

# Propensity Scoring: Theory and Applications

Leo Guelman, PhD (candidate)  
RBC Insurance

Michael Kim, PhD  
CNA Insurance

CAS RPM Seminar  
Dallas, TX  
March 9 -11, 2015

# Antitrust Notice

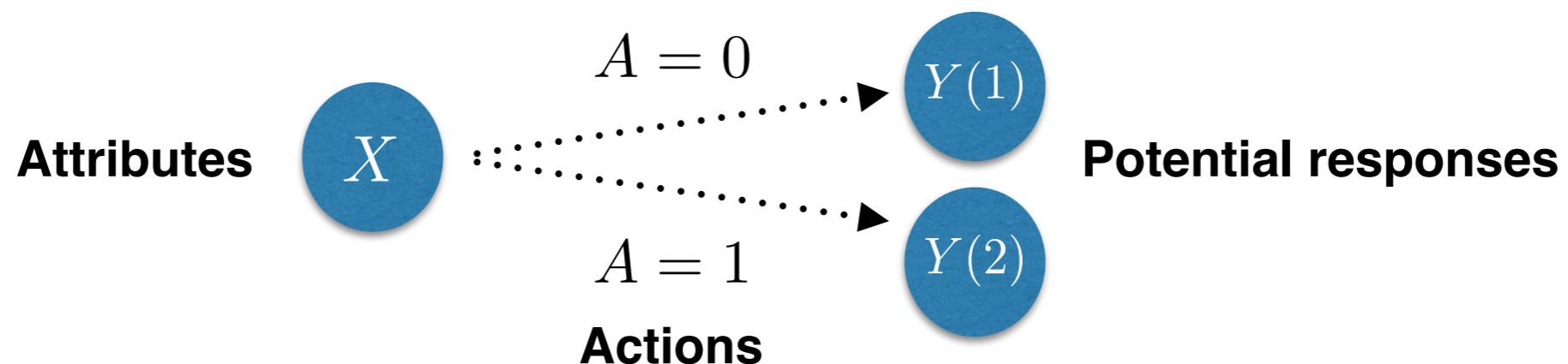
- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such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

# From Predictive to Causal Modeling

- **Predictive Modeling** has been established as a core strategic capability of many top insurers.
- **Common goal:** to predict a response variable using a collection of attributes under static conditions — i.e., assumes “business as usual” conditions.



- **Causal Modeling goes one step further:** the interest is in estimating/predicting the response under **changing conditions** — e.g., induced by alternative actions or “treatments”.



# Key Questions Faced by Decision Makers

Question	Estimand
Does the action affect the outcome?	<b>Average Treatment Effect (ATE)</b>
Does the action affect the outcome differently for different customer types?	<b>Subgroup Treatment Effect (STE)</b>
What is the impact of the action at the individual customer-level?	<b>Individual Treatment Effect (ITE)</b>

The **ITE** is an *unobserved* quantity, as a customer can never be observed simultaneously under more than one action — this is known as “*fundamental problem of causal inference*”.

# Causal Modeling with Experimental Data

- **Randomization**: the “gold standard” for scientific research.
  1. Randomly sample subjects from the population.
  2. Randomly assign subjects to treatment and control conditions.
  3. Estimate the ATE:

$$ATE = E[Y|A = 1] - E[Y|A = 0] .$$

- As the sample size grows, the client attributes  $\mathbf{X}$  will tend to be “balanced” between treatment and control groups — subjects become “**exchangeable**”.
- Even in an experimental setup, much can go wrong which requires statistical correction (e.g., [Barnard](#), [Frangakis](#), [Hill](#), and [Rubin 2003](#)).

# Example: Price-Elasticity Estimation

- **Objective:** Estimate the expected portfolio renewal rate among auto policyholders under alternative rate increases.
- In this context, the **rate increase plays the role of the treatment** (e.g., 5% vs. 10% rate increase), and the response represents the **renewal outcome (Y/N)**.
- Under randomized assignment of policyholders to rate changes, the **ATE (price-elasticity here)** can be computed straightforwardly.

Rate Change	+5%	+10%
N (policies)	10,000	10,000
Retained policies	9,200	8,700
Retention Rate	92%	87%

→ ATE = 87% - 92% = (5%)

# Experimentation: Challenges and Threats

- Most insurance data come from “business as usual” conditions:
  - High costs associated with experimentation.
  - Legal and/or regulatory constraints.
  - Violation to ethical standards.
  - Lack of planning.
- In the absence of randomization, subjects are no longer exchangeable and thus direct comparisons can be misleading ([Rosenbaum and Rubin, 1983](#)).
- Assume the following extreme scenario (“X” indicates NO available data):

Rate Change	+5%	+10%
Age < 25 yrs	✓	✓
Age >= 25 yrs	✓	X

>= 25 yr old clients exposed to a 5% rate increase do not have a counterfactual in the 10% rate change group.

# Regression-based Estimation: What Can Go Wrong?

- Need to adjust any difference in the outcome for differences in the client attributes under alternative treatments.
- The **standard regression approach** estimates:

$$Y = \alpha + \tau A + \beta X + \epsilon$$

where the slope  $\tau$  of the treatment indicator is an estimator of the average treatment effect.

- In the absence of experimental data, **the standard regression approach is unreliable** ([Berk, 2004](#)):
  - **Regression-based methods mask non-overlap problems**, and they extrapolate inferences in regions of the predictors where certain treatment haven't been observed.
  - The **problem is worse with a large number of predictors** (Big Data), as we cannot easily see non-overlap problems.
  - **Standard statistical software can be deceptive** (does not issue any warnings about non-overlap problems).

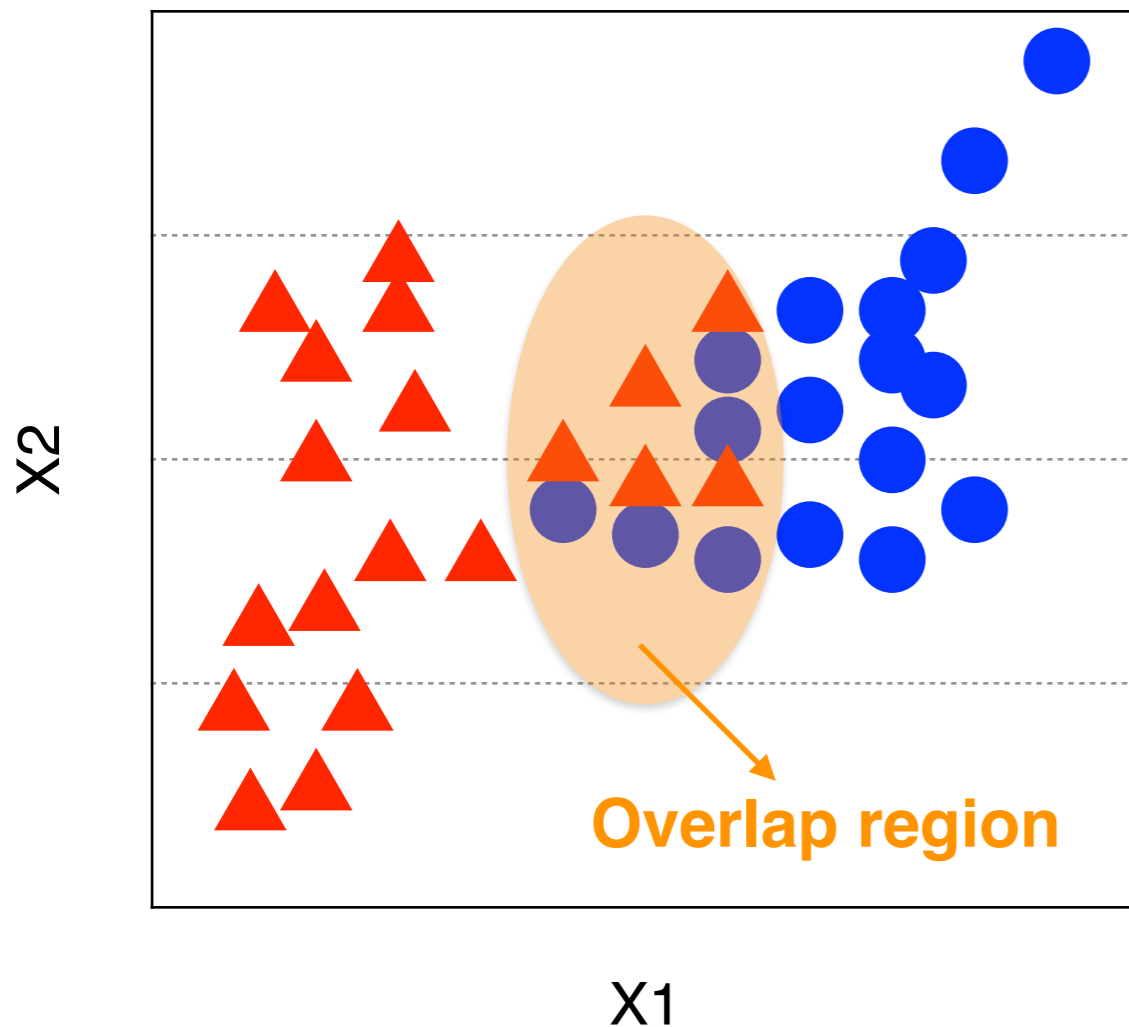


# Causal Inference with Observational Data

- An **observational study** attempts to draw inferences about the effect of treatments in the absence of experimental data (a.k.a. observational data).
- [Rosenbaum and Rubin \(1983\)](#) proposed **propensity score matching** as a method to remove the bias in the estimation of treatment effects from observational data.
- These methods have become increasingly popular in a wide variety of fields (from economics to medicine).
- **Key concept:** Under certain data conditions, we can approximate a randomized experiment from observational data.

# Key Data Conditions: **Common Support**

**Common support** (a.k.a. overlap) requires that similar customers received different treatments.

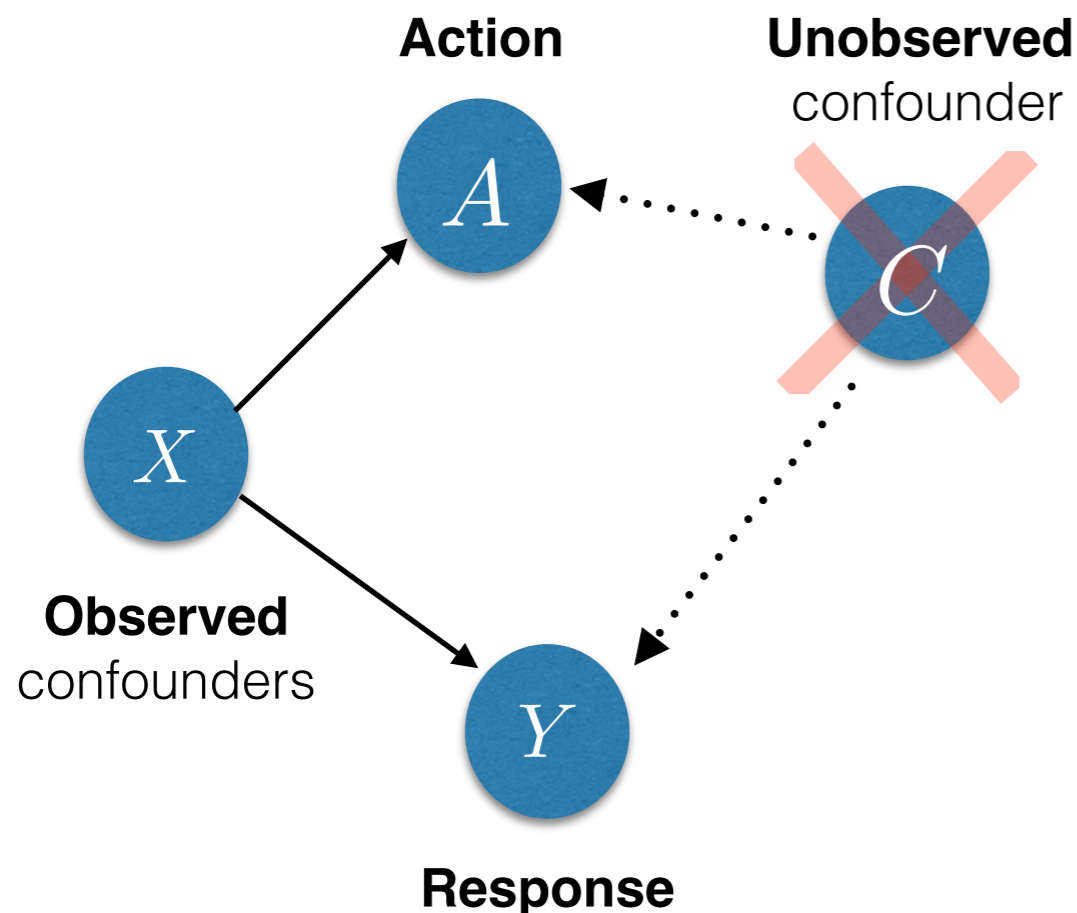


- Estimates of treatment effects are only reliable within the overlap region.
- Outside the overlap region, causal effect estimates involves risky extrapolation.

**Figure:** Distribution of customers colored by action.

# Key Data Conditions: **Unconfoundedness**

**Unconfoundedness** requires that historic actions were entirely based on the observed attributes **X**.



- Confounders are variables associated with both the treatment and the outcome.
- Unobserved confounders will bias treatment effect estimates (omitted variable bias).
- Unconfoundedness is untestable and irreversible by statistical methods.

# Propensity Score

- In randomized experiments, subjects are assigned to actions using some sort of **random mechanism**.
- In the absence of randomization (and assuming the key data conditions hold) subjects are assigned to actions on the basis of their attributes **X**.
- To approximate a randomized design from observational data **we need to understand the assignment mechanism**.
- This is answered by the **Propensity Score**, which is defined as the conditional probability of assignment to treatment given the attributes:

$$\pi(X) = \text{Prob}(A = 1|X).$$

# Matching: Conceptual Framework

- As a result of the non-random treatment assignment mechanism, subjects exposed to different actions have different distributions of  $\mathbf{X}$  — i.e., their attributes are **not balanced**.
- **Goal of matching:** Achieve balance on attributes.
- If we pair subjects that have **exact the same attribute values** but differ only in the treatment they received, we could achieve **perfect balance**.
- Perfect balance is not feasible even for moderate number of attributes in  $\mathbf{X}$  (or if  $\mathbf{X}$  contains continuous attributes), so we need alternative methods.
- **The key idea:** Pair subjects that differ in the treatment they received, but have **approximately the same probability** of being assigned to the same treatment — i.e., the same propensity score.

# Propensity Score — Balancing Property

- What allows us to pair (match) subjects based ONLY on the propensity score?
- **Balancing Property:** Treatment assignment **A** and attributes **X** are conditional independent given the propensity score.

$$A \perp \mathbf{X} | \pi(\mathbf{X})$$

- In words, if we match subjects on the propensity score, the **distribution of X** will be similar across treatment groups in the matched sample.

# Matching Algorithms

Matching algorithms have many variants. There are 3 key choices:

1. The **definition of distance** between two subjects in terms of their attributes.
2. The choice of the **algorithm** used to form the matched pairs and make the distance “small” (**greedy vs. optimal matching**).
3. The **structure of the match** — i.e., the number of treated and control subjects that should be included in each match set.

# Application to P&C: Price-Elasticity Estimation

## Objectives

1. Obtain auto insurance price elasticity estimates at the portfolio level (Average Treatment Effect).
2. Identify client subgroups with varying price sensitivities (Subgroup Treatment Effect).

## Data considerations

- As it is often the case, no access to experimental data.
- Clients were historically exposed to rate change levels based on (i) a pricing modeling exercise, (ii) regulatory constraints, (iii) competitive analysis, and (iv) general business objectives.



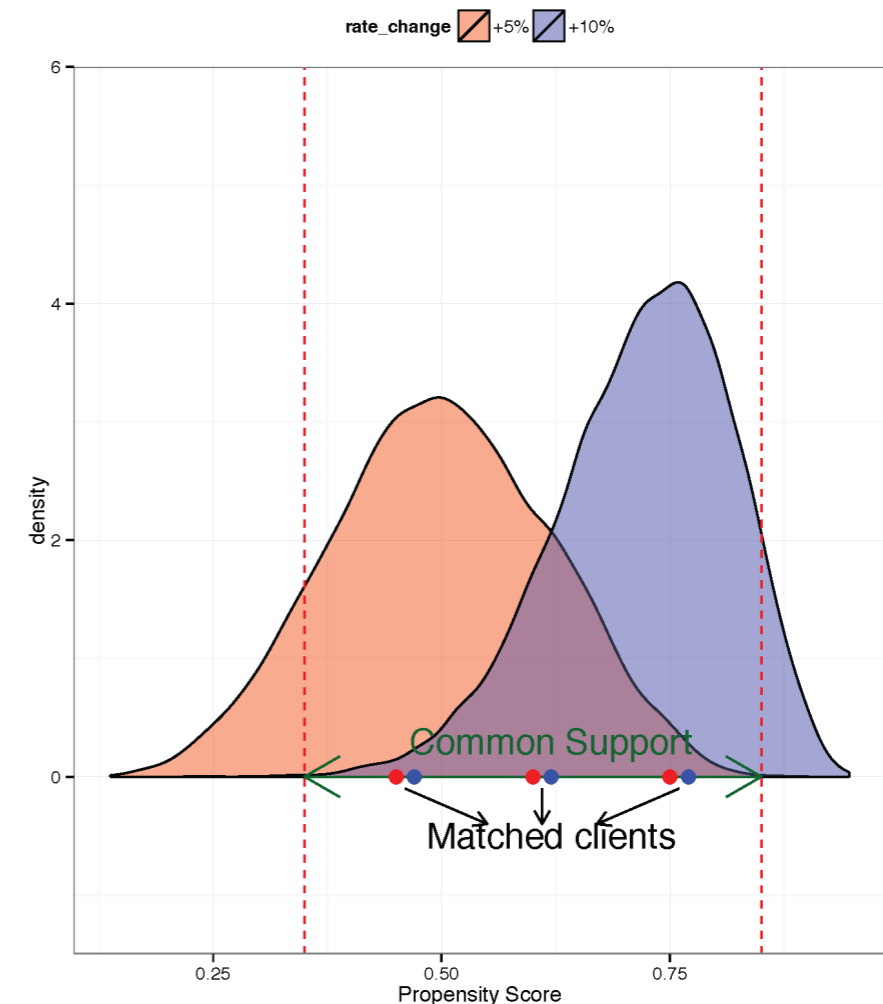
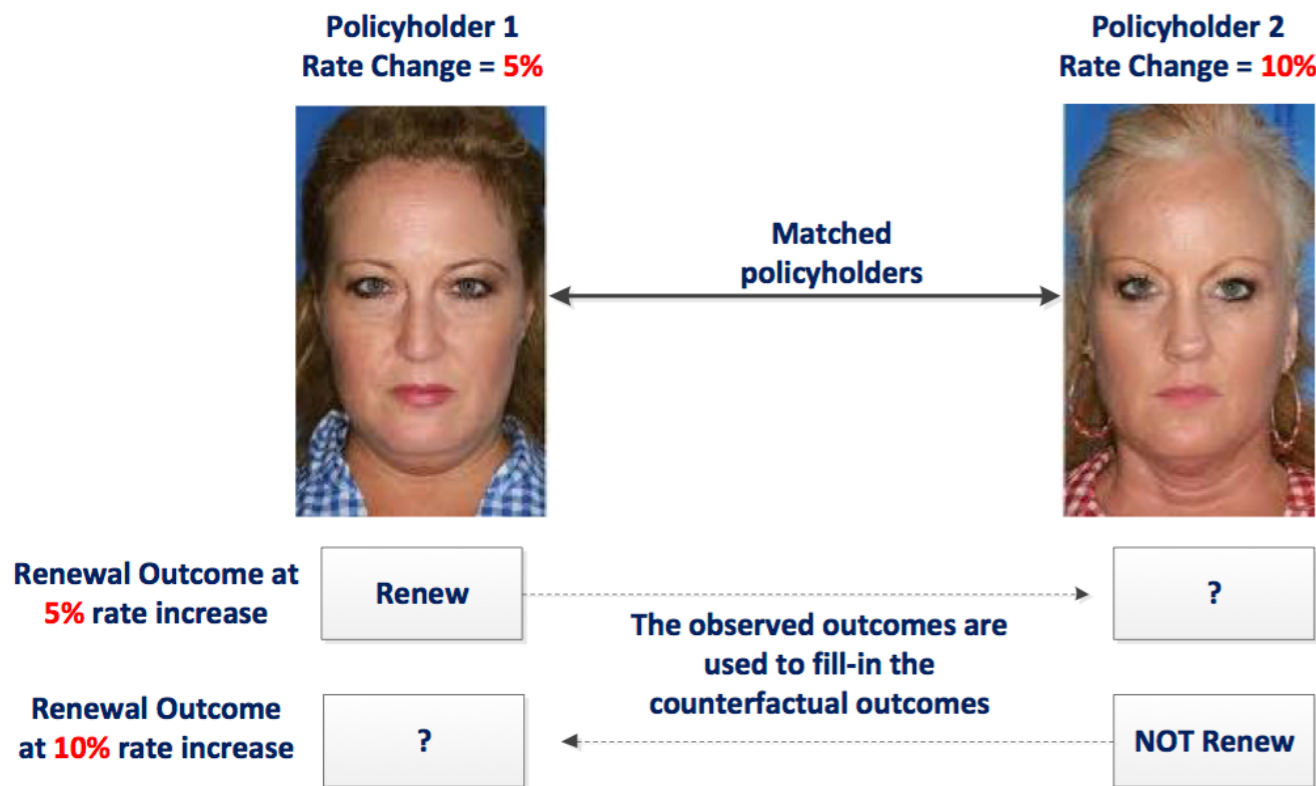
# The starting point: Client-by-rate change table

- The entries  $r_{\ell a}$  below denote the observed renewal outcome  $\in \{0, 1\}$  of policyholder  $\ell = \{1, \dots, L\}$  when exposed to rate change level  $A = a$ .
- To simplify, the rate change is binned into five ordered values  $A = \{1 < \dots < 5\}$ .
- Dots indicate counterfactual outcomes, which are unobserved.
- The price elasticity estimation problem is equivalent to the problem of filling in the missing values in the client-by-rate change table with reliable estimates.

Table 1: **Client-by-Rate change table**

Client	Rate Change Level				
	Level 1	Level 2	Level 3	Level 4	Level 5
1	.	$r_{12}$	.	.	.
2	.	.	$r_{23}$	.	.
3	$r_{31}$	.	.	.	.
4	.	.	.	$r_{44}$	.
5	.	$r_{52}$	.	.	.
6	.	.	.	.	$r_{65}$
...	...	...	...	...	...
L	.	.	.	.	$r_{L5}$

# Propensity Score Matching



- “**Clone policyholders**”: similar in terms of the relevant lapse predictors — i.e., about the same age, driving record, live in the same neighbourhood, etc.
- But exposed to different rate change levels.

- **Propensity score**: Probability of assignment to a 10% relative to a 5% rate increase.
- Distribution of propensity score is shown for each rate change group.
- Clients are matched only in the common support (overlap) region.

# Filling the client-by-rate change table

**Step 1:** Replace the actual renewal outcomes with probability estimates — by estimating  $E[r_{\ell a} | X, a]$ .

**Step 2:** Infer the counterfactual renewal outcomes from the matched pairs (as far as the overlap situation permits).

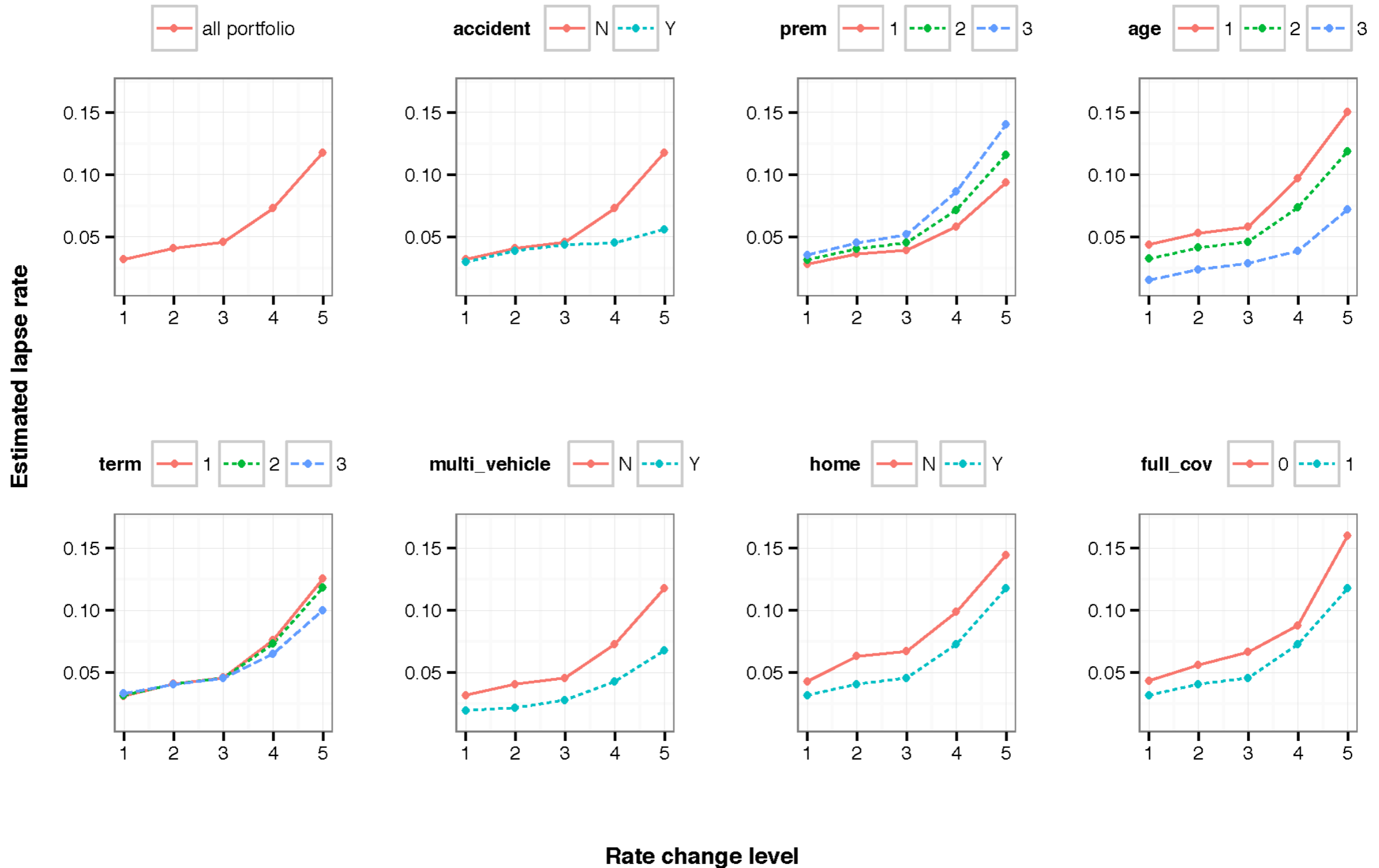
**Step 3:** Develop a “global model” of the response  $\Rightarrow$  fit the observed + counterfactual renewal estimates on **X** and **A**.

Client	Rate Change Level				
	Level 1	Level 2	Level 3	Level 4	Level 5
1	.	$\hat{r}_{12}$	.	.	.
2	.	.	$\hat{r}_{23}$	.	.
3	$\hat{r}_{31}$	.	.	.	.
...	...	...	...	...	...
L	.	.	.	.	$\hat{r}_{L5}$

Client	Rate Change Level				
	Level 1	Level 2	Level 3	Level 4	Level 5
1	$\hat{r}_{11}$	$\hat{r}_{12}$	$\hat{r}_{13}$	$\hat{r}_{14}$	$\hat{r}_{15}$
2	$\hat{r}_{21}$	$\hat{r}_{22}$	$\hat{r}_{23}$	$\hat{r}_{24}$	.
3	$\hat{r}_{31}$	.	.	.	.
...	...	...	...	...	...
L	$\hat{r}_{L1}$	$\hat{r}_{L2}$	$\hat{r}_{L3}$	$\hat{r}_{L4}$	$\hat{r}_{L5}$

Client	Rate Change Level				
	Level 1	Level 2	Level 3	Level 4	Level 5
1	$\hat{\hat{r}}_{11}$	$\hat{\hat{r}}_{12}$	$\hat{\hat{r}}_{13}$	$\hat{\hat{r}}_{14}$	$\hat{\hat{r}}_{15}$
2	$\hat{\hat{r}}_{21}$	$\hat{\hat{r}}_{22}$	$\hat{\hat{r}}_{23}$	$\hat{\hat{r}}_{24}$	$\hat{\hat{r}}_{25}$
3	$\hat{\hat{r}}_{31}$	$\hat{\hat{r}}_{32}$	$\hat{\hat{r}}_{33}$	$\hat{\hat{r}}_{34}$	$\hat{\hat{r}}_{35}$
...	...	...	...	...	...
L	$\hat{\hat{r}}_{L1}$	$\hat{\hat{r}}_{L2}$	$\hat{\hat{r}}_{L3}$	$\hat{\hat{r}}_{L4}$	$\hat{\hat{r}}_{L5}$

# Price Elasticity Functions



# Case Study: CNA Medical Case Management

# Background

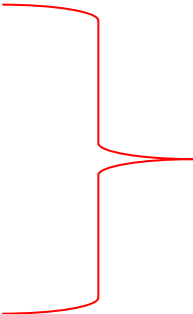
- Nurse medical case management (MCM) is a common method intended to improve claim outcomes for injured workers<sup>1</sup>
- MCM is intended to provide care coordination, more efficient medical care review, and quality services<sup>1</sup>
- Research indicates MCM may be associated with a decrease in workers' compensation related costs; however, the causal relationship is not known<sup>2</sup>

<sup>1</sup>Beernacki EJ, Tsai SP, Shan P. Ten Years' Experience Using an Integrated Workers' Compensation Management System to Control Workers' Compensation Costs. *Journal of Occupational and Environmental Medicine*. 2003. (45) 5: 508-516.

<sup>2</sup>Jarvis KB, Phillips RB, Morris EK. Cost per Case Comparison of Back Injury Claims of Chiropractic versus Medical Management for Conditions with Identical Diagnostic Codes. *Journal of Occupational Medicine*. 1991. 33(8): 847-852.

# Background

- CNA Insurance utilizes medical case management for workers' compensation claims
- Historical savings estimates and descriptive statistics on medical and indemnity costs were unreliable, thus a different analytic approach to understand the impact of MCM and drivers of cost was necessary
- As mentioned, randomization of intervention helps to reduce selection bias and provides a stronger association and helps to establish causality
- However, a Randomized Control Trial (RCT) of MCM at CNA Insurance not feasible due to several constraints
- Alternative to RCT – Observational/Retrospective Data
  - Descriptive Analyses
  - Case/Control Study
  - Qualitative Study
  - Cross-Sectional Study



While prevalent in literature, issues with selection bias, spurious results, unbalanced case/control groups, weaker associations

# Objective

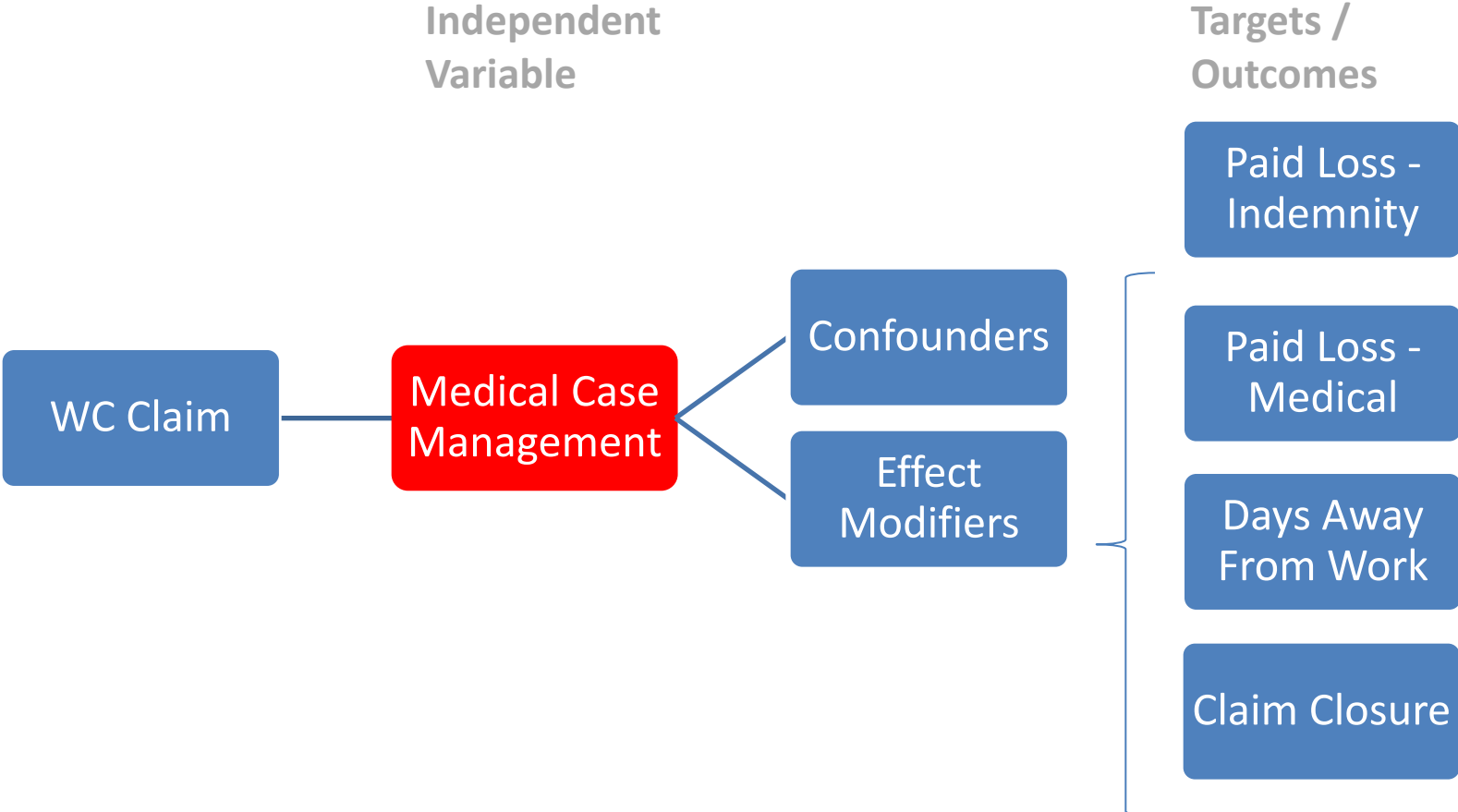
- Can we approximate a RCT with observational data and apply it to measure the impact of MCM on workers' compensation claim outcomes?



# Methodology - Overview

- Data Preparation
  - Variable Selection
  - Longitudinal Data
- Analysis
  - Propensity Scoring with Nearest Neighbor Matching
  - Multivariate regression with Generalized Estimating Equations (GEE)

# Data Preparation – Variable Selection / Conceptual Framework



Unit of Observation

Binary 1/0 Indicator

Binary, Continuous, Categorical

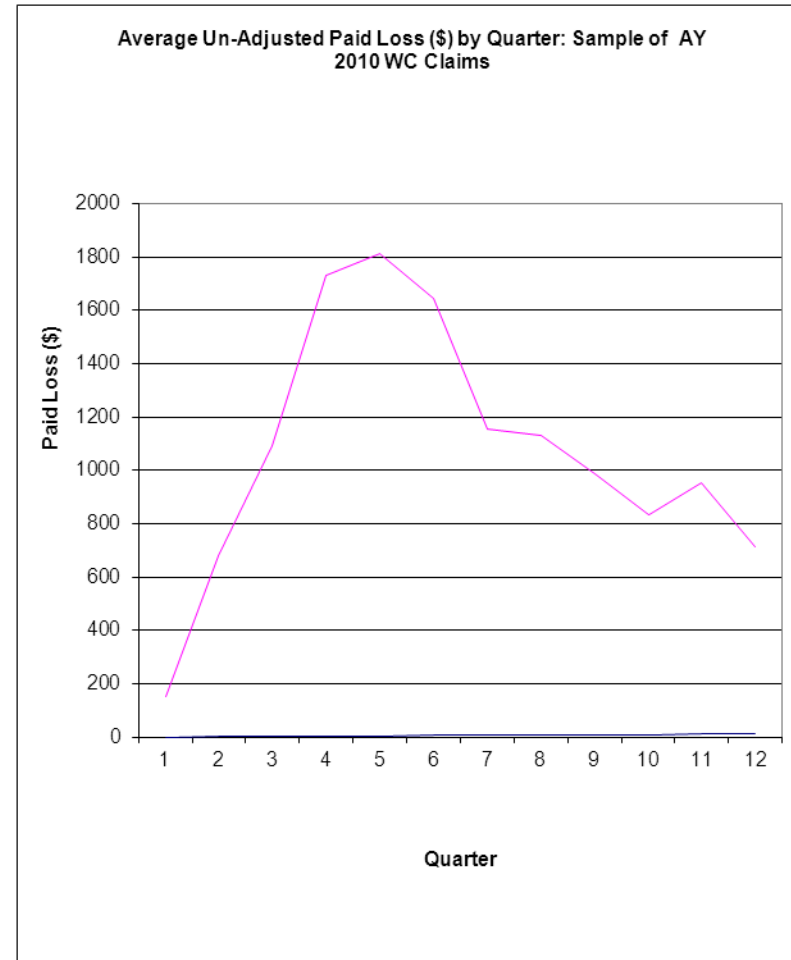
Binary, Continuous, etc.

# Data Preparation – Variable Selection

- Static
  - Variables that are typically unchanged over time
  - Important for propensity scoring
  - Examples include Loss State, Loss Year
- Dynamic
  - Variables that change over time
  - Typically, repeated observations of a variable for a single claim
  - Examples include Paid Loss, Days Away from Work

# Methodology – Longitudinal Data

- Workers' compensation claims are dynamic and can be long tailed
- Repeated observations on a claim will yield clustered observations with correlated data<sup>1</sup>
- Utilization of ordinary models without accounting for autocorrelation will produce incorrect results<sup>2</sup>
- Given this assumption, use dynamic variables accounting for time (quarters for this analysis)



1. Liang K, Zeger S.L. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986. 73(1): 13-22.
2. Ghisletta P. An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics*. 2004. 29(4): 421-437.

# Methodology – Longitudinal Data

“Wide” Format

Claim #	State	LOSS YEAR	Body Part	Paid Loss @ 3	Paid Loss @6	Paid Loss @ n	MCM
1	AL	2010	Elbow	0	1000	5000	1
2	GA	2011	Back	1000	5000	10000	0
3	RI	2010	Back	0	0	1000	1

**Wide format ideal for propensity scoring of static variables pre-treatment**

“Long” Format

Claim #	Quarter	Paid Loss	State	Body Part	MCM
1	3	0	AL	Elbow	1
1	6	1000	AL	Elbow	1
1	N	5000	AL	Elbow	1
2	3	1000	GA	Back	0
2	6	5000	GA	Back	0
2	N	10000	GA	Back	0
3	3	0	RI	Back	1
3	6	0	RI	Back	1
3	N	1000	RI	Back	1

**Long format ideal for longitudinal data analysis of both static and dynamic variables**

# Analysis – Propensity Scoring

- Propensity Score = Probability of Receiving MCM (when they didn't) given a vector of observed variables
- 2 Key Assumptions
  - Conditional Independence/Unconfoundedness
  - Common Support
- Steps
  1. Variable Choice – Static or pre-treatment
  2. Model Choice (Probit or Logit)
  3. Choose Matching Algorithm
  4. Calculate Propensity Score (0-1 Scale)
  5. Check for Balance
  6. Analysis

# Propensity Scoring – Variable Choice

1. **Variable Choice**
2. Model Choice
3. Choose Matching Algorithm
4. Calculate Propensity Score (0-1 Scale)
5. Check for Balance
6. Analysis

- To calculate the probability of medical case management, we need to pick variables that each case will have recorded *prior* to case management assignment
- Static Variables

# Propensity Scoring – Model Choice

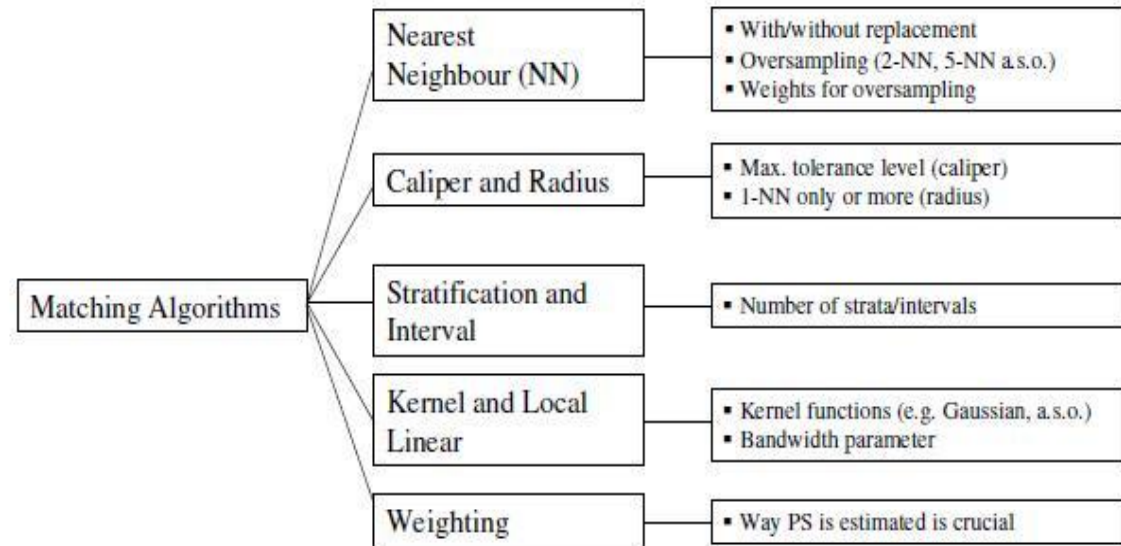
1. Variable Choice
2. **Model Choice**
3. Choose Matching Algorithm
4. Calculate Propensity Score (0-1 Scale)
5. Check for Balance
6. Analysis

- Once we select our variables, we must calculate the odds of medical case management
- Our independent variable, medical case management, will serve as our target variable temporarily
- Can fit a full logistic or probit model or conduct variable selection prior to running the model
  - It is essential to randomize your data prior to propensity scoring
  - Save predictions ( $\hat{y}$ )



# Propensity Scoring – Matching Algorithm

1. Variable Choice
2. Model Choice
3. **Choose Matching Algorithm**
4. Calculate Propensity Score (0-1 Scale)
5. Check for Balance
6. Analysis



**The matching algorithm will be incorporated into an option when the propensity score is calculated with the predicted probabilities in step 2**

# Propensity Scoring – Calculate Propensity Score

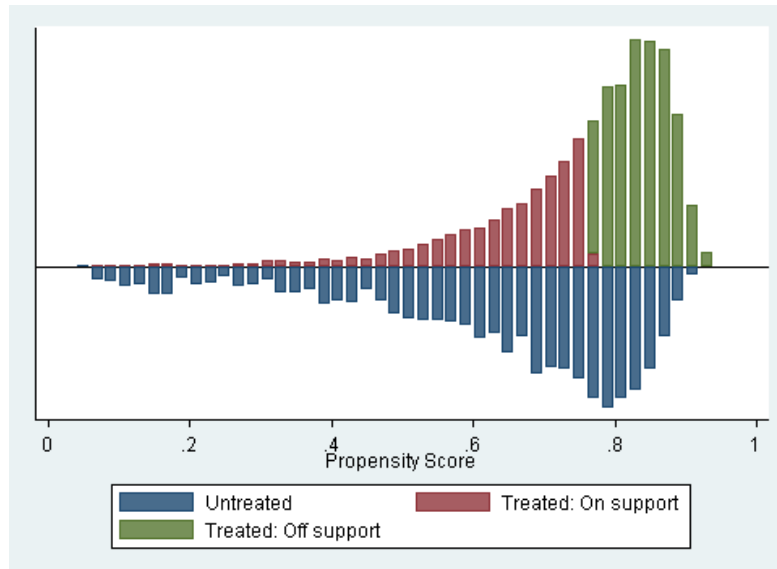
- After running the logistic/probit model, using the predicted probabilities, calculate propensity score (programming considerations on slide xx) with matching algorithm specified and other options. This list is not exhaustive, but represents common elements to expect

Options	Definition	Output	Definition
Matching Algorithm	NN, radius and caliper, kernel, llr, mahalanobis	$\hat{y}$	Predicted log odds
With/Without Replacement	With = replacement; Without = nearest neighbor p-score matching only	p-score	Propensity score (0-1 range)
Common Support	Drops treatment observations whose p-score is higher than the maximum or less than minimum p-score of the controls	Treated	Case Management/ No Case Management (0/1)
Trim	Imposes common support by dropping a percent of treatment observations at which p-score density of control observations is the lowest	Support	1 = common support 0 = off the support
Odds	Matches on odds ratio of propensity score	Weight	NN-Matching = Frequency with which observation is used as a match ; K-Nearest Neighbor = Holds normalized weight ; Kernel / LLR = overall weight given to the matched observation
Index	Use latent variable index instead of probability	Nk	If 1-1 NN Match , observation # of k-th matched control observation
Descending	Perform 1-1 NN in descending order	nn	NN Match, for every treatment observation, it stores # of matched control observation
		pdif / mdif	1-1 NN Match / Mahalanobis, absolute distance to its matched control in terms of propensity score

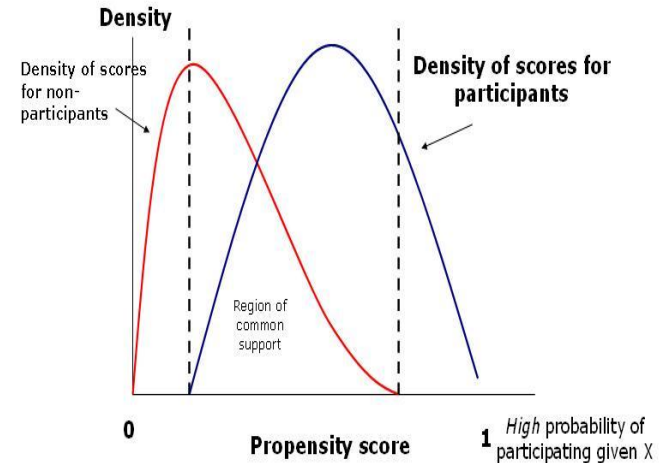
# Propensity Scoring – Check for Balance

- The question is, is our matching procedure able to balance the distribution of relevant covariates?
  - Manual inspection, descriptive statistics of matched claims and covariates
  - Propensity score graph

1. Variable Choice
2. Model Choice
3. Choose Matching Algorithm
4. Calculate Propensity Score (0-1 Scale)
- 5. Check for Balance**
6. Analysis



Distribution of claims and propensity scores for CNA claim data



Illustrative example of common support and balance (not CNA data)

# Propensity Scoring – Analysis

1. Variable Choice
2. Model Choice
3. Choose Matching Algorithm
4. Calculate Propensity Score (0-1 Scale)
5. Check for Balance
6. **Analysis**

- After cases and controls are established, can now analyze data
- Standard model building rules apply
- For this work, longitudinal data and generalized estimating equations was utilized
  - Robust method for correlated data over time<sup>1-3</sup>
  - Allows for relaxed distributional assumptions<sup>1-3</sup>

1. Liang K, Zeger S.L. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986. 73(1): 13-22.
2. Ghisletta P. An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics*. 2004. 29(4): 421-437.
3. Zeger S.L., Liank. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986. 42: 121-130.

# Propensity Scoring Syntax and Programming

	STATA 10.0	R
Logistic Regression	<i>xi: logistic Y<sub>1</sub> X<sub>1</sub> + ... X<sub>n</sub></i>	<i>logit &lt;- glm(Y<sub>1</sub> + X<sub>1</sub> + ... X<sub>n</sub>, data=mydata, family="binomial")</i>
Propensity Score and Matching	<i>psmatch2 Y<sub>i</sub> <math>\hat{y}</math>, common noreplacement</i>	Packages = MatchIt, Matching, twang, cem, optmatch, PSAgraphics
Check for Balance	Descriptive Statistics <i>psgraph</i>	Descriptive Statistics Package= PSAgraphics
Analysis	User Choice, same model building rules apply;	User Choice, same model building rules apply
Results Validation	Residual analysis, quasi- information criterion <sup>1</sup>	Residual analysis, quasi- information criterion <sup>1</sup>

1: Cui J. QIC program and model selection in GEE analyses. *The Stata Journal*. 2007. 7(2): 209-220.

# Conclusions

- Causal modeling is a more appropriate framework than predictive modeling when the objective is to assess the potential outcomes from alternative actions.
- When possible, planning a randomized experimental controlled design is the best approach to draw conclusions from the effect of alternative actions.
- As most insurance databases are derived from “business as usual conditions”, inferences about treatment effects require special modeling considerations.
- Under certain data conditions, propensity score matching can be used to remove the bias in the estimation of treatment effects from observational data.
- There are a number of matching algorithms one can use after calculating a propensity score which will depend on your data.
- Propensity scoring is a first step to create a balanced cohort; however, one should consider appropriate model building methodologies, as well.