



Predictive Modeling for Workers' Compensation Claims

Buckeye Actuarial Continuing Education




**Joe Anderson, Director of Analytic Services
Progressive Medical, Inc.**

**NOTE: Data in this presentation has been adjusted to
protect proprietary information**

April 1, 2013

WHAT IS PREDICTIVE MODELING?

Predictive Modeling is making decisions with statistics and data.

Company	Goal of predictive model	Result
	Identify new mothers as quickly as possible to get them in the habit of shopping at Target.	Delivered coupons to young mothers before their family even knew they were expecting.
	Determine which movies its customers will like, based on what they have already rated	Improved their predictions by 10%; a \$1,000,000 prize was awarded.
	Choose the best baseball players available for the next season, with a limited budget.	20 consecutive wins; the book and film <i>Moneyball</i> are based on this.

Sources:

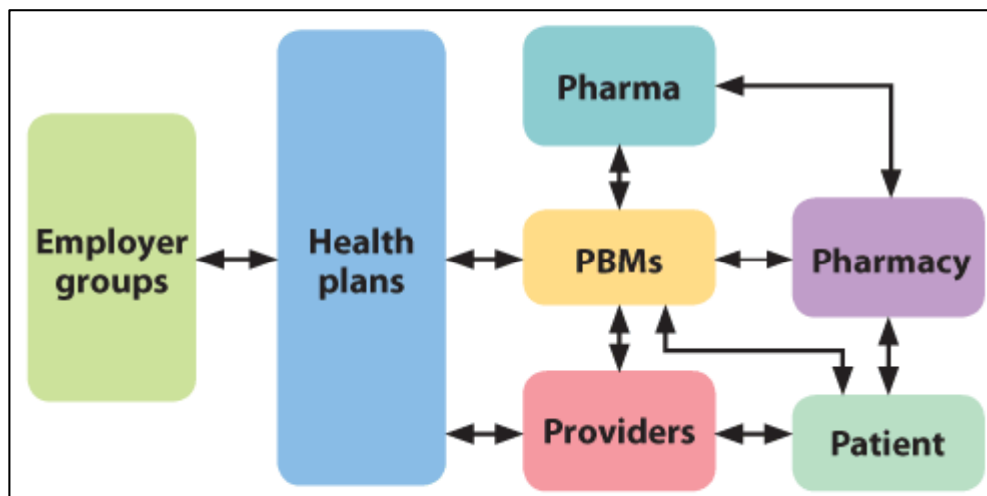
Duhigg, C., *How Companies Learn Your Secrets*, *The New York Times Magazine*. 2012 February 16

Lohr, S., *A \$1 Million Research Bargain for Netflix, and Maybe a Model for Others*, *The New York Times*, 2009 September 21

Mahler, J., *Smaller Markets and Smarter Thinking*, *The New York Times*, 2011 October 14

HOW CAN WE USE IT?

- As a PBM, we see some of the data going through the system, but not all of it.



- Each company in the industry needs can use analytics with their own data:
 - Imagine if Netflix wants to know whether you'll enjoy the movie *Moneyball*.
 - Netflix doesn't know if you have read the book *Moneyball*, if you studied statistics or if you are an Oakland Athletics fan.
 - They do know if you like other baseball movies, other Brad Pitt movies and other movies based on nonfiction books.

THE DATA MINING PROCESS

- Identifying the problem
- Transforming data into information
- Taking action
- Measuring the outcome

THE PROBLEM

A solution is needed that reduces prescriptions most efficiently.

Prescription Drug Deaths and Increasing Costs

- More people are dying from prescription drug use.
- Prescription drug prices are rising.
- Workers' compensation in particular has seen increases in use of prescription pain killers.

Time Constraints on Nurses, Adjustors, Clinicians

- Cannot examine or intervene on every claim
- Cannot determine which claims will have high long-term costs
- Too many "false positives" from individual clinical triggers (i.e. only 10% of claims with morphine equivalence of 90mg result in high long-term costs)

THE SOLUTION: MULTIVARIATE STATISTICAL MODEL TO PREDICT HIGH-COST CLAIMS

**Correlate early
data about an
injured worker...**

Workers injured in 2007

**... with resulting
long-term spend of
that injured worker.**

Resulting pharmacy costs in 2009-2010
(DEPENDENT VARIABLE)



THE SOLUTION: EARLY STEPS IN THE PROCESS

- Start with the data and a theory about how to use it.
- Add “expert opinion” after gaining familiarity with the data.
- This avoids “confirmation bias” and allows the analyst to develop their own area of expertise before confronting ideas about the data that may not be true.
- Example: Strength of medications vs. duplication of medications (“duplicate therapies”)

TRANSFORMING DATA INTO INFORMATION

- Example 1: Converting “Injury Type” to “Injury Risk”
- Example 2: Determining whether an injured worker had surgery
- Example 3: Whether to use regression, decision tree, neural network, etc.

RISK SCORES FOR EACH INJURED WORKER

Claim	Risk Score	Reason
Allison	6.5	Multiple Neck Injury, High Total Medication Use (Including Narcotics)
Bob	5.4	Continued Medication Use, High Risk Prescriber: Allergy and Immunology Specialist
Cindy	5.0	Multiple Prescribers in Early Months, High Days Supply of Various Medications
Dwayne	4.5	High Risk State and Moderate Injury Risk: Dislocated Disc
Elaine	3.9	Prescriber Risk: Pain Management Specialist, High Narcotics Use To-Date
Frank	3.1	Moderate Injury Risk, Demographic Risk and Prescriber Risk: Pain Management Specialist

EXAMPLE 1: FROM “CATEGORY” TO “CATEGORY RISK”

- We expect that an injured worker’s injury type will predict their long-term cost.
- We cannot use the words “lower back strain” in a statistical model; we need to assign a number to that.
- We can do that through binary variables or through another method.

COMPARING DEPENDENT VARIABLE TO CATEGORIES

- Each observation should have a category assigned to it and a dependent variable associated with it.

Claimant	Injury Type	Long-Term Spend
Andrew	Ruptured Disc	\$10,000
Bart	Hand Laceration	\$5,000
Claire	Lower Back Strain	\$2,500
Diego	Fractured Arm	\$1,000
Eleanor	Lower Back Strain	\$1,000

CALCULATE THE AVERAGES

- Each category will have an average value of the dependent variable associated with it.

Injury Type	Average
Concussion	\$ 643,588
Punctured Lung	\$ 18,812
Dislocated Disc	\$ 64,590
Carpal Tunnel Syndrome	\$ 4,556
Multiple Physical Injuries	\$ 42,508
Hernia	\$ 8,698

DETERMINING SIGNIFICANCE

- Using the number of observations with that category and the variance within the category vs. the total number of observations and the total variance, we can find which injuries are different with statistical significance.

Injury Type	Average	Significant?
Concussion	\$ 643,588	Yes
Punctured Lung	\$ 18,812	Yes
Dislocated Disc	\$ 64,590	No
Carpal Tunnel Syndrome	\$ 4,556	Yes
Multiple Physical Injuries	\$ 42,508	No
Hernia	\$ 8,698	Yes

USING SIGNIFICANCE

- If the category is different, with statistical significance, use the category's value, otherwise use the average value.

Injury Type	Average	Significant?	Value for Model
Concussion	\$ 643,588	Yes	\$643,588
Punctured Lung	\$ 18,812	Yes	\$ 18,812
Dislocated Disc	\$ 64,590	No	\$ 30,000
Carpal Tunnel Syndrome	\$ 4,556	Yes	\$ 4,556
Multiple Physical Injuries	\$ 42,508	No	\$ 30,000
Hernia	\$ 8,698	Yes	\$ 8,698

EXAMPLE 2: WHO HAD SURGERY?

- As a PBM, we do not have data on whether injured workers have surgery or not.
- You may have other problems with lack of data. For example, you may want to know whether a certain person engages in certain high-risk behaviors.
- A good analytics model will incorporate all known data and support it by purchasing other data where available.

FINDING SPECIALTIES

- Use what we do know: who was the first prescriber for that injured worker?

Claimant	First Prescriber
Andrew	Dr. Kelly
Bart	Dr. Lewis
Claire	Dr. Mahajan
Diego	Dr. Newman
Eleanor	Dr. Ogden

FINDING SPECIALTIES

- We can buy data on which prescribers are associated with which specialty.

Claimant	First Prescriber	Prescriber Specialty
Andrew	Dr. Kelly	Orthopedics
Bart	Dr. Lewis	General Surgery
Claire	Dr. Mahajan	General Surgery
Diego	Dr. Newman	General Practice
Eleanor	Dr. Ogden	Orthopedic Surgery

SPECIALTY RISK

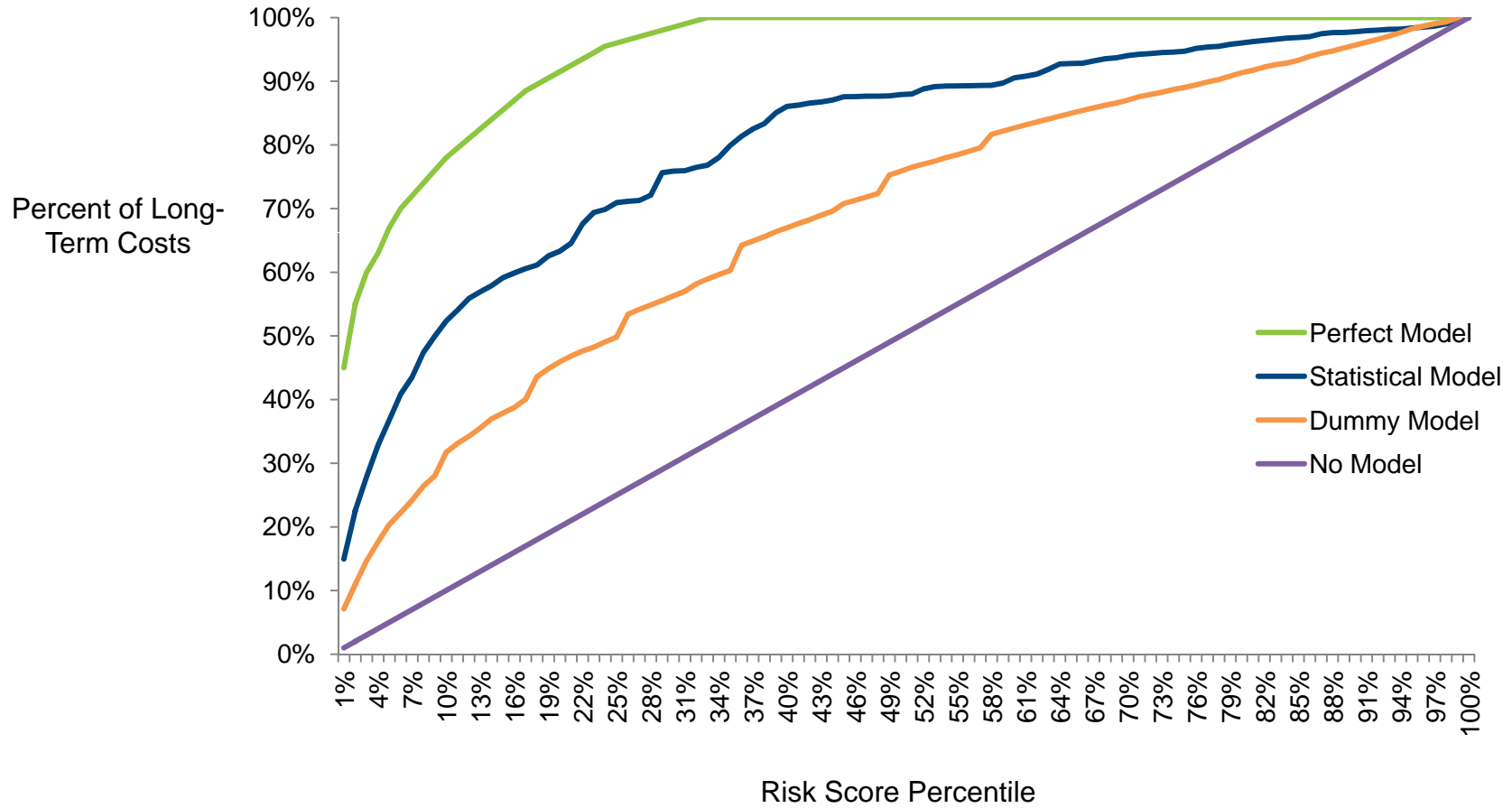
- We give each specialty a risk, which is a proxy for whether a claimant had surgery.

Claimant	First Prescriber	Prescriber Specialty	Prescriber Specialty Risk
Andrew	Dr. Kelly	Orthopedics	2.2
Bart	Dr. Lewis	General Surgery	1.5
Claire	Dr. Mahajan	General Surgery	1.5
Diego	Dr. Newman	General Practice	1.0
Eleanor	Dr. Ogden	Orthopedic Surgery	1.8

EXAMPLE 3: TYPE OF MODEL TO USE

- The statistical methodology is a choice, like whether to use a hammer, glue or tape to hang something on a wall.
- Each type of statistical model (regression, neural network, decision tree) has its advantages:
 - Regression is the most well-known; most statisticians understand its outputs. Its components can be broken apart.
 - Decision tree is easiest to explain to a layperson and handles interactions better than regression.
 - Neural network can have higher predictive power than regression but is more of a “black box” than the other two.

MEASURING EFFECTIVENESS



TAKING ACTION

- We intervene on claims that are at risk for high pharmacy cost:
 - Claims Professional Outreach
 - Physician Outreach
 - Drug Utilization Evaluation
 - Peer-to-Peer Review

MEASURING THE OUTCOME

*Statistical Confidence that
Intervention Changes this Outcome*

