A Complementary Approach for Product Management and Book of Business Segmentation:

Turning Data into Knowledge

Milliman

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.

Technology Focus

- Focused on building Business Solutions
    - Application specific products
    - Not limited to project based engagements
    - Looking for repeatable business problem/business solutions
- Segment focus
    - Personal lines
    - Workers Comp

Milliman

Team's Business Experience

- Predictive Modeling based software business
  - Supplier Performance Management Application
  - Worked with Fortune 500 Manufacturing Companies
  - Aggregated data from Manufacturer's and 3rd party data (D&B)
  - After 9/11 event aerospace industry slowed down
    - Large number of small businesses went bankrupt
    - Clients came to us asking if we could use data to predict negative financial outcome
  - Successfully built and deployed Financial Stress Score

  - Company acquired by D&B

Milliman

Machine Learning

- The most exciting phrase to hear in science, the one that heralds new discoveries, is not “Eureka” but “That’s funny...”

—Isaac Asimov (1920–1992)

Milliman

Machine Learning

- Complementary to more traditional actuarial approaches
  - Observes/identifies patterns in data
  - Determines accuracy/repeatability of patterns
  - Can be developed to recalibrate based on predicted versus actual outcomes
- No such thing as “Bad Data”
  - Just Useful and Useless Data
  - The more data the better
  - More sources the better
- Lowest level detail even better

Machine Learning and Regularization

- New approach to predictive modeling
- Bringing analysis to the data (as opposed to bringing the data to the analysis)
- Less emphasis on "hypothesis": enabled by the use of Regularization in the predictive algorithms
- Regularization prevents over-fitting and the negative effects of multiple multi-collinearity.
- Mathematically proven to result in better predictive performance on yet-unseen data (future cases not included in the training set)
- Allows jumping into predictive modeling without lengthy upfront investment to ensure that the "right" set of predictive variables and training set instances are used

Regularized predictive algorithms

- Let x denote the vector of input variables and y the predicted
- Given a set of observations (x_i, y_i) we seek an estimator function, f, such that $f(x)$ is a good estimator of y
- Regularized Least Squares:

$$\min_f \frac{1}{l}\sum_{i=1}^{l}(y_i - f(x_i))^2 + \lambda \|f\|_K^2$$

- Similar expressions for different algorithms (e.g., Logistic Regression, Decision Trees, Neural Networks etc.)
- Vapnik & Chervonenkis proved that Regularized algorithms prevent overfitting of the training data and yield better performance on yet unseen data.

Machine Learning

- Outline
- Examples
- Q&A

Example - Homeowners

Data Set

- Approximately 400,000 Homes
  - 300K – training set
  - 100K – test set
- National coverage
- 5 years of data
- Non -CAT

Identify top factors driving losses

- Book's performance had been in decline
- Client needed results to be useful and manageable from an underwriting perspective
  - 100 factors too many
  - 1 factor too few
- Client requested 3 factors

Approach

- Built model to identify factors correlating to losses
- Factors observed included traditional/expected variables
  - Location
  - Construction type
  - Etc.
- Model also identified unexpected nonlinearities

5 Segments:

Segment	Var 1	Var 2	Var 3	Machine Learning Score	Count of Instances	Loss Ratio 2010
1		Low		0.231	5857	0.313
2		Hi	Low	0.405	5347	0.353
3	Hi	Hi	Hi	0.487	22903	0.433
4	Low	Hi	Hi	0.549	12718	0.450
5		Hi	Med	0.583	14795	0.466

Milliman

Top 3 Variables

- Identified 3 variables that were not well represented in previous underwriting models
- These variables consistently correlated to losses
- Due to restrictions will only discuss one of 3 variables

Milliman

Variable #3 – Age of Home

- Observed “Non-Linear” results
- Homes of different ages had losses that did not consistently correspond to their age
- Further examination indicated that location and age was consistent predictor of loss
- Client confirmed that they had done studies related to building code enforcement that aligned with results

Loss Ratio Lift: 1.5x

Milliman

Example - Workers Comp

Data Set

- Approximately 400,000 Homes
  - 300K – training set
  - 100K – test set
- National coverage
- 5 years of data
- Non -CAT

Return to Work Studies

The Menninger Foundation – "Window of Suggestibility"

Study findings strongly suggest that early intervention is a variable that can make a major difference in outcomes.

- Personality characteristics (especially those relating to independence) *begin to change 60 days after injury.*

"Window of Suggestibility"

60

DAYS

PIE principles - Military combat stress reaction (CSR)

- Proximity - treat the casualties close to the front and within sound of the fighting
- Immediacy - treat them without delay and not wait till the wounded were all dealt with
- Expectancy - ensure that everyone had the expectation of their return to the front after a rest and replenishment

Analytics in Action

Predictive Modeling:

workers™ comp CLAIM DNA

Data Analytics:

NationalWorkersComp Clearinghouse®

Business Challenges

- Talent Crisis
- Defusing Exploding Claims

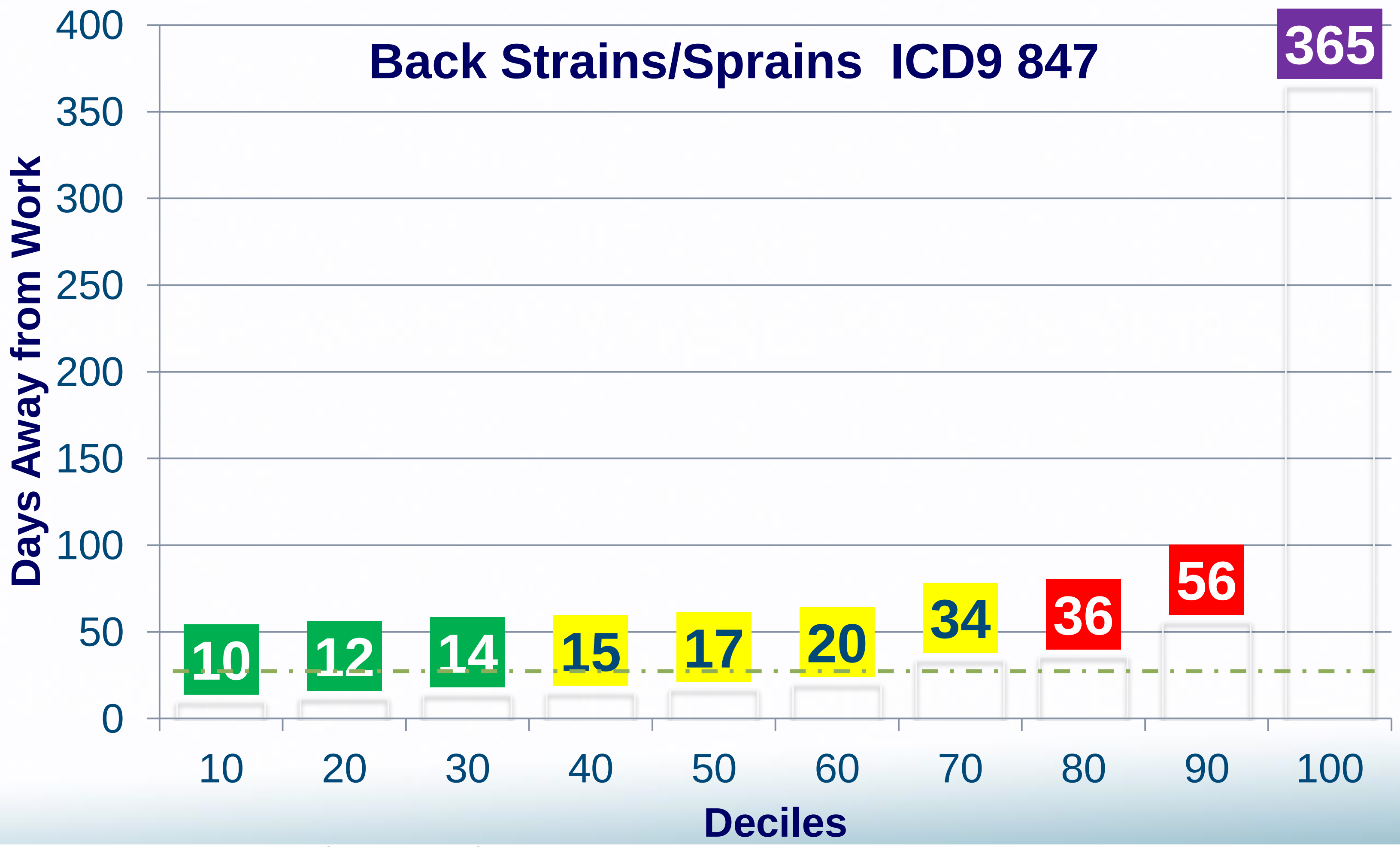
Achieve Better Outcomes

- Accurate Projections
- Over Exaggerated Claims

20

RTW Claims Data

Retrospective



Source: ODG WorkLossData Institute

Claim Triage/Claim Indicators

Three employees – same employer – same diagnosis
ICD9: 847.2

TRIAGE

Isabella	Ethan	Jacob
•Age 37	•Age 27	•Age 51
•Female	•Male	•Male
•Divorced	•Single	•Married
•Three Children	•2ndShift/USW	•One Child
•Office	•Lift Truck Driver	•3rd Shift
•Family Doctor	•Chiropractor Tx	•Emergency Room
•Return to Work	•Out of Work	•Return to Work

Claim Triage/Claim Indicators

Three employees – same employer – same diagnosis
ICD9: 847.2

TRIAGE

Isabella	Ethan	Jacob
•Age 37	•Age 27	•Age 51
•Female	•Male	•Male
•Divorced	•Single	•Married
•Three Children	•2ndShift/USW	•One Child
•Office	•Lift Truck Driver	•3rd Shift
•Family Doctor	•Chiropractor Tx	•Emergency Room
•Return to Work	•Out of Work	•Return to Work

Claim Triage/Claim Indicators

Three employees – same employer – same diagnosis
ICD9: 847.2

	Isabella	Ethan	Jacob
TRIAGE	•Age 37 •Female •Divorced •Three Children •Office •Family Doctor •Return to Work	•Age 27 •Male •Single •2ndShift/USW •Lift Truck Driver •Chiropractor Tx •Out of Work	•Age 51 •Male •Married •One Child •3rd Shift •Emergency Room •Return to Work
HIDDEN	• 30 mile commute •(+) MD TX patterns •(+) Claim experience •Rx – NSAIDs		

Claim Triage/Claim Indicators

Three employees – same employer – same diagnosis
ICD9: 847.2

	Isabella	Ethan	Jacob
TRIAGE	•Age 37 •Female •Divorced •Three Children •Office •Family Doctor •Return to Work	•Age 27 •Male •Single •2ndShift/USW •Lift Truck Driver •Chiropractor Tx •Out of Work	•Age 51 •Male •Married •One Child •3rd Shift •Emergency Room •Return to Work
HIDDEN	• 30 mile commute •(+) MD TX patterns •(+) Claim experience •Rx – NSAIDs	•5 mile commute •Lives alone •Co-Morbid 1: Smoke •(-)Claim filing zip code •(+) Chiro TX patterns •Rx- none	

Claim Triage/Claim Indicators

Three employees – same employer – same diagnosis
ICD9: 847.2

	Isabella	Ethan	Jacob
TRIAGE	•Age 37 •Female •Divorced •Three Children •Office •Family Doctor •Return to Work	•Age 27 •Male •Single •2ndShift/USW •Lift Truck Driver •Chiropractor Tx •Out of Work	•Age 51 •Male •Married •One Child •3rd Shift •Emergency Room •Return to Work
HIDDEN	• 30 mile commute •(+) MD TX patterns •(+) Claim experience •Rx – NSAIDs	•5 mile commute •Lives alone •Co-Morbid 1: Smoke •(-)Claim filing zip code •(+) Chiro TX patterns •Rx- none	•20 mile commute •Co-Morbid 1: BMI +2 •Co-Morbid 2: Smoke •2nd Injury within 3 years •Stay at Home Spouse •College-aged Child •Rx – Percocet (MD dispensed

Claim Triage/Claim Indicators

Three employees – same employer – same diagnosis
ICD9: 847.2

	Isabella	Ethan	Jacob
TRIAGE	•Age 37 •Female •Divorced •Three Children •Office •Family Doctor •Return to Work	•Age 27 •Male •Single •2ndShift/USW •Lift Truck Driver •Chiropractor Tx •Out of Work	•Age 51 •Male •Married •One Child •3rd Shift •Emergency Room •Return to Work
HIDDEN	• 30 mile commute •(+) MD TX patterns •(+) Claim experience •Rx – NSAIDs	•5 mile commute •Lives alone •Co-Morbid 1: Smoke •(-)Claim filing zip code •(+) Chiro TX patterns •Rx- none	•20 mile commute •Co-Morbid 1: BMI +2 •Co-Morbid 2: Smoke •2nd Injury within 3 years •Stay at Home Spouse •College-aged Child •Rx – Percocet (MD dispensed

Data Collection

- Varied Insurance Carrier Claim Systems
- Legacy vs Home Grown vs 3rd Party Vendor
  - State Fund – 328 elements
  - State Fund – 671 elements
  - Carrier - 1401 elements
  - TPA - 514 elements
  - IAIABC FROI/SROI Release 1 - 64 elements
  - IAIABC FROI/SROI Release 3 - 254 elements

Body Mass Index

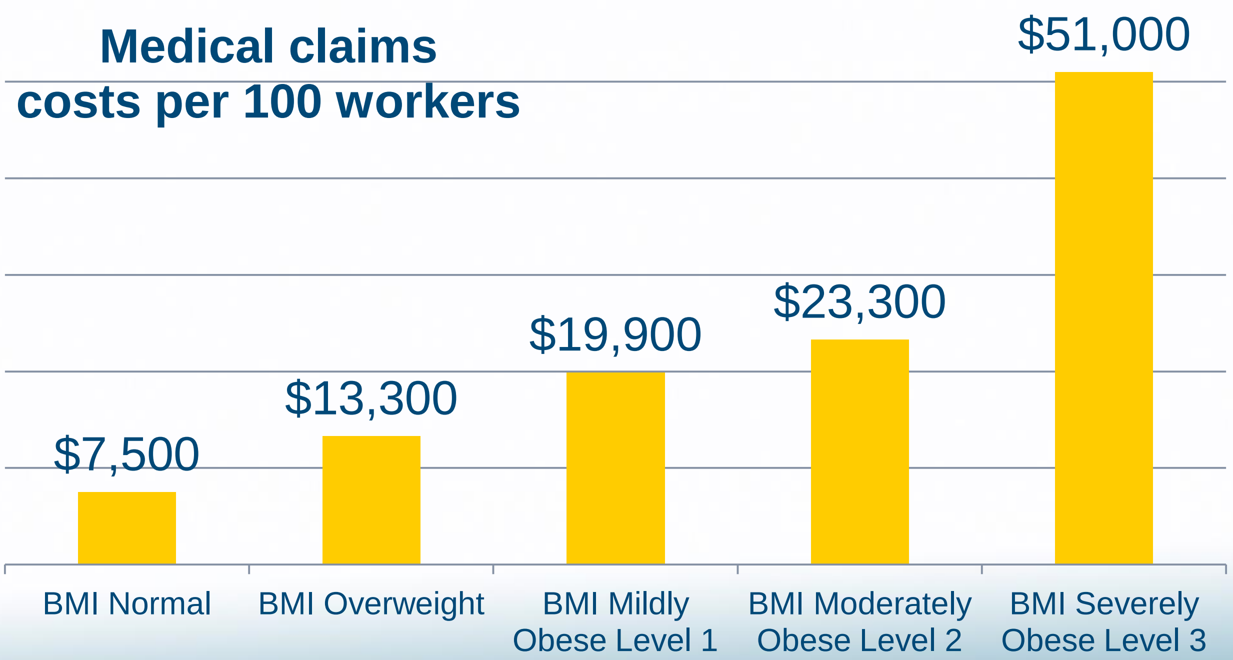
$$\text{Weight} \times 703 \div \text{Height}^2 = \text{BMI}$$

BMI Classification

- •18.5 or less Underweight
- •18.5 to 24.99 Normal Weight
- •25 to 29.99 Overweight
- •30 to 34.99 Obesity (Class 1)
- •35 to 39.99 Obesity (Class 2)
- •40 or greater Morbid Obesity

Work comp medical claims costs rose with injured workers' BMI

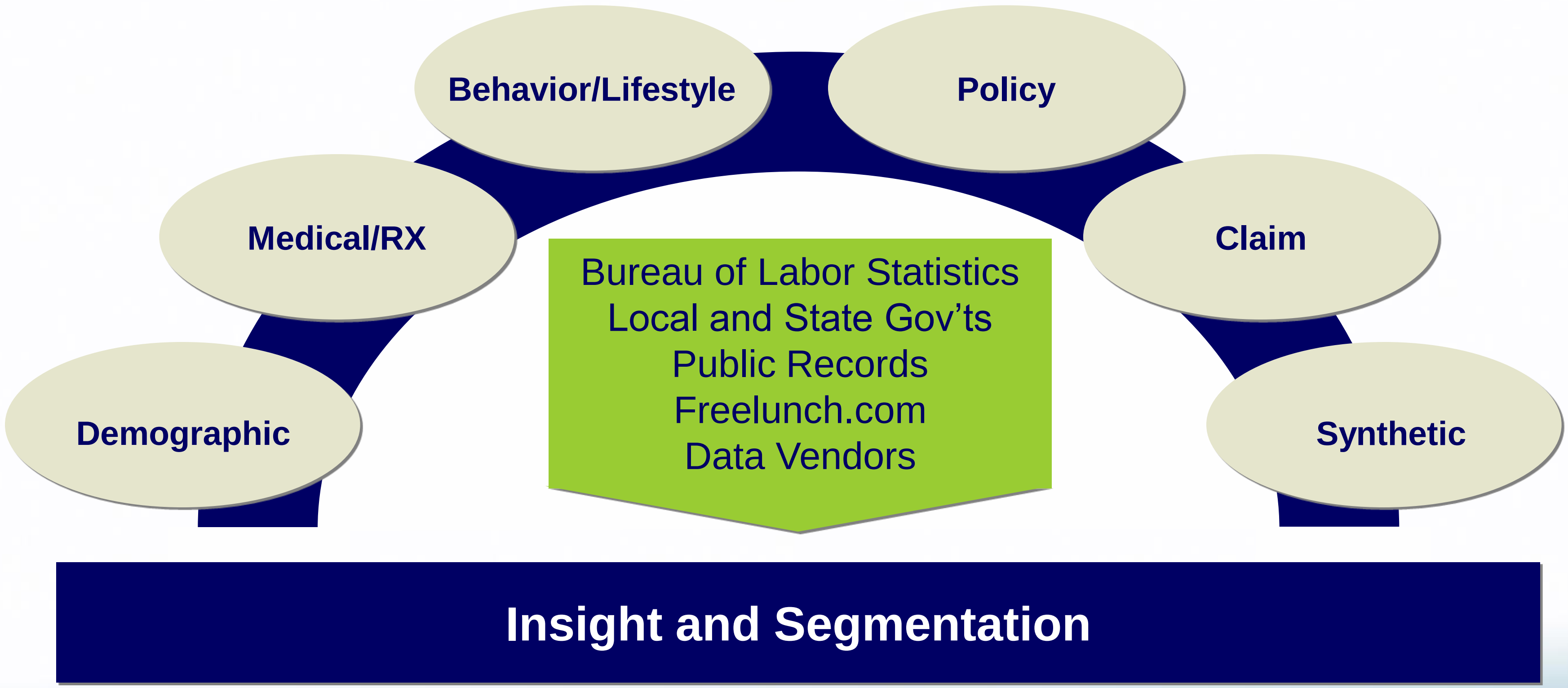
The Duke study indicated nearly six workers' comp claims were filed per 100 workers of normal range BMI, compared with more than 11 claims filed per 100 of the heaviest workers.



29

Data Types and Sources

Distinct and disparate 3rd party data sets provide "lift" and segmentation.



Text Mining

- Text mining refers to the process of deriving relevant and usable text that can be parsed and codified into a word or numerical value.
- Text mining can identify co-morbid conditions and/situations that will have profound impact on the outcome of a claim.

SAMPLE KEY WORDS/PHRASES
▪Diabetes/insulin/injections
▪Packs day/coughing
▪Pain killers/anti-depression
▪Children/school
▪Pain unchanged
▪Home Alone
▪Homemaker wife went to work
▪c/o, CXR, FB, FX
▪CBT – Cognitive Behavior Therapy
▪SNOMED

Text sources: Adjuster notes, medical reports, independent medical exams, etc.

Other Variables of Interest

Specific variables of interest can be based upon recent WCRI Benchmark Studies for medical and prescription cost.

MEDICAL

- Number of visits / claim
  - By specialty
- Number of services/visit
  - By specialty
- Number of physical/occupational therapist visits
  - By specialty
- Number of MRI's within 28 Days from DOI for ICD9 847
- Number of hospital visits for ICD9 847

DRUG TESTING

CPT codes:
- 80100
- 80101
- 80101QW
- G0430
- G0430QW
- G0431
- G0431QW

PHARMACY

- By dispensing point:
  - pharmacy
  - physicians office
- By therapeutic class of drug
  - pain medications
  - gastrointestinal agents
  - sleep inducing, antidepressants and anti-anxiety medications
  - anti-infective
- By generic or brand name
- Average # of pills per claim per prescription
- Average # of prescriptions per claim with prescriptions
- Average # of visits to a dispensing point
- Average # of prescriptions filled per visit
- Average # of pills per prescription

"RED" Flags as potential synthetic variables

Every service company provides RED flags as a way to garner referral business.

Claimant

1. Number of days worked and amount of salary inconsistent with occupation;
2. Injured worker disputes average weekly wage due to additional income (i.e., per diem and/or 1099 income);
3. Cross-outs, white-outs and erasures on documents;
4. Injured worker files for benefits in a state other than principle location of the alleged industrial injury or occupational disease;
5. Injured worker-listed occupation is inconsistent with employer's stated business;

http://www.ohiobwc.com/basics/guidedtour/generalinfo/empgeneralinfo22.asp

Medical Provider

1. Injured worker does not recall having received the billed service;
2. Provider's medical reports read almost identically even though they are for different patients with different conditions;
3. Much higher health-care costs than expected for the allowed injury type;
4. Frequency of treatments or duration of treatment period is greater than expected for allowed injury type, especially for older (non-catastrophic) claims;
5. Frequent billing in older (non-catastrophic injury) claims;

Attorney

1. Representation letter received within a few days of the incident.
2. Attorney consistently deals with same medical providers.
3. Attorney consistently willing to compromise for low dollar amounts.
4. Attorney is single practitioner with offices in several cities.
5. First notice of claim comes from attorney or medical clinic

http://www.untied.com/feature/redflag.pdf

Chronic Pain

1. Continued pain or increased pain 3 months post injury
2. Injured Worker referred to a Pain Management Program
3. Injured Worker referred for spine surgery
4. Injured Worker has seen 2 or more care providers for same diagnosis or symptoms
5. Pain mediation is prescribed by more than one medical provider

Case Management Associates, Inc.

Overview of Constructing a Predictive Model

Predictive Variables → Target Variable

DATA

Train | Test | Validate

Predictive Model $y= b_0 + b_1(x_1)+b_n(x_i)$

Commonly referred to as a "Scoring engine" to estimate the unknown value y based on known values (x_i).

Types of Models:

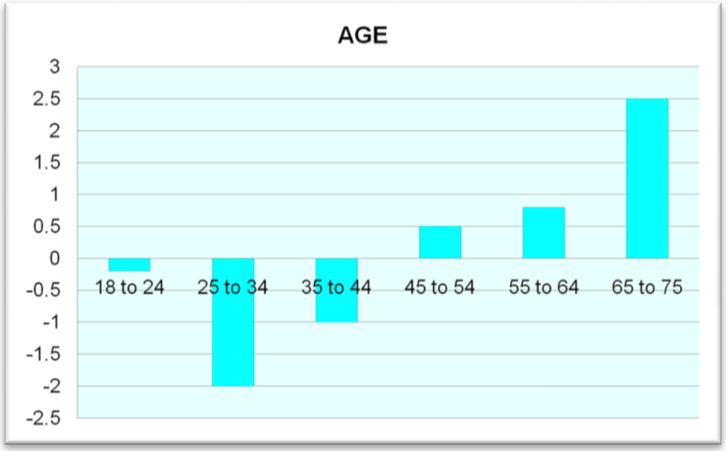
- Linear regression models
- Time series models
- Classification and regression trees
- Neural networks

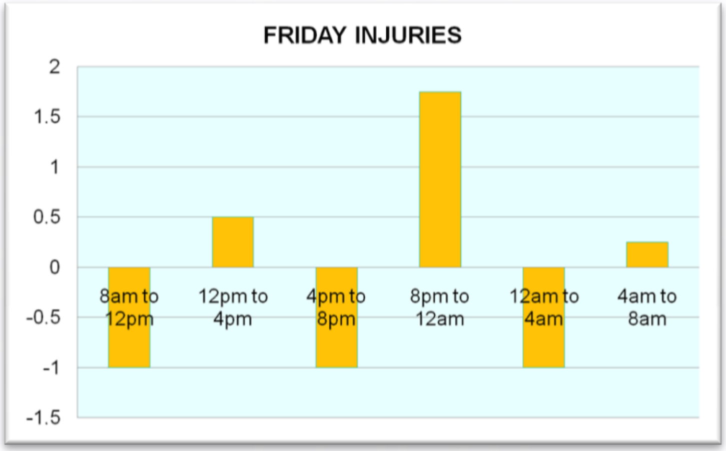
Milliman
34

Univariates

A univariate is an exercise that allows comparison of one variable against a targeted outcome. The strongest are selected for use in modeling.

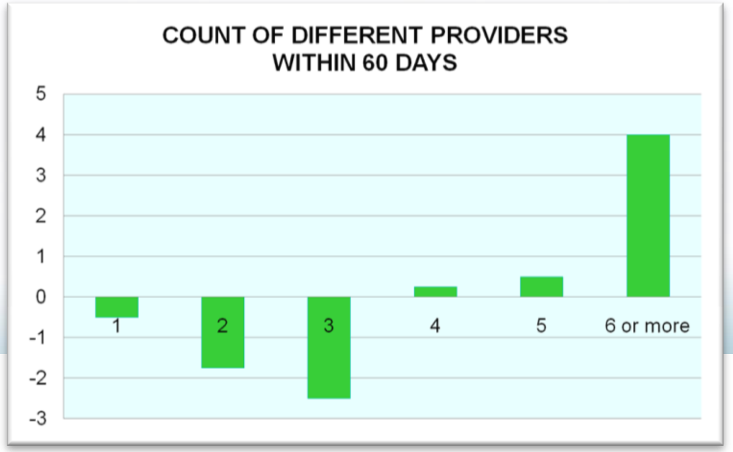
Demographic

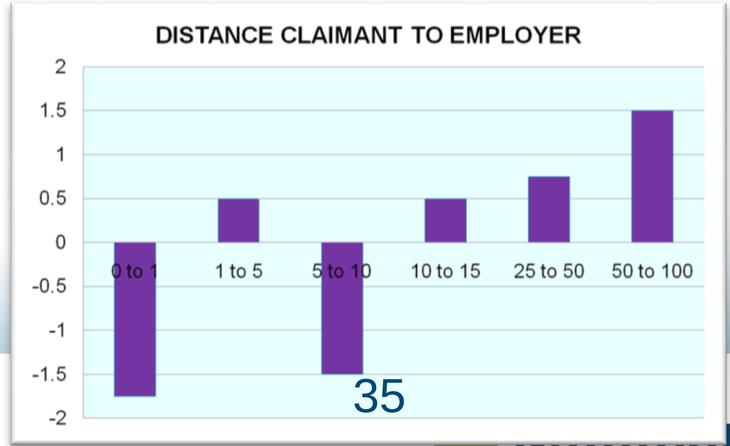




Sample univariates for demonstration purposes

Synthetic





35

Univariate to Multivariate

Univariate

$$y = b_0 x$$

Claim Variables

1. Age
2. Gender
3. Date of Injury
4. Time of Injury
5. Treating Physician
6. Rx
7. ICD-9/10

External Variables

1. US Census
   - Income by Zip Code
2. Claim History by GIS Code
3. Employment by GIS Code

Synthetic Variables

1. Employee Distance to:
   - Employer
   - Physician
   - Attorney
2. Physician Changes

Other Variables

1. Clearinghouse
- National WC Claims DB
- Millions of Claims
- Multiple Industries
- Groupers

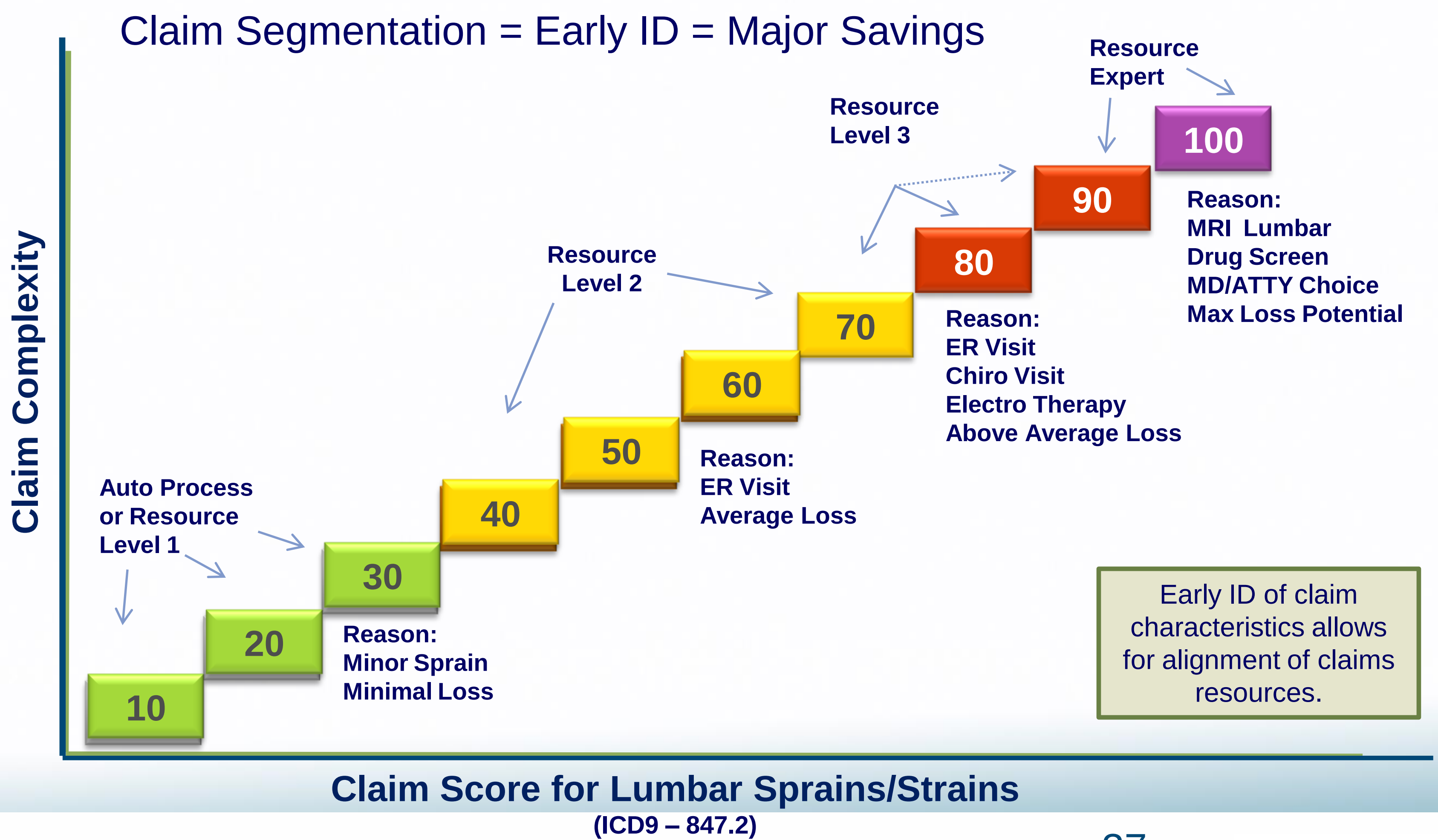
Multivariate

$$y = b_0 + b_1(\mathrm{age})_1 + b_2(\mathrm{dist})_2 + b_3(\mathrm{ICD9})_3 + b_4(\#\mathrm{Rx})_4 ..... + b_n(\mathrm{Variable})_n =$$

(select the 50 to 75 strongest)

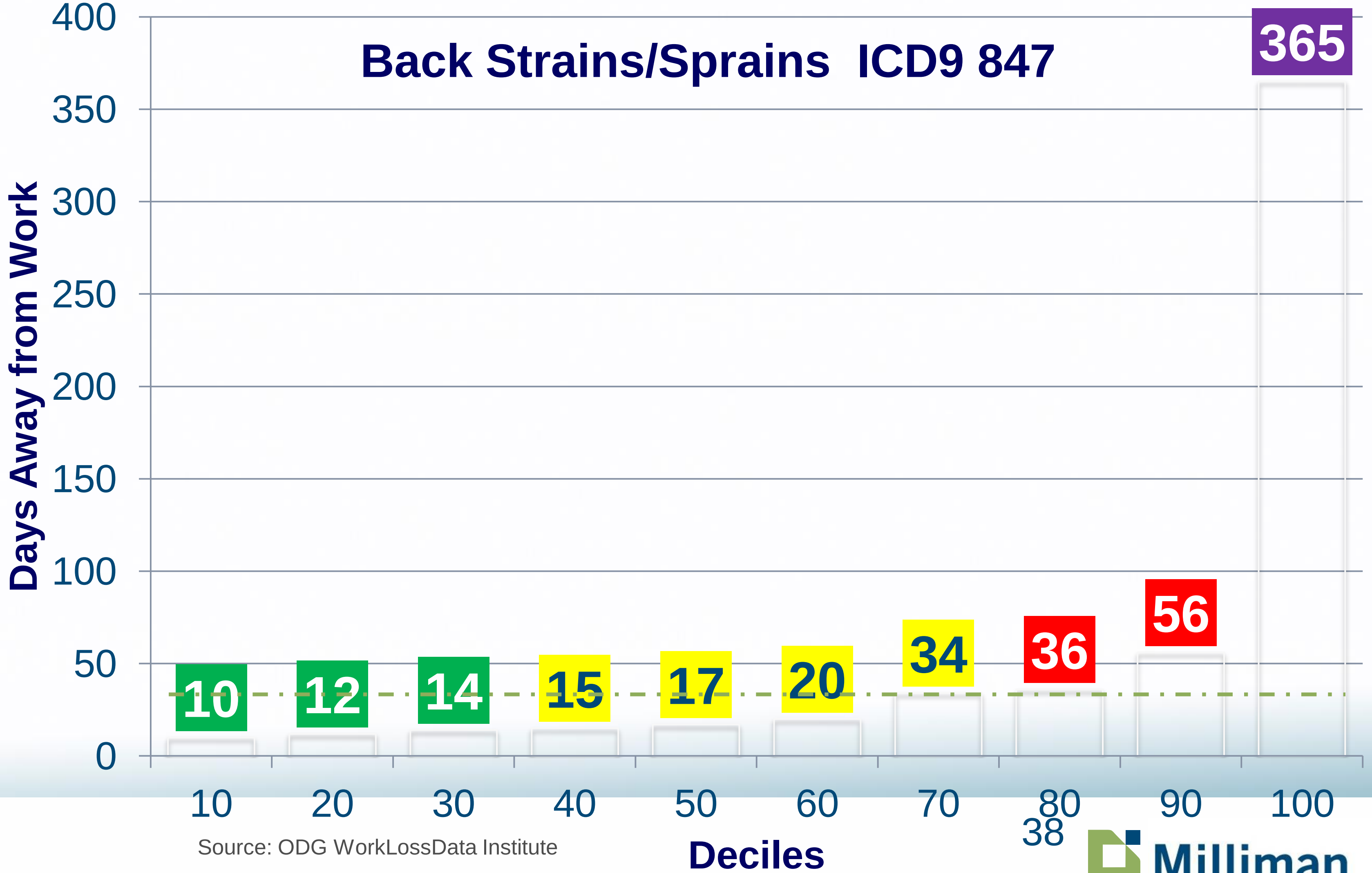
SCORE
1 to 100

Claim Complexity ID Model



RTW Claims Data

Retrospective

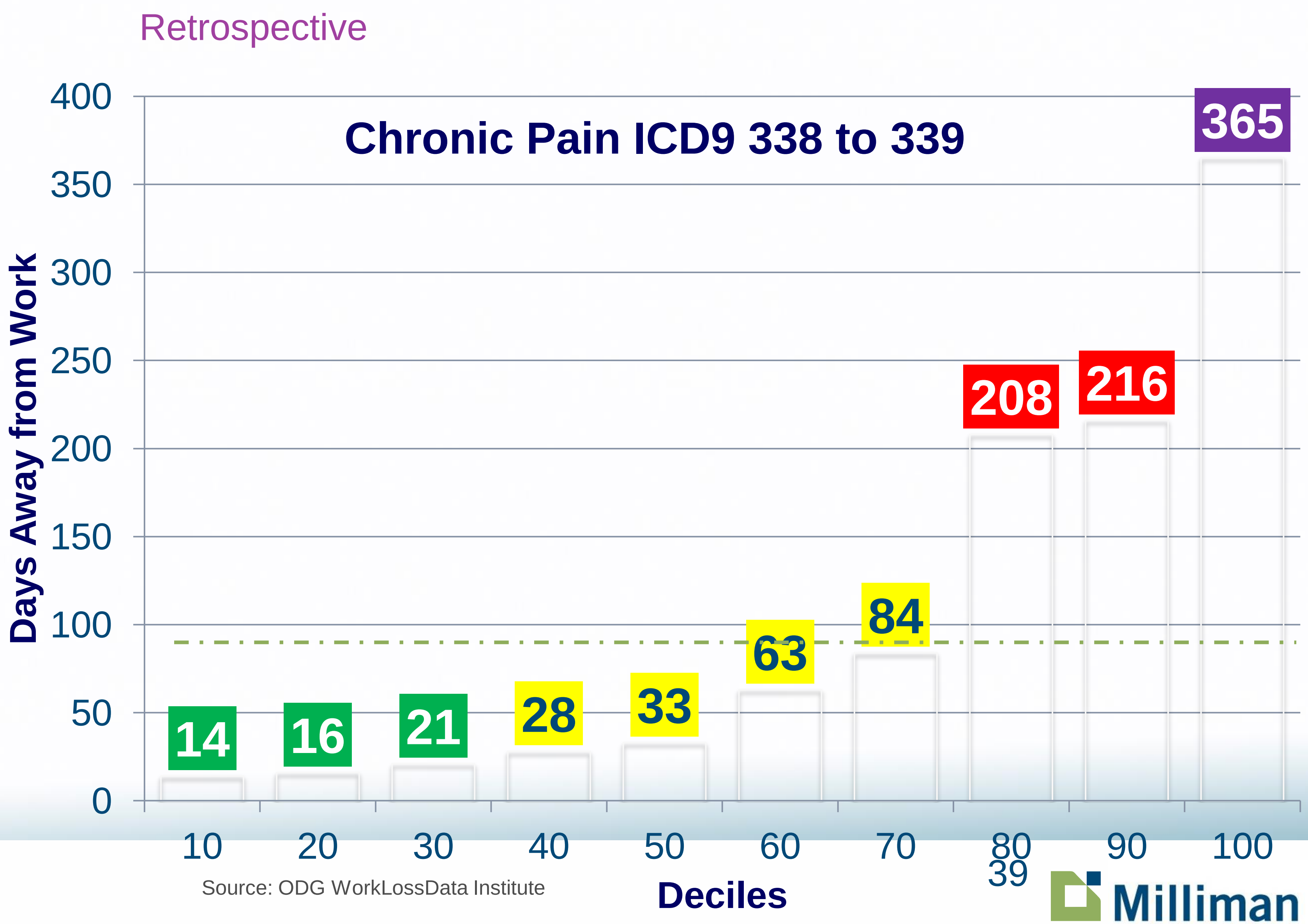


Source: ODG WorkLossData Institute

38

Milliman

RTW Claims Data

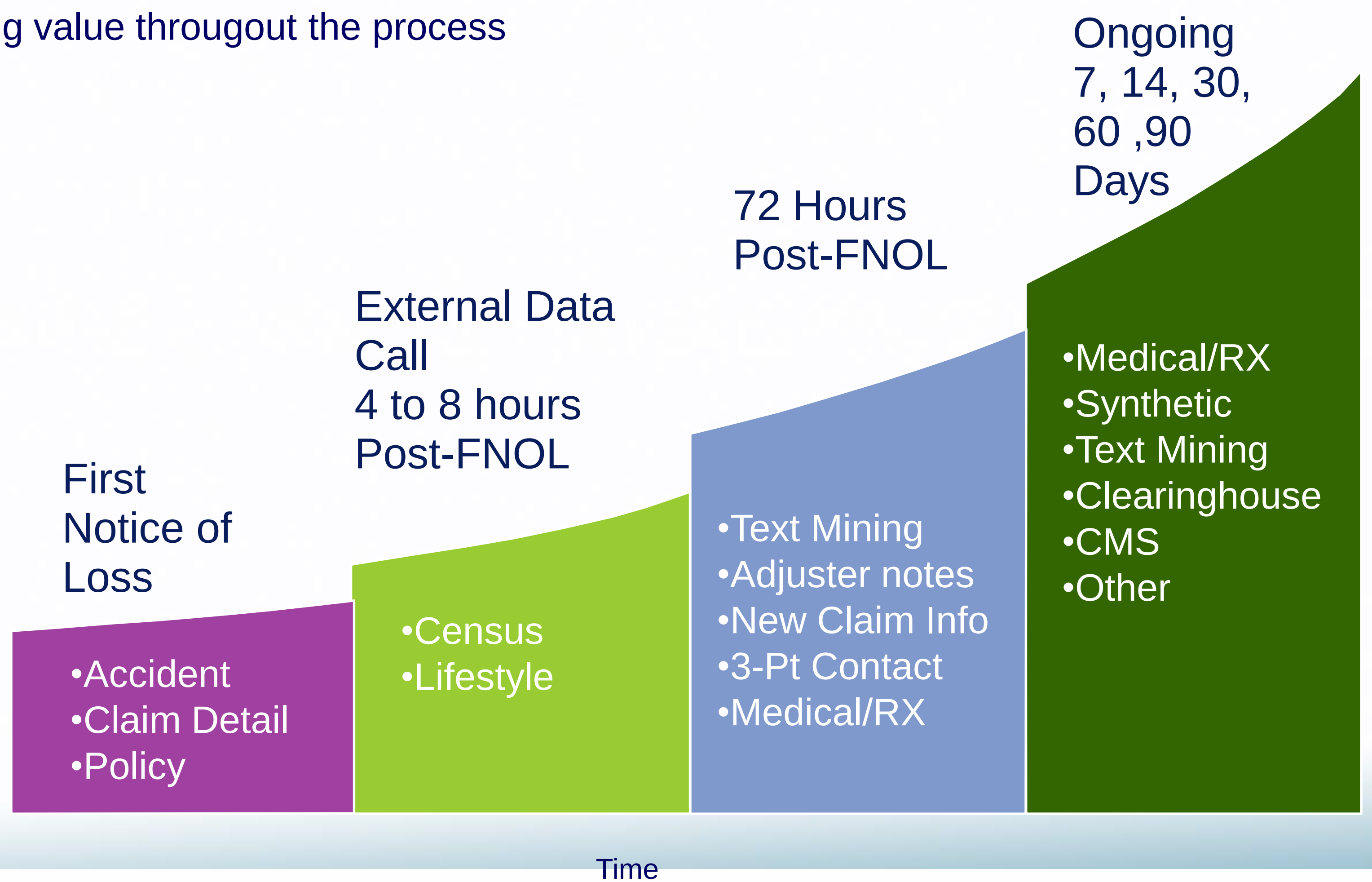


Source: ODG WorkLossData Institute

39

Model Scoring Timing

Bringing value througout the process



Milliman

Implementing Model Output

Predictive models bring order and logic in tackling claim issues.

Segmentation	Business Rules	Claim Strategies
• Identify minor or routine vs. • Severe and complicated cases.	• Codify the model output • Develop consistent best practices	Actionable claim strategies achieve successful outcomes.

Scores are Silent - Actions are **Loud**

41

Predictive Modeling Examples

Identify, measure, manage and reduce claim financial risk

Identifies the high-risk, high-cost claim

Offers better opportunities to manage to a positive outcome

Predictive models are algorithms that can prospectively identify certain types of cases, strategies and assignment patterns.

1. Claim Complexity ID
2. High Cost Claim ID/MSA
3. RTW ID
4. Nurse Case Management
5. Exaggerated Lost Time ID
6. Case Reserving
7. Medical Provider Performance
8. Improved Negotiation Strategies
9. Loss Control Efforts to Outcomes
10. Robust Claim Metrics

.

Predictive Modeling Examples

Identify, measure, manage and reduce claim financial risk

Identifies the high-risk
hig

Carrier triage process identified only 50% of total High Cost cases

Predictive models are algorithms that can prospectively identify certain types of cases, strategies and assignment patterns.

1. Claim Complexity ID
2. High Cost Claim ID/MSA
3. RTW ID
4. Nurse Case Management
5. Exaggerated Lost Time ID
6. Case Reserving
7. Medical Provider Performance
8. Improved Negotiation Strategies
9. Loss Control Efforts to Outcomes
10. Robust Claim Metrics

.