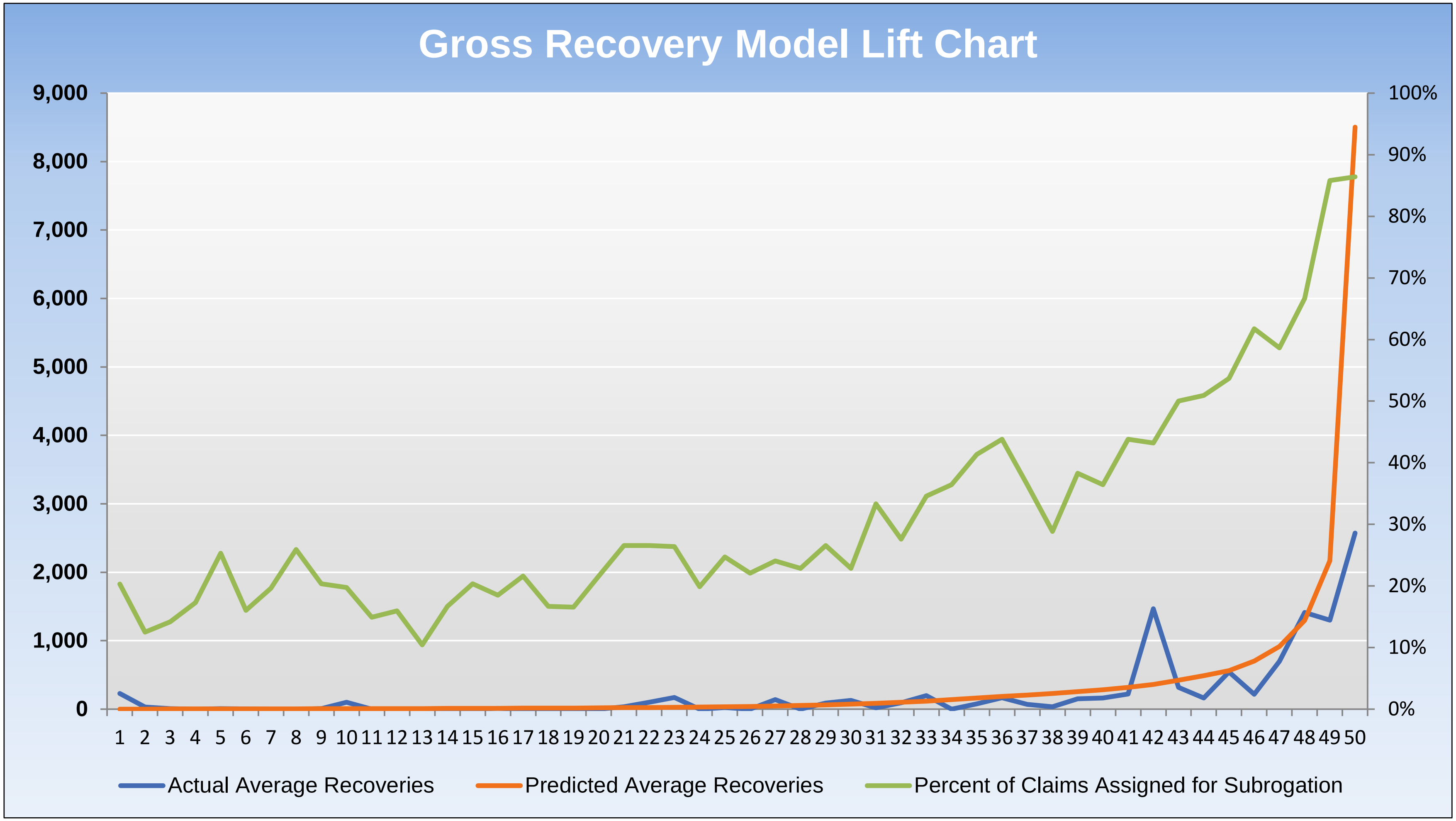
Model Performance – Percent of Loss Recovered



Model Performance – Post Implementation



Model Performance – Post Implementation



Claim Models

Binomial Regression Models

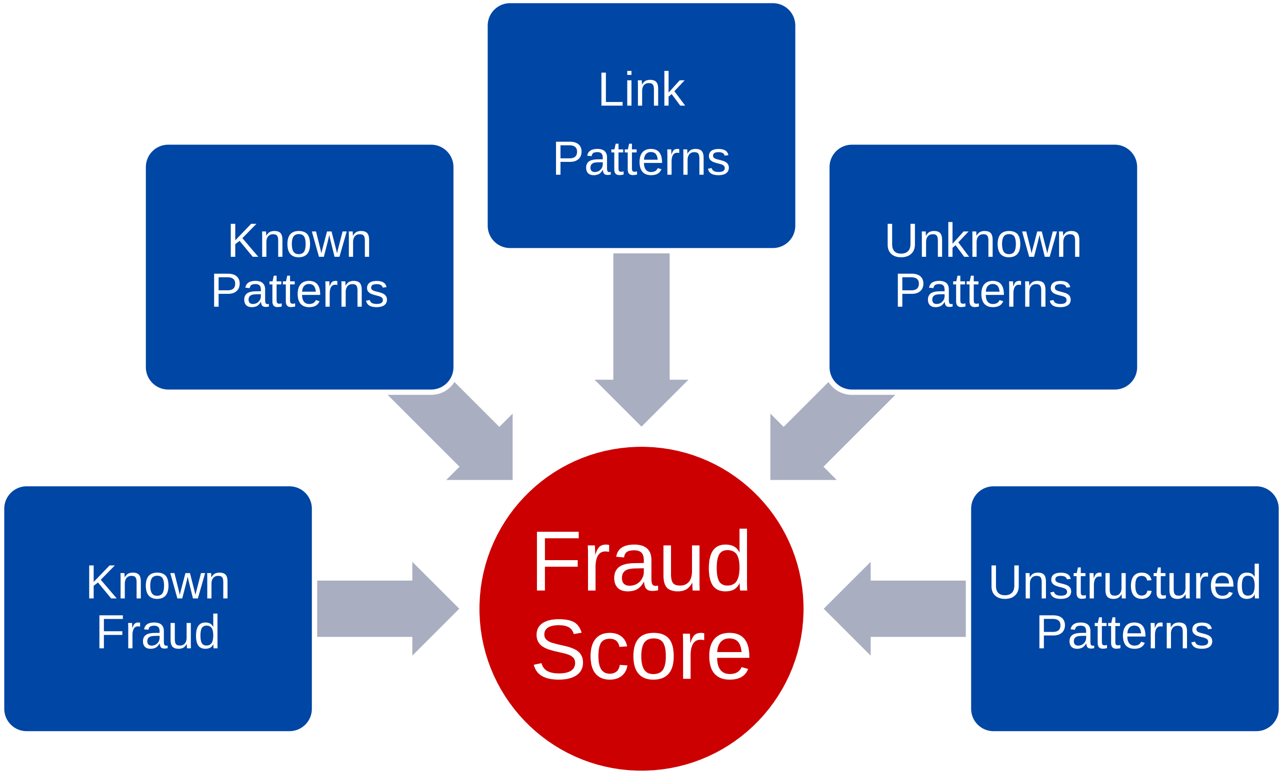
Subrogation Identification

— Fraud Detection —

Claim Triage

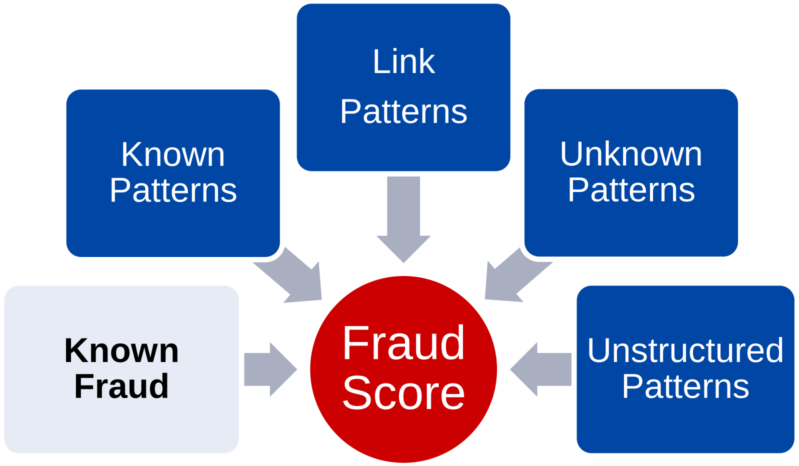
Severity Models

Fraud Detection



There is no guaranteed method for detecting fraud. Using a variety of techniques improves our odds of correctly detecting fraud.

Known Fraud

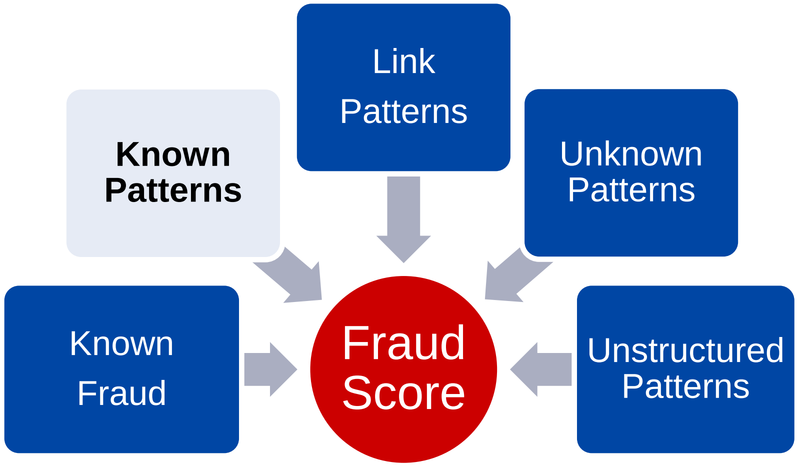


The National Insurance Crime Bureau (NICB) maintains databases of claim participants suspected of engaging in fraudulent activity.

	Claim	NICB	Match Score
Name	Ron MacTosky	Ronald McTofski	63
Address	1760 Gabbro Rd Solon, OH 44124	1762 Gabbro Trail Solon, OH 44124	89
SSN	No data	078-05-1120	-
		Match Score =	**92**

Fuzzy matching can be used to join names, addresses and other personally identifiable information from internal claims to data provided by NICB.

Known Patterns



Claim adjusters are trained to look for "Red Flags" and refer claims with Red Flags to the SIU.

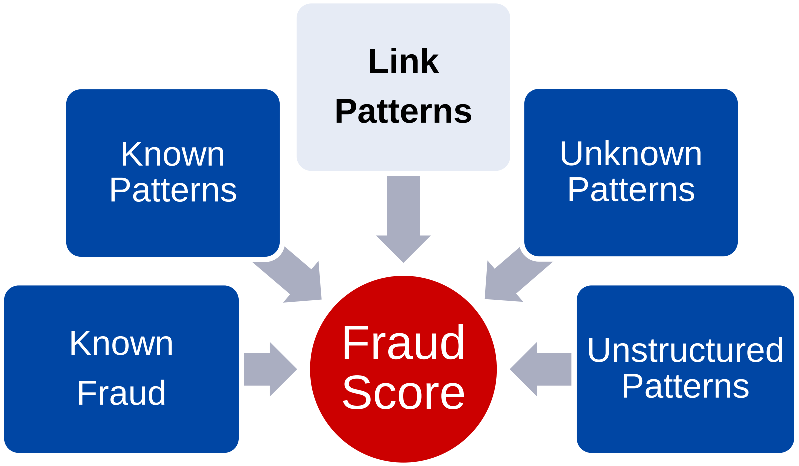
Some example Red Flags:

- Loss occurs a few days after policy was purchased.
- Loss occurred when nobody would be present.
- Loss involves a large amount of cash.
- Insured is experiencing financial difficulties.

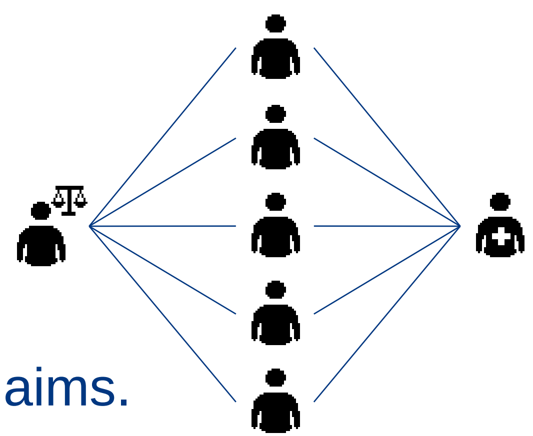
This can be modeled using logistic regression, if you have the data.

Link Patterns

Searches for relationships that link otherwise unrelated claims together based on the name, address, and other personally identifiable information of claim participants.

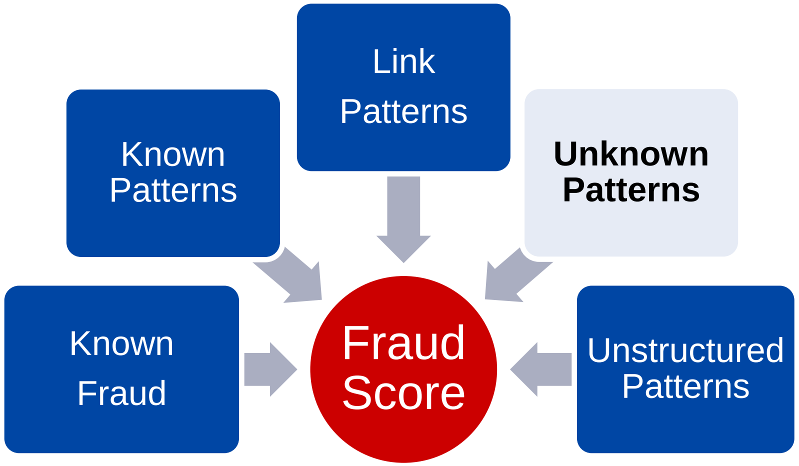


Claims that link to suspicious claims are suspicious by association.



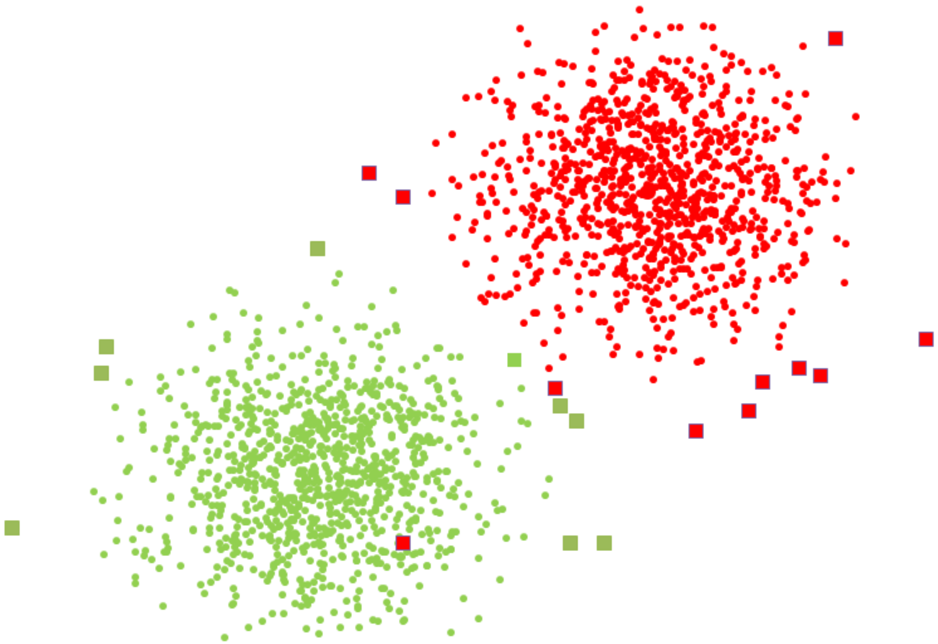
Patterns in links can also implicate claims.

Unknown Patterns

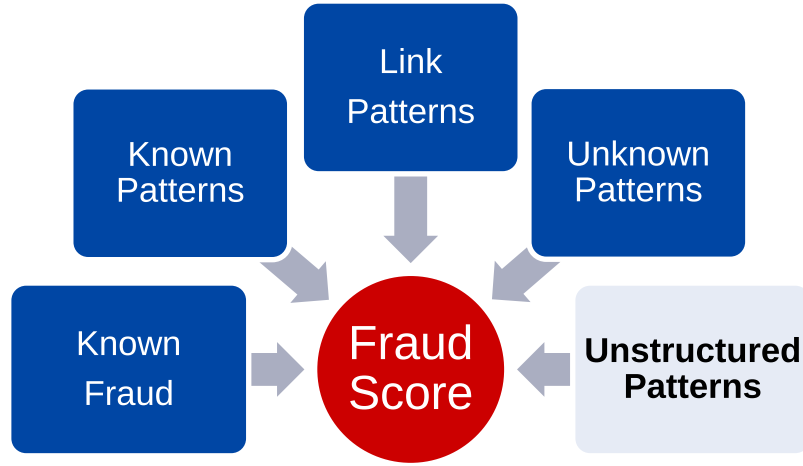


- Outliers are claims with extreme or rare attributes
- Anomalies are claims that don’t conform to usual claim patterns.

Unusual patterns are a sign of suspicious activity, and warrant further attention from the SIU.



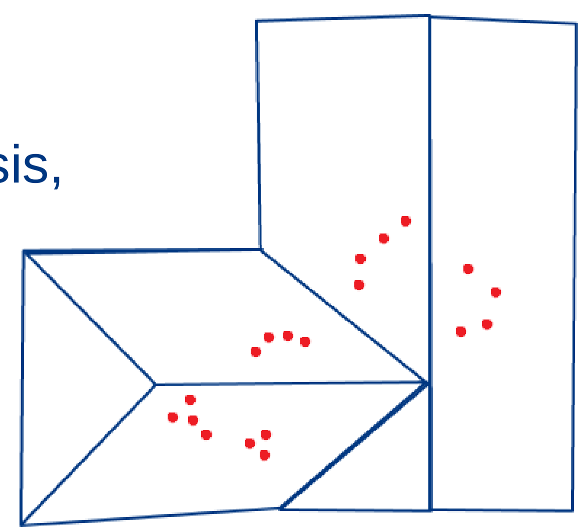
Patterns in Unstructured Data



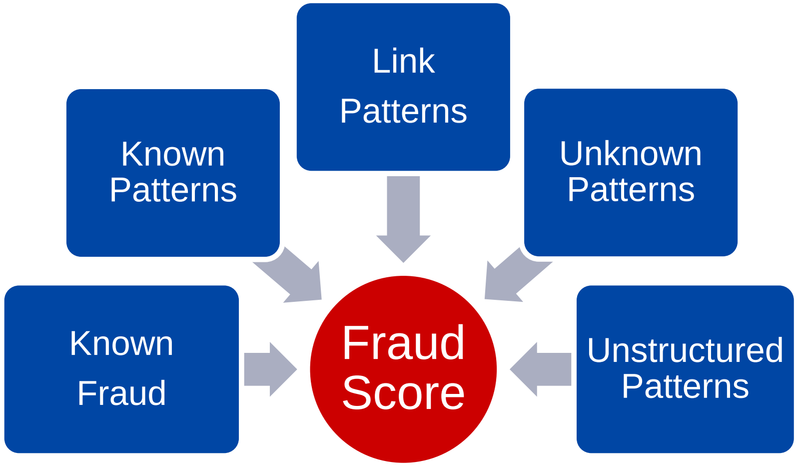
Unstructured claim data includes:

- Claim adjuster notes
- E-mails
- Photographs of damage
- Scans of receipts
- Audio from insurer’s call center and 911 calls
- Video of damage

Text mining, speech recognition, sentiment analysis, object recognition, facial recognition, movement analysis and other machine learning techniques find patterns in unstructured data that can be used to detect fraud.



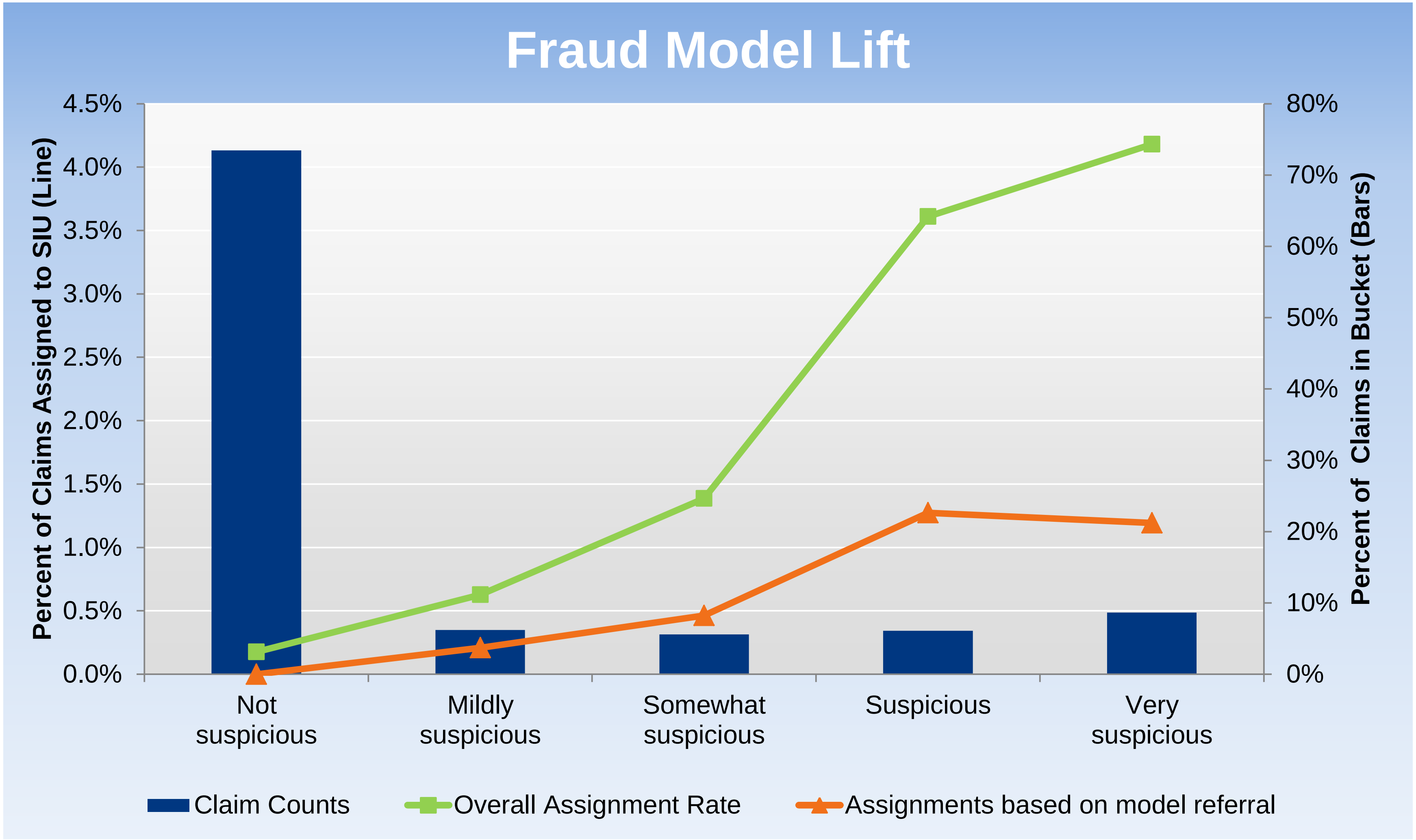
Fraud Scores



Results from all models are brought together to determine which claims need to be referred to the SIU.

Claims are reported to SIU when they:

- score higher that a predefined threshold on a model
- trigger a business rule
- have fraud score that recently increased significantly



Fraud Model Lift
Percent of Claims Assigned to SIU (Line)
4.5%
4.0%
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
Percent of Claims in Bucket (Bars)
80%
70%
60%
50%
40%
30%
20%
10%
0%
Not suspicious
Mildly suspicious
Somewhat suspicious
Suspicious
Very suspicious
Claim Counts
Overall Assignment Rate
Assignments based on model referral

Claim Models

Binomial Regression Models

Subrogation Identification

Fraud Detection

— **Claim Triage** —

Severity Models

Triage Model at Lemonade

"...A.I. Jim, Lemonade's claims bot, reviewed Brandon's claim, cross-referenced it against his policy, ran 18 anti-fraud algorithms on it, approved it, sent wiring instructions to the bank for the transfer of $729..."

"... A.I. Jim is still learning, and he often escalates claims to real Jim."

"Lemonade Sets a New Record." lemonade.com/blog/lemonade-sets-new-world-record. Accessed 4/15/2017.

"The algorithms powering AI Jim 'understand' the nature of claims, their severity, and whether the user is in a state of emergency."

"The Secret Behind Lemonade's Instant Insurance." lemonade.com/blog/secret-behind-lemonades-instant-insurance/. Accessed 4/15/2017.

Triage Models at Other Insurers

"ESIS' predictive modeling services are designed to help improve the outcomes of individual claims by scientifically identifying files with a propensity for severity.

...**Multiple models are run to score claim severity at intake**, and then run again within the first three months to re-score the claim."

"ESIS Advanced Claims Analytics in Action." chubb.com/us-en/business-insurance/esis-advanced-analytics-in-action.aspx. Accessed 4/15/2017.

"...Through the use of sophisticated analytics tools, the employer's insurer identifies this claim's potential for volatility and quickly assigns the claim to an elite team of medical professionals....**The use of predictive analytics to identify the nonobvious factors that can improve claim outcomes is an increasing area of focus for leading insurers**..."

"Predictive Modeling Improves Claim Outcomes While Lowering Costs." thehartford.com/sites/the_hartford/files/Claims-Predictive-Modeling.pdf. Accessed 4/15/2017.

How should we triage?

Assume we want

- claims less than $25,000 assigned to Low Severity Adjusters, and
- claims $25,000 or more assigned to High Severity Adjusters.

How should we model this?

- Model the ultimate severity and direct claims based on E[Loss]
- Model the Probability(Loss<$25,000) and direct claims based on that

What if the loss is distributed as

$$Loss = \begin{cases} \$0, & 80\%\ of\ the\ time \\ \$100{,}000, & 20\%\ of\ the\ time \end{cases}$$

Triage Models

The triage model is composed of multiple models with different purposes:

• Likelihood to close for less than $5,000	Logistic regression
• Likelihood to close for more than $25,000	Logistic regression
• Likelihood to close for more than $250,000	Logistic regression
• Likelihood of more than 5 payments	Logistic regression
• Likelihood to still be open after 60 days	Logistic regression
• Expected value of loss amount	Gamma GLM
• Expected Variance of loss amount	Log-linear variance Model
• Salvage/Subrogation/Litigation/Fraud Models	

Triage Model Performance

Current Process assignment	Proposed Process assignment (the model)	Percent of claims	Percent of claims greater than $25K	Median Severity of claims >$25K
Low Sev	Low Sev	74%	16%	$51,000
High Sev	Low Sev	15%	27%	$49,000
Low Sev	High Sev	8%	45%	$98,000
High Sev	High Sev	3%	73%	$148,000

The Claim Triage Model would assign fewer claims to the more experienced adjusters, and the claims would tend to be larger more often.

Claim Models

Binomial Regression Models

Subrogation Identification

Fraud Detection

Claim Triage

— Severity Models —

GLM Severity Models

If including claims closed without payment,

- Tweedie distribution.

Excluding claims closed without payment,

- Gamma distribution with inverse link
- Inverse Gaussian with inverse squared link
- Normal distribution with log link

Accelerated Failure Time Models

In Survival Analysis, AFT models are able to

- Predict the time from a starting point (entering a study)
  to an event (death),
  and some data may be censored (left the study)

This parallels Severity Modeling, where we want to

- Predict the dollars from a starting point (the deductible)
  to an event (claim closed),
  and some data may be censored (reached a limit or remain open)

AFT appropriately adjusts for

- left truncation from deductibles,
- right censoring from limits and open claims.

Generalized Linear Mixed Models

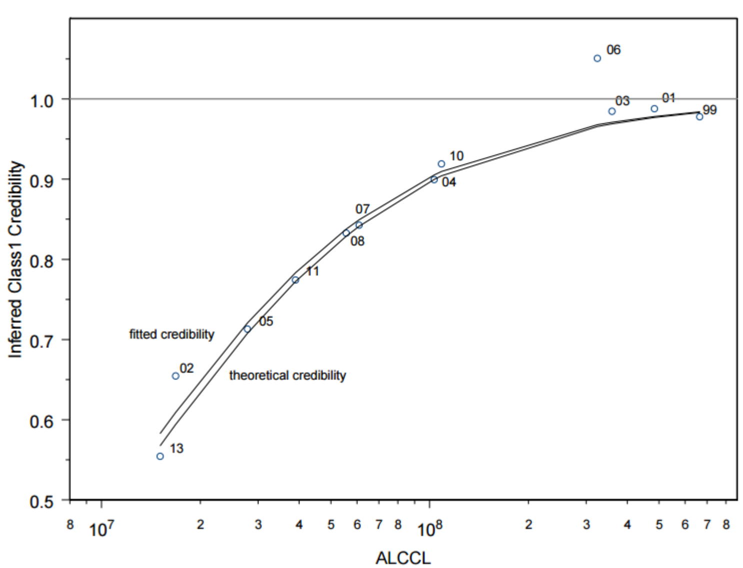
Similar to GLM, but enables specified parameters to be shrunk towards the mean in a manner that very closely mimics Buhlmann-Straub credibility.

“Fixed effects” are treated same as in GLM.

Coefficients for “Random effects” with low credibility are adjusted to be closer to zero.

The algorithms behind GLMM do not “think” in terms of credibility, and sometimes imply more than 100% credibility to a particular subgroup. But there are ways around that.

Inferred Credibility vs. ALCCL



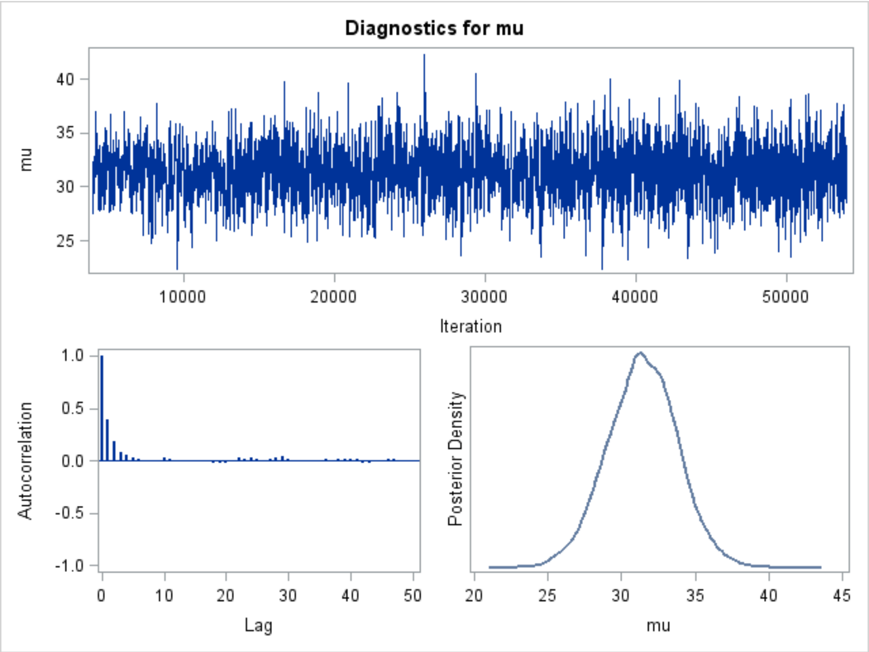
Klinker, F., “Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting” CAS E-Forum, Winter 2011 (Volume 2), p. 17.

Bayesian Models

- Produces joint posterior samples of model parameters which can be used to incorporate parameter risk in prediction confidence intervals.

- "Power priors" can be used to incorporate information from similar historical data when analyzing current data.

- Able to perform
  - o Generalized Linear Models
  - o Generalized Linear Mixed Models
  - o Accelerated failure time models



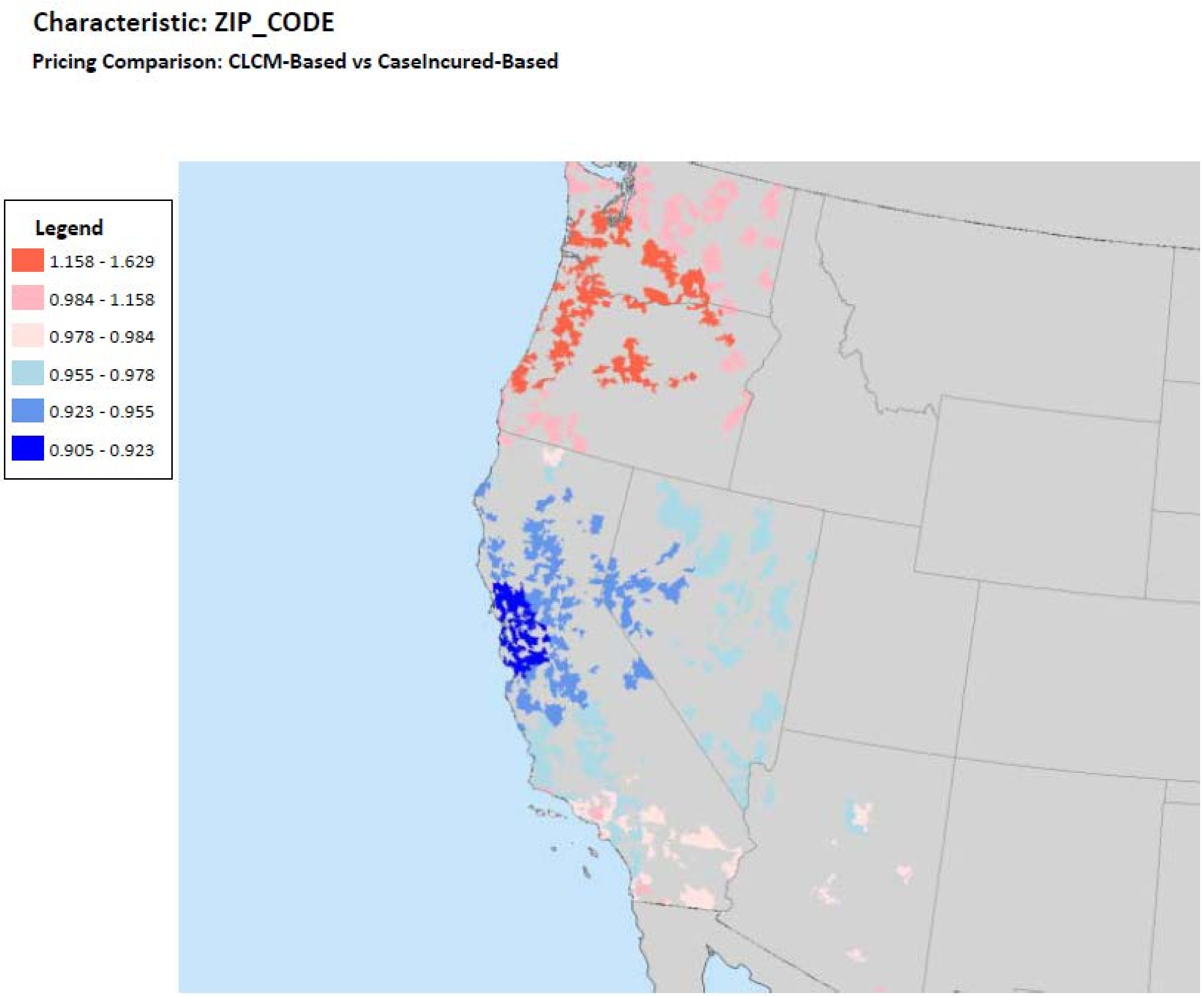
Areas of Potential Benefit Identified by the Working Party

- Claims Management
- **Pricing**
- Actuarial Reserving
- Enterprise Risk Management
- Reinsurance

Pricing

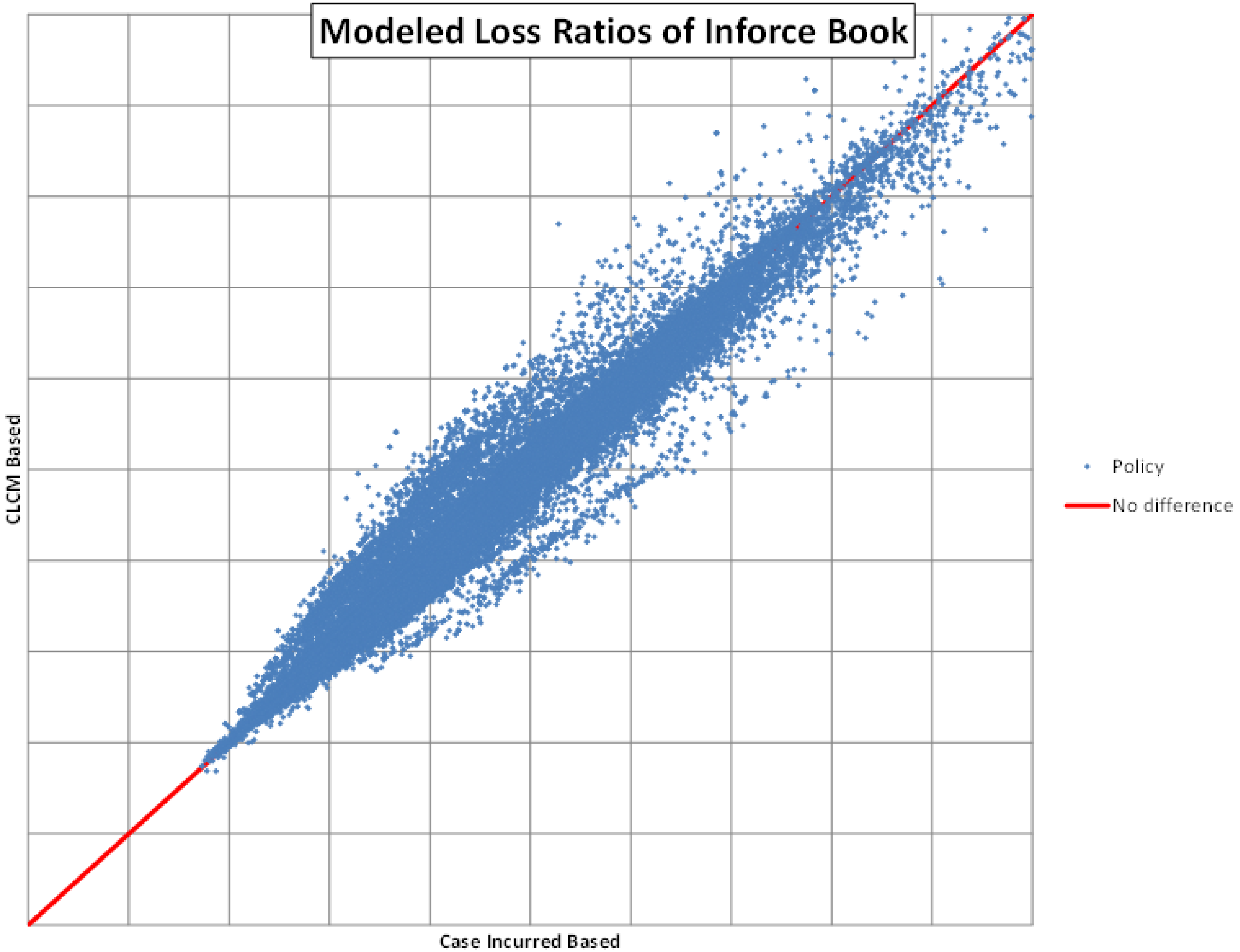
- Current approaches tend to assume that different risks develop the same way.
- More refined models of development using claim and exposure characteristics as predictive variables can have significant impact on pricing indications

Pricing



Characteristic: ZIP_CODE
Pricing Comparison: CLCM-Based vs CaseIncured-Based
Legend
1.158 - 1.629
0.984 - 1.158
0.978 - 0.984
0.955 - 0.978
0.923 - 0.955
0.905 - 0.923

Example 1



Modeled Loss Ratios of Inforce Book
CLCM Based
Case Incurred Based
Policy
No difference

Example 2


Modeled Loss Ratios of Inforce Book
CLCM Based
Case Incurred Based
Policy
No difference

Areas of Potential Benefit Identified by the Working Party

- Claims Management
- Pricing
- **Actuarial Reserving**
- Enterprise Risk Management
- Reinsurance

Actuarial Reserving

- Allocation
- Shock to confirmation bias
- Mix Shifts

Actuarial Reserving

	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10
AY 1	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y
AY 2	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	
AY 3	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y		
AY 4	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y			
AY 5	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y				
AY 6	**X**Y	**X**Y	**X**Y	**X**Y	**X**Y					
AY 7	**X**Y	**X**Y	**X**Y	**X**Y						
AY 8	**X**Y	**X**Y	**X**Y							
AY 9	x**Y**	x**Y**								
AY 10	x**Y**									

Areas of Potential Benefit Identified by the Working Party

- Claims Management
- Pricing
- Actuarial Reserving
- **Enterprise Risk Management**
- Reinsurance

Enterprise Risk Management

- Commonly: Top-down, starting with silos, correlation
- Desirable to be able to pass the “use test” at a transactional level.
- While top-down approaches can certainly be used in this way, they may be overly simplistic
- Building detailed stochastic bottom-up models underlying development and pricing, can add even more value to the process of ERM, particularly when reconciled with the top-down approach

Areas of Potential Benefit Identified by the Working Party

- Claims Management
- Pricing
- Actuarial Reserving
- Enterprise Risk Management
- **Reinsurance**

Reinsurance

- High layer transactions can benefit
  - behavior of losses below the layer
  - potential for entry into the layer
  - impact of limits
  - etc.
- Loss Portfolio – same as actuarial reserve benefits
- Proportional/Quota Share
  - Better understanding of a shifting book
- Challenge getting data from cedants

A unified, sophisticated and continuously updated model that covers not only the claim generation but also the entire claim “lives” that could underlie both pricing and reserving and provide useful inputs for claim departments.

Monte Carlo Simulation

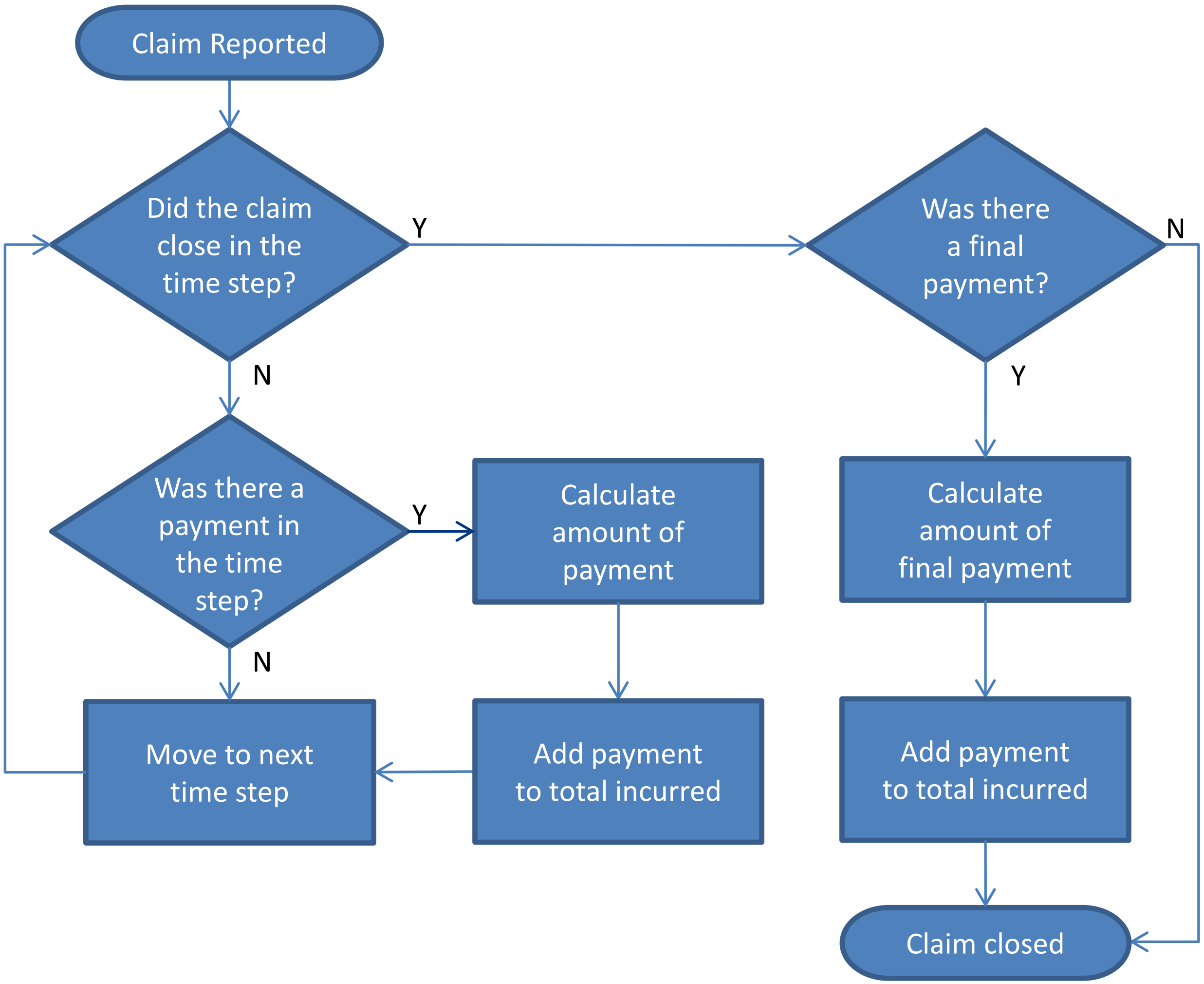
"Monte Carlo simulation is now a much-used scientific tool for problems that are analytically intractable and for which experimentation is too time-consuming, costly, or impractical."

"Introduction To Monte Carlo Simulation." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2924739. Accessed 4/15/2017.

Monte Carlo simulation is used in

- climate forecasting
- computational biology
- engineering
- manufacturing
- physical sciences
- project management
- research and development
- transportation
- ... and numerous other industries, including insurance.

Simple Example for a Claim Life Cycle Model



Parameterization and Projection

- Each component model can be parameterized by any number of predictive modeling techniques
- Use the data underlying all triangle cells
- Age of the claim could be treated as a variable or separate models by age could be built
- Simulation is a useful approach to integrating the model projections out to a distribution of ultimate projections for all open claims

Simulation example

		Time Step									
Claim #	Sim.	1	2	3	4	5	6	7	8	9	10
001	1	$C									
001	2	-	$	C							
001	3	C									
001	4	$	$	C							
002	1	$	-	-	$	-	$	C			
002	2	-	-	$	$	$	$	$	$	$C	
002	3	$	$	$	-	-	-	-	$C		
002	4	$	-	-	$C						

$ = simulated payment. C = simulated claim closed.

Simple Example for Emergence/True IBNR