Comparative Effectiveness Research Methods Training

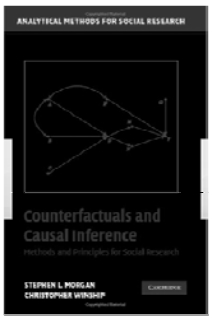
Module 3: Propensity Score Theory

J. Michael Oakes, PhD
Associate Professor
Division of Epidemiology
University of Minnesota
oakes007@umn.edu

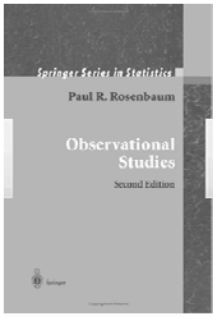
A little (more) about me.

Image from cartoonbank.com removed.

Image description: Picture depicts wealthy guests at a party with focus on 2 women talking to each other with caption, “And it was so typically brilliant of you to have invited an epidemiologist.”

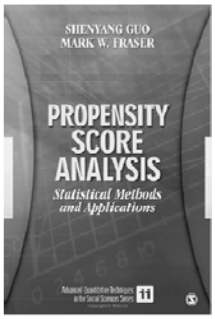


(2007)



(2010)

CCTS | Center for Clinical & Translational Science



(2010)

CCTS | Center for Clinical & Translational Science

Module #3 Outline

1. Review of Core Ideas
2. Confounding
3. Multiple Regression
4. Propensity Score Methods
5. Issues & Assumptions
6. Review
7. Questions

CCTS | Center for Clinical & Translational Science

1. Core Ideas

CCTS | Center for Clinical & Translational Science

More formally, T has a causal effect on Y for person *i* if

$$Y_{i\ T=0} \neq Y_{i\ T=1}$$

But we can only observe one of these outcome for any *i*

At the population level, we use probabilities and assuming exchangeability,

$$\text{Prob}[Y_{T=0} = 1] \neq \text{Prob}[Y_{T=1} = 1]$$

CCTS | Center for Clinical & Translational Science

Ideal Experiment

Randomize a bunch of folks to two different conditions/environments and observe their outcomes at later time.

Assume no problems with:
measurement, attrition, contamination, interference, etc

CCTS | Center for Clinical & Translational Science

Ideal Experiment

The process of allocating subjects to conditions (i.e., randomization) is independent of the outcome variable.

In other words, treatment assignment mechanism does not depend on potential outcomes.

We can identify effects because Tx and Cx groups are exchangeable: we've a defensible counterfactual substitute.

CCTS | Center for Clinical & Translational Science

Analysis of Experimental Data

Simple!

- Bivariate regression (ie, t-test)
- Non-parametric test (eg, permutation test)

CCTS | Center for Clinical & Translational Science

Even if ethical, randomized experiments are very expensive and difficult to conduct.

Most of our work must rely on observational study designs.

Observational designs pose many many many problems for causal inference.

CCTS | Center for Clinical & Translational Science

Central problem of Obs Study

Selection or differences between Tx and Cx groups unrelated to the Tx (ie, confounding)

It's an identification problem:

- Is observed effect attributable to Tx or differences between groups?

CCTS | Center for Clinical & Translational Science

2. Confounding

- A mixing of effects
- Lack of exchangeability between Tx and Cx groups
- Imbalance of background characteristics between groups that effects outcomes

CCTS | Center for Clinical & Translational Science

Confounding

Surgeons who do high-risk surgeries have higher patient mortality rates than family docs who treat sore knees.

- Are the surgeons less skilled or are the groups being treated different?

CCTS | Center for Clinical & Translational Science

Absent randomization the central design and analytic task is to remove influence of confounding and regain exchangeable groups. How?

A. Restrict

B. Match

C. Adjust

CCTS | Center for Clinical & Translational Science

A. Restriction

Design study to collect or keep only data on exchangeable subjects. Delete other data.

- Works well, but how to define exchangeable?
- What values of what variable(s) should you limit analysis to? Age, smoking status, SES, health-risk score, ???

CCTS | Center for Clinical & Translational Science

B. Match

Match Tx and Cx subjects on observed confounding variable (eg, age). This yields conditional exchangeability.

- Works well, but how precise should “age” be? Year, year+month, age_cat?
- Can only match on a few (eg, 1-3) variables before “curse of dimensionality”

CCTS | Center for Clinical & Translational Science

C. Adjust

Use multiple regression to "control for" or "adjust for" potential confounders.

Can simultaneously adjust for many potential confounders.

CCTS

3. Multiple Regression

If the differences between groups is large, the average value applied to each group with adjustment may represent "no man's land", a place where no actual observations exist. Given this scenario, the interpretation of the estimate becomes speculative rather than soundly based. *Heroic modeling assumptions are required*.

William Cochran (1957)

CCTS

Analysis of Experimental Data

$$Y = \alpha + \beta_1 T + \varepsilon$$

$$\hat{\beta}_1 \Rightarrow \bar{\Delta} = \text{ average causal effect}$$

T is (0,1) treatment indicator which, for large samples, is independent of background characteristics by study design (ie, randomization)

CCTS

Absent Randomization

$$Y = \alpha + \beta_1 T + \mathbf{\beta Z} + \varepsilon$$

Covariates, *Z*, serve to
adjust groups for confounding…

CCTS

Absent Randomization

Unless specification of the model, including X, is perfect, bias results

$$\hat{\beta}_1 = \overline{\Delta} + \text{BIAS}$$

CCTS

Simple (mean-centered) regression model

$$y \mid x = \beta_1 x + e \qquad (1)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \qquad (2)$$

CCTS

Substitute (1) into (2) and get

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} x_i e_i}{\sum_{i=1}^{n} x_i^2} \qquad (3)$$

If x and e are correlated (due to confounding), the value of far right-hand term of equation (3) will not go to zero as sample size approaches infinity. The result will be that our treatment effect estimate of will be biased:

$$\beta_1 \neq \beta_1$$

CCTS | Center for Clinical & Translational Science

What goes in *Z* ?

- **Small p-values in bivariate models?**
- **10% rule?**
- **Stepwise procedures?**
- **Everything?**
- **Try stuff (e.g., interactions) until I get "good" p-value for main effect?**
- **Theory?**
- **Too many Z's can overcome data (dimensionality)**

CCTS | Center for Clinical & Translational Science

In regular regression, unless you specify elements of Z in advance you end up capitalizing on chance and your standard errors and corresponding p-values are too small… so conclusions are often wrong.

Regression screening of random data will yield statistically significant "effects" when there are none.

Avoid this!

CCTS | Center for Clinical & Translational Science

4. Propensity Score Methods

An approach to confounder control that better mimics the experimental approach.

Introduced by Rosenbaum and Rubin in 1983

CCTS | Center for Clinical & Translational Science

Propensity Score, *p(z)*

In the analysis of treatment effects, suppose that we have a binary treatment *T*, an outcome *Y*, and background variables Z. The propensity score is defined as the conditional probability of treatment given background variables:

$$p(z) \equiv \Pr(T = 1 \mid Z = z)$$

Propensity score

CCTS | Center for Clinical & Translational Science

Propensity Score, *p(z)*

English: Propensity score is defined as the probability of being treated given a subject's background characteristics.

CCTS | Center for Clinical & Translational Science

Ignorable TAM

Let *Y*(0) and *Y*(1) denote the potential outcomes under control and treatment, respectively. Then treatment assignment is (conditionally) unconfounded if treatment is independent of potential outcomes conditional on *Z*. *This is an assumption!*

CCTS | Center for Clinical & Translational Science

Ignorable TAM

$$TAM \perp Y(0), Y(1) \mid Z$$

CCTS | Center for Clinical & Translational Science

Ignorable TAM

$$TAM \perp Y(0), Y(1) \mid Z$$

$$TAM \perp Y(0), Y(1) \mid p(z)$$

Propensity score

CCTS | Center for Clinical & Translational Science

Ignorable TAM

$$TAM \perp Y(0), Y(1) \mid Z$$

$$TAM \perp Y(0), Y(1) \mid p(z)$$

As with randomized experiments, the expected result of an independent TAM is balance in confounders between treated and untreated groups, thus yielding perfect (?) counterfactual substitutes.

CCTS | Center for Clinical & Translational Science

Simplified Tasks

(1) Set outcome variable (Y) aside

(2) Model treatment/exposure (0,1) with logistic regression or perhaps better models

(3) Calculate/estimate predicted value of exposure from model; this is propensity score!

(4) Use propensity score in analysis to estimate treatment effect

CCTS | Center for Clinical & Translational Science

Model for Treatment?

We want to first model the probability of being treated, not the outcome (eg, health).

Prob(Tx)= (things that predict getting treated)

CCTS | Center for Clinical & Translational Science

Use logistic regression

Logistic regression is like "regular" regression but used when the outcome variable (eg, Y) is not continuous but dichotomous (0,1). It yields predicted probabilities.

Regular regression $Y = \alpha + \beta_1 T + \varepsilon$

Logistic regression $\text{logit}(Y) = \alpha + \beta_1 T$

CCTS | Center for Clinical & Translational Science

$$\text{logit}(T) = \alpha + \boldsymbol{\beta}\mathbf{Z}$$

$$\text{Prob}(T = 1 \mid \mathbf{Z}) = \frac{\exp(\alpha + \boldsymbol{\beta}\mathbf{Z})}{1 + \exp(\alpha + \boldsymbol{\beta}\mathbf{Z})}$$

CCTS | Center for Clinical & Translational Science

So what goes in **Z**?

Use all potential predictors of Tx, except those that are outcomes of Tx.

You can "play" with specification since you've set outcome Y aside. Greatly reduced threat of capitalizing on chance.

CCTS | Center for Clinical & Translational Science

The propensity score is nothing more than the predicted probability of being treated, which comes directly from the logistic regression model.

Each observation in your data will have a propensity score variable with range 0-1

Some observations may have been treated (T=1) with low propensity score of 0.01, while others not treated (T=0) with high propensity score of 0.90

CCTS

Uses of propensity scores

(1) Use as "regular" covariate in regression model

(2) Stratify/classify data by range of p-score and estimate effects within and then average

(3) Match those actually treated and those not actually treated on their p-score

(4) Use as a weight in more sophisticated models

CCTS

Propensity score matching

Match treated person to their counterfactual, which is a non-treated person with a similar propensity score as treated person.

Calculate difference, *d* , from the observed outcome, Y, for each index person and their matched counterfactual

Calculate average d across all observations

$$\bar{d} = ATT \approx ACE = ATE$$

CCTS

Matching methods

- **Nearest Neighbor**
  - match treated to counterfactual with closest p-score
- **Nearest Neighbor within Caliper** (I like best)
  - match treated with closest within range (ie, caliper)
- **Kernel, Local Linear, Manhalanobis, Optimal**

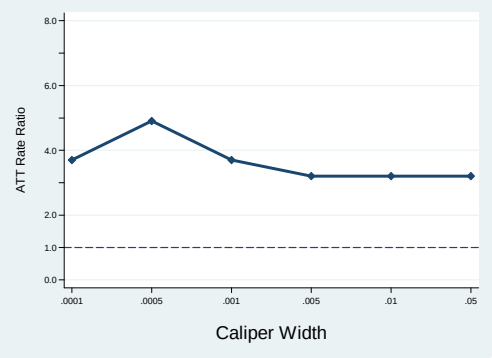
CCTS

How wide a caliper?

Suggestion is no more than 25% of the standard deviation of your observed propensity scores.

$$\varepsilon \leq 0.25 * \sigma_{pscore}$$

CCTS



CCTS

How do you know matching worked?

Assess balance between treatment and control group prior to treatment.

If balance then ignorable TAM, by assumption!

CCTS | Center for Clinical & Translational Science

Estimate Difference in Balance

Estimate difference in means b/w Tx and Cx groups for covariate X before and after matching.

Best to standardized difference so that you can compare them, and assess decrease in differences, which would be an increase in balance wrt that covariate.

CCTS | Center for Clinical & Translational Science

Before matching

$$d_X = \frac{\left| M_{Xt} - M_{Xp} \right|}{S_X}$$

d_X is that absolute standardized difference in covariate means
M_{Xt} is the mean of variable X for the treatment group
M_{Xp} is the mean of variable X for the potential control group
S_X is the pooled standard deviation of the groups

CCTS | Center for Clinical & Translational Science

Gou & Fraser, p. 157

After matching

$$d_{Xm} = \frac{\left|M_{Xt} - M_{Xc}\right|}{S_X}$$

d_{Xm} is that absolute standardized difference in covariate means after matching
M_{Xc} is the mean of variable X for the control group after matchng

CCTS | Center for Clinical & Translational Science

Balance Assessment

Absolute reduction in imbalance due to matching = $d_X - d_{Xm}$

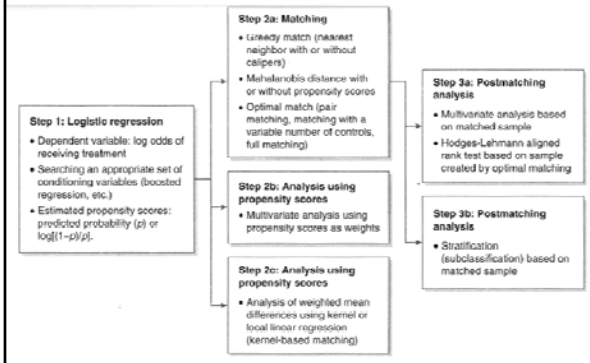
% reduction in imbalance due to matching = $100\left[\frac{(d_X - d_{Xm})}{d_X}\right]$

CCTS | Center for Clinical & Translational Science

Steps to Propensity Score Matching

1. Fit logistic regression of treatment (not outcome!)
2. Estimate propensity score for each person
3. Match across exposure on estimated p-score
4. Throw away off-support observations
5. Assess balance between groups
6. Re-estimate pscore if balance not obtained
7. Estimate causal model **(eg, t-test or other methods)** with outcome
8. Bootstrap standard error of causal effect

CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science

Gou & Fraser, Fig 5.1

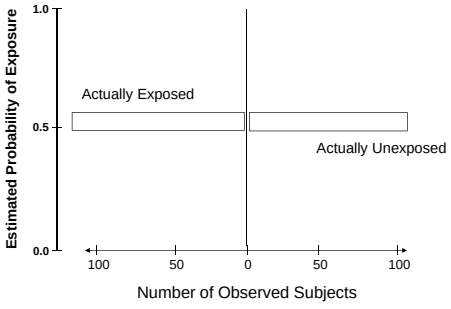
Off-support?

Rubin – If complete separation in propensity score? "You can say nothing about causal effects."

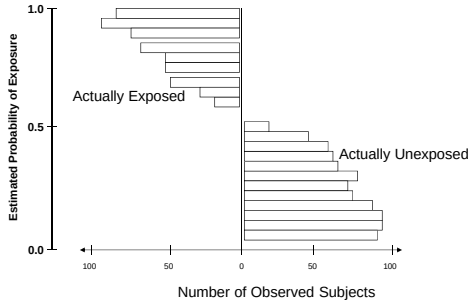
Rosenbaum – Sharply distinct treatments that could happen to anyone.

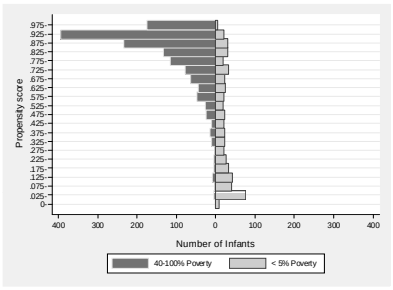
If your substitute (ie, comparison group) does not reflect treatment group, then all inferences are based on (off-support) model assumptions.

CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science





Propensity of Am Ind living in high-poverty Mpls neighborhoods

Cutting Edge?

- Estimation of propensity scores (CART?)
- Estimation of "bounds" around estimated treatment effects
- Propensity score model calibration
  - Validation sample
  - Qualitative info
- SE estimation
- Imputation for missing covariates
- Inducing confounding by misspecifying Z

5. Issues & Assumptions

1. Exposure model specification
2. Matching algorithm (nearest, greedy, etc)
3. Unobservables!
4. Missing values of covariates
5. Matching with replacement (ie, imputation)
6. Precision/bias tradeoff with respect to "support"
7. Which treatment effect estimator (ATT, ITT, ACE…)
8. Clustering
9. SUTVA violations

CCTS | Center for Clinical & Translational Science

Because they require us to think about the ideal experiment we would have liked to have conducted, propensity score methods are a better tool than multiple regression. Setting aside the outcome variable, Y, until it's time to assess differences between observed outcomes and counterfactual substitutes, is an invaluable addition to the practice of applied research.

CCTS | Center for Clinical & Translational Science

Remember…

- We *impose* a causal model on the world/phenomena
  - It's a cognitive thing… a belief subject to scientific scrutiny
- We are bombarded with "data" and so must select some for consideration
  - Humans are excellent at confirmation bias… finding "data" to support our belief; we struggle with data that undermines our beliefs.

CCTS | Center for Clinical & Translational Science

Add this watermark

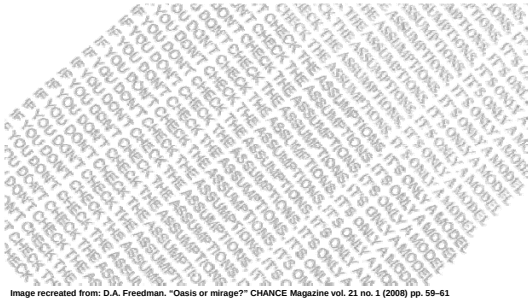


Image recreated from: D.A. Freedman. "Oasis or mirage?" CHANCE Magazine vol. 21 no. 1 (2008) pp. 59–61

CCTS | Center for Clinical & Translational Science

5. Review

I. Review of Core Ideas
   a. Causal inference
   b. Ideal experiments
   c. Effect identification

II. Confounding
   a. An imbalance
   b. TAM not ignorable
   c. Restriction
   d. Matching
   e. Adjustment

III. Multiple Regression
   a. Model specification
   b. Model Y

IV. Propensity Score Methods
   a. Mimic ideal experiment
   b. Model Tx, not Y
   c. Use as covariate
   d. Subclassify
   e. Match (several ways)
   f. Weight
   g. Balance assessment
   h. Support assessment
   i. Cutting edge issues

V. Assumptions & Issues
   a. Black-box mechanisms
   b. Unobservables

CCTS | Center for Clinical & Translational Science

6. Questions

1. Why is confounding a problem for causal inference?
2. What is a treatment assignment mechanism?
3. Name three uses of propensity scores
4. When done correctly, do propensity score methods address unobservables?
5. What does "off-support" mean?

CCTS | Center for Clinical & Translational Science

References & Resources

1. Cochran WG. Analysis of Covariance: Its Nature and Uses. *Biometrics*.1957;13:261-281.
2. Freedman, DA. Oasis or mirage? CHANCE Magazine. 2008; 21: (1): 59–61.
3. Guo S, Fraser MW. Propensity Score Analysis: Statistical Methods and Applications. Thousand Oaks, CA: SAGE Publications; 2010:157.
4. Guo S, Fraser MW. Propensity Score Analysis: Statistical Methods and Applications. Thousand Oaks, CA: SAGE Publications; 2010:Fig 5.1.