

# Causal Inference: Recent Methodological Developments for Estimating Causal Effects

Donald B. Rubin  
Harvard University

# Statistics 140, W.G. Cochran, 1968

## Classical Experimental Design

Clear separation between:

- “Science” (and ...)

	X	Y(0)	Y(1)
1			
Units			
N			

X = Covariates unaffected by treatments

Y(0) = Potential outcomes under control treatment

Y(1) = Potential outcomes under active treatment

Notation due to Neyman (1923) in context of randomized experiments

# Classical Experimental Design

- Clear separation between Science and what we do to learn about Science:
  - Randomized assignment of treatments
  - $W$  = Vector of  $N$  treatment indicators
- This distinction and the consequential clarity should be maintained and should not be forgotten when designing observational studies
- Cochran & Stat 140, great advantage to start here

# Causal Inference is a Missing Data Problem

RCM (Holland, 1986) for work in 1970's

Maintains critical distinction from experimental design

- Same notation for science whether try to learn about it from randomized experiment or observational study.
- Earlier, “ $Y_{obs}$ ” used in nonrandomized studies with  $W$  a predictor in regressions, paths, arrows

$$Y_{obs} = \{Y_{obs,i}\}$$

$$Y_{obs,i} = W_i Y_i(1) + (1 - W_i) Y_i(0)$$

Entangles Science and assignments

# Assignment Mechanism

Creates missing potential outcomes

$$\Pr(W|X, Y(0), Y(1)) \quad (\text{AM})$$

Randomized experiments are:

unconfounded:  $\text{AM} = \Pr(W|X)$ , and

probabilistic:  $1 > \Pr(W_i|X_i) > 0$  for all  $i$

Earlier, words describing AM but no explicit mathematical notation or expressions (e.g., Roy, 1953)

Same is true for potential outcomes before Neyman's (1923) notation (e.g., Fisher, 1918)

# Design Observational Studies to Approximate Randomized Trials

1. Hide outcome data until the design phase is complete
  2. Think very carefully about decision makers and the key covariates that were used to make treatment decisions
  3. If key covariates are not observed or very noisy, usually best to give up and seek better data source
  4. Find subgroups (subclasses or matched pairs) in which the treatment and control groups have balance – essentially the same distribution of observed covariates
    - Not always possible to achieve balance
    - Inferences are limited to subgroups where balance is achieved
- #1 - #4 combine to create an objective design that approximates a randomized trial in each subclass that is balanced with respect to observed covariates

## Cochran (1968) – Illustrative Example with One Key Covariate

- Population: Male smokers in U.S.
- Treatment = cigar/pipe smoking
- Control = cigarette smoking
- Outcome = death rate/1000 person years
- Decision maker is the individual male smoker
- Reason for a smoking male to choose cigarettes versus cigar/pipe?
- **Age** is a key covariate for selection of smoking type for males

# Subclassification to Balance Age

- To achieve balance on age, compare:
  - “young” cigar/pipe smokers with “young” cigarette smokers
  - “old” cigar/pipe smokers with “old” cigarette smokers
- Or better, compare:
  - Young, middle aged, old
  - Even more age subclasses
- Design phase, no outcome data, objective:
  - Approximates a randomized trial within subclasses
- Now look at outcome data

# Comparison of Mortality Rates for Two Smoking Groups in U.S.

	Cigarette Smokers	Cigar/Pipe Smokers
Mortality Rates per 1000 person-years, %	13.5	17.4
Adjusted Mortality Rates using subclasses, %		
2 age subclasses	16.4	14.9
3 age subclasses	17.7	14.2
9-11 age subclasses	21.2	13.7

Source: Cochran WG. The effectiveness of adjustment of subclassification in removing bias in observational studies. Biometrics 1968; 24:295-313.

But 20 four-class covariates  $\Rightarrow$  over million million subclasses

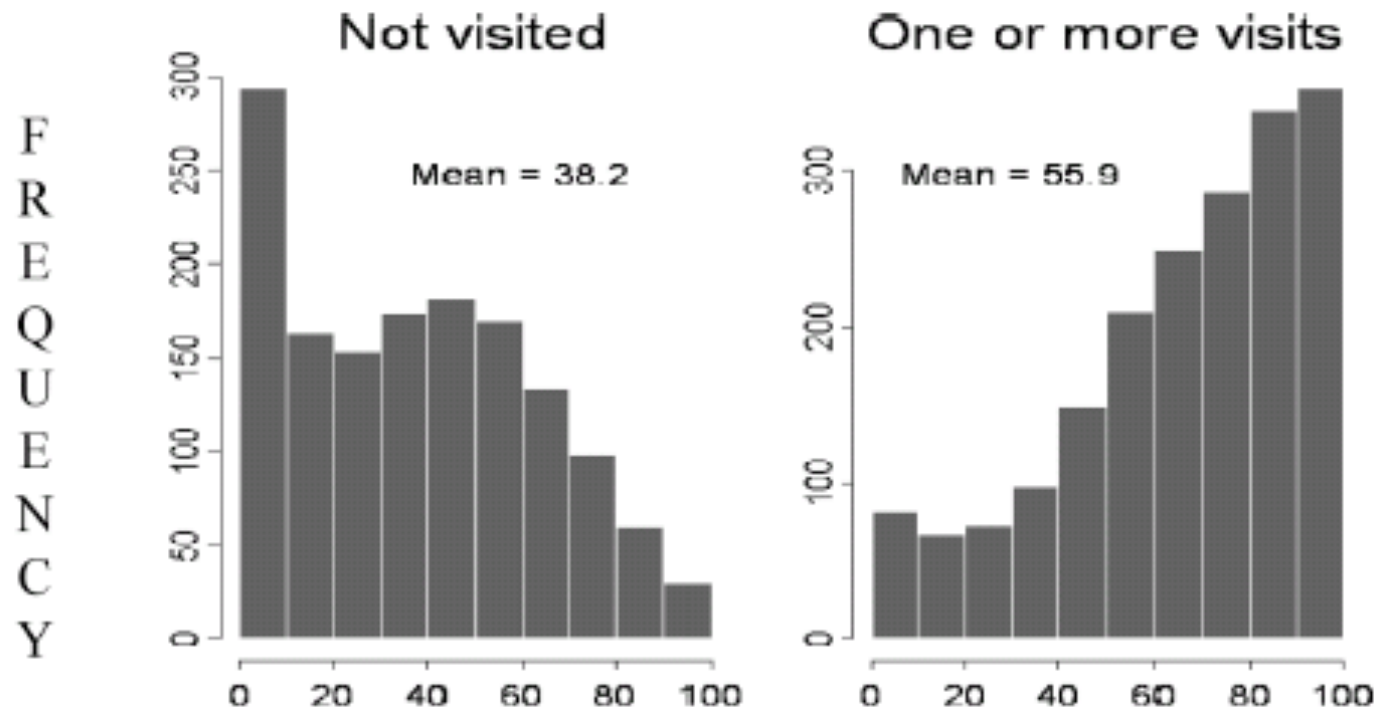
# Propensity Score Methods

- Rosenbaum and Rubin. “The Central Role of the Propensity Score in Observational Studies.” *Biometrika* 1983.
- Observational study analogue of complete randomization
- The propensity score is the probability of treatment versus control as a function of observed covariates
  - Model the reasons for treatment versus control at the level of the decision makers
  - For example, logistic regression model to predict cigarette versus cigar/pipe smoking with age, education, income, etc. as predictors
- Then subclassify (or match) on the propensity score as if it were the only covariate, e.g., 5-10 subclasses
- If correctly done, this creates balance within each subclass on **ALL** covariates used in estimating the propensity score

# Diagnostics for Accessing Balance

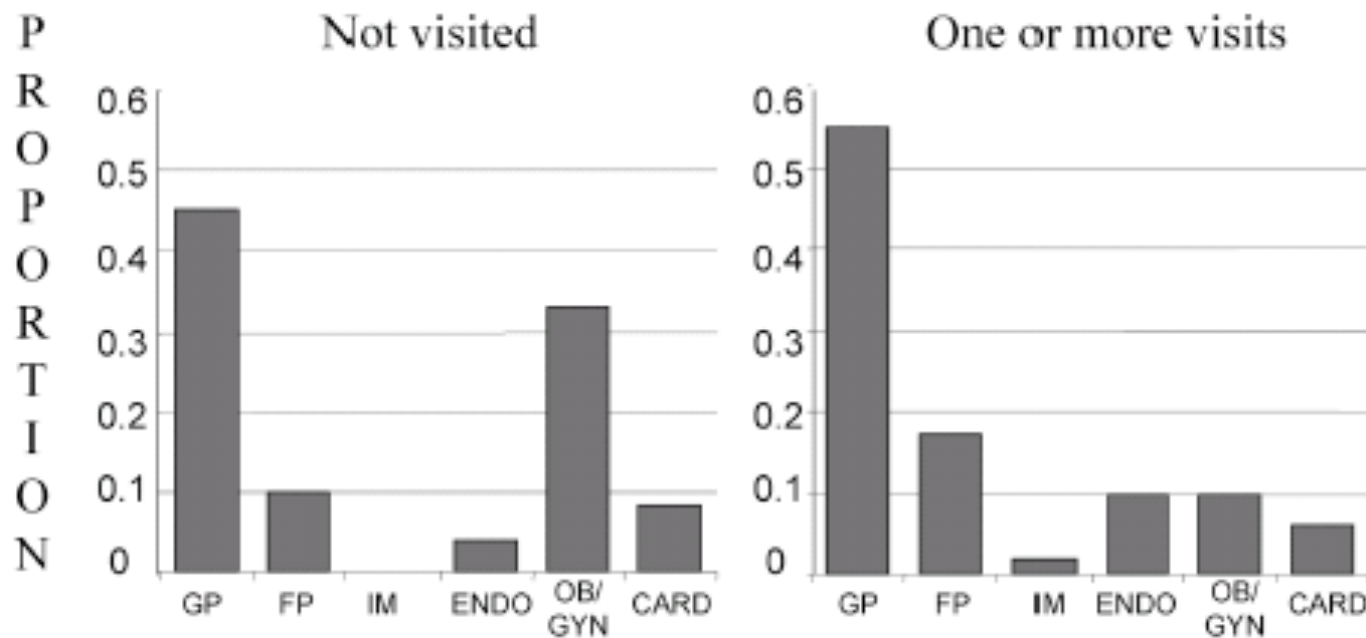
- Assessing balance simpler in large samples, just as with randomized experiments
- To illustrate diagnostics, use a marketing application that involved a weight loss drug
- Units = doctors
- Treatment = sales rep “visits” doctor to discuss
- Control = no visit
- Decision-makers = sales reps
- Key covariates = prior Rx's, medical specialty, years in practice, size of practice, etc.

# Histograms for background variable: Prior Rx Score (0-100) at Baseline



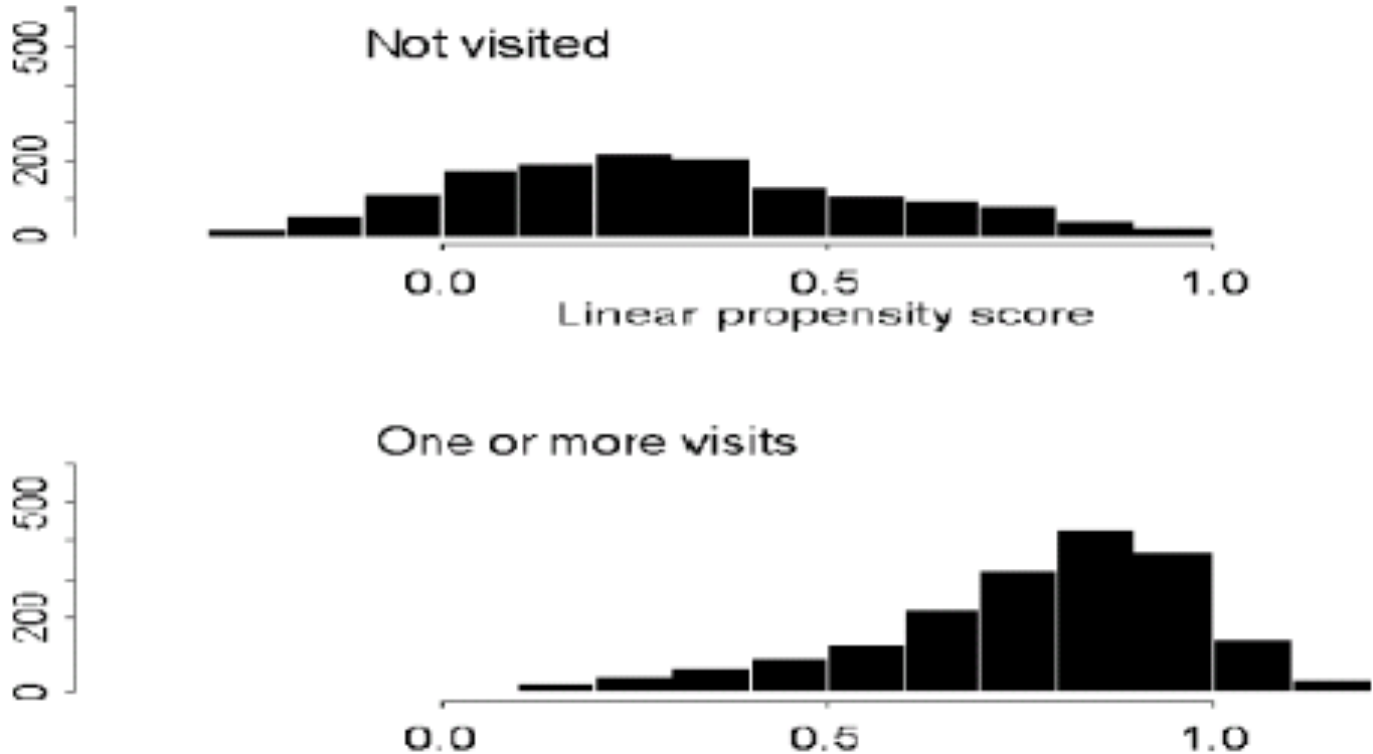
Source: Rubin DB and Waterman RP. Estimating Causal Effects of Marketing Interventions Using Propensity Score Methodology. *Statistical Science* 2006; 21(2):206-222.

# Histograms for background variable: Specialty



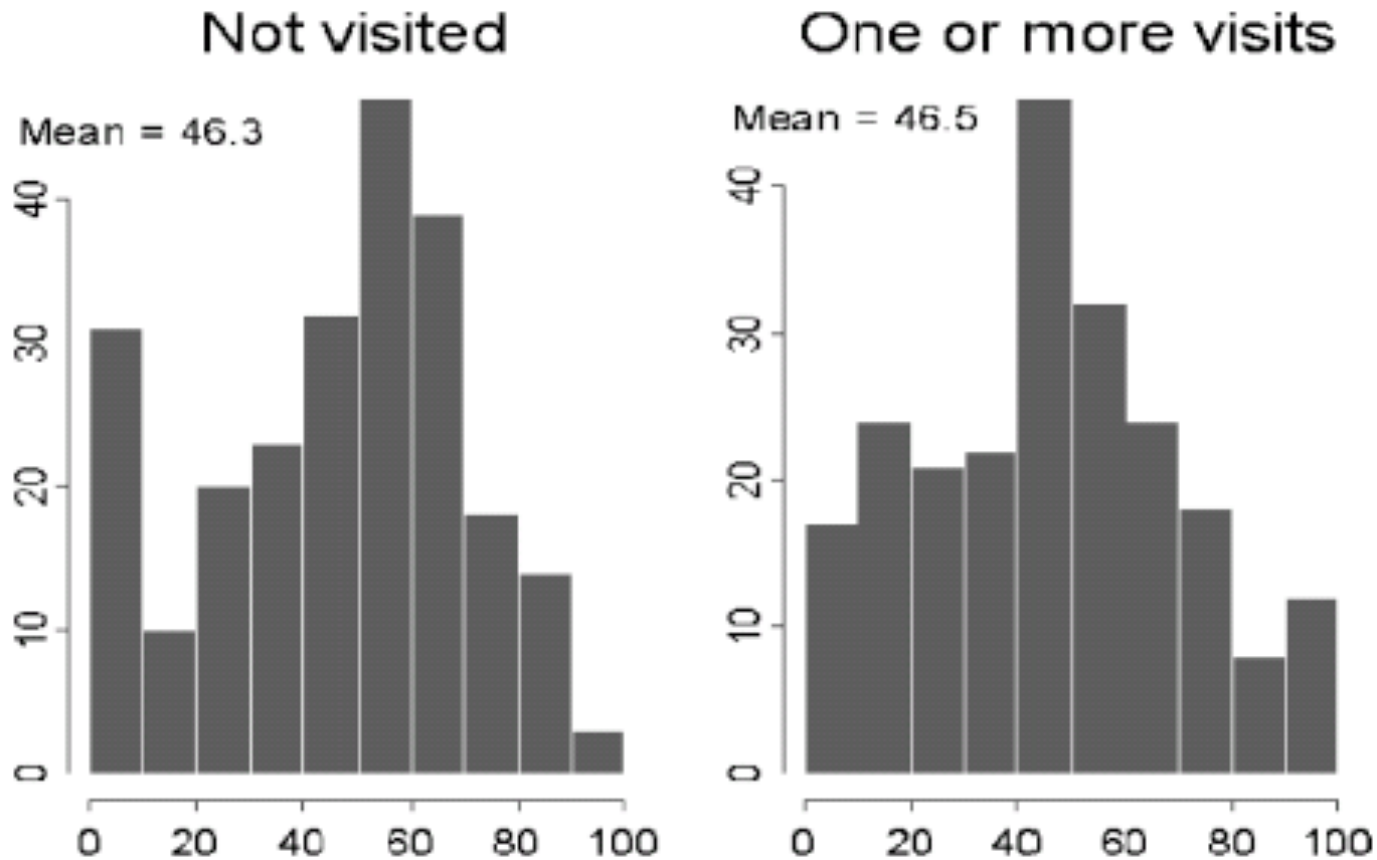
Source: Rubin DB and Waterman RP. Estimating Causal Effects of Marketing Interventions Using Propensity Score Methodology. *Statistical Science* 2006; 21(2):206-222.

# Histograms for summarized background variables: Linear Propensity Score



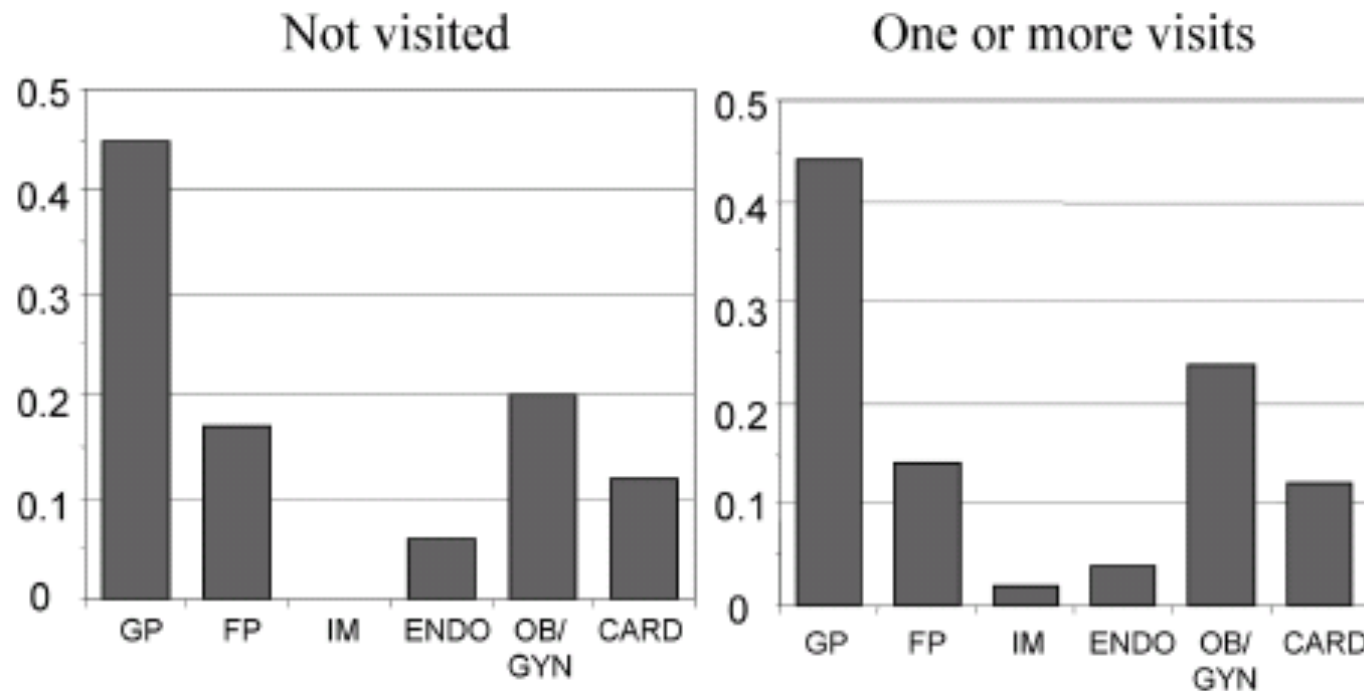
Source: Rubin DB and Waterman RP. Estimating Causal Effects of Marketing Interventions Using Propensity Score Methodology. *Statistical Science* 2006; 21(2):206-222.

# Histograms for a variable in a subclass of propensity scores: Prior Rx Score



Source: Rubin DB and Waterman RP. Estimating Causal Effects of Marketing Interventions Using Propensity Score Methodology. *Statistical Science* 2006; 21(2):206-222.

# Histograms for a variable in a subclass of propensity scores: Specialty



Source: Rubin DB and Waterman RP. Estimating Causal Effects of Marketing Interventions Using Propensity Score Methodology. *Statistical Science* 2006; 21(2):206-222.

# Marketing Example: Achieved Balance

- Within each narrow subclass of propensity scores, the treatment and control groups will be as balanced as if randomly divided
- Claim: This holds for all subclasses in which there are both treated and control subjects, and holds for all covariates that were used to estimate the propensity score
- Works best when the propensity score subclasses have large sample sizes and are relatively narrow
- Five to ten propensity score subclasses often fully adequate to balance all covariates
- No outcome data used in the design stage

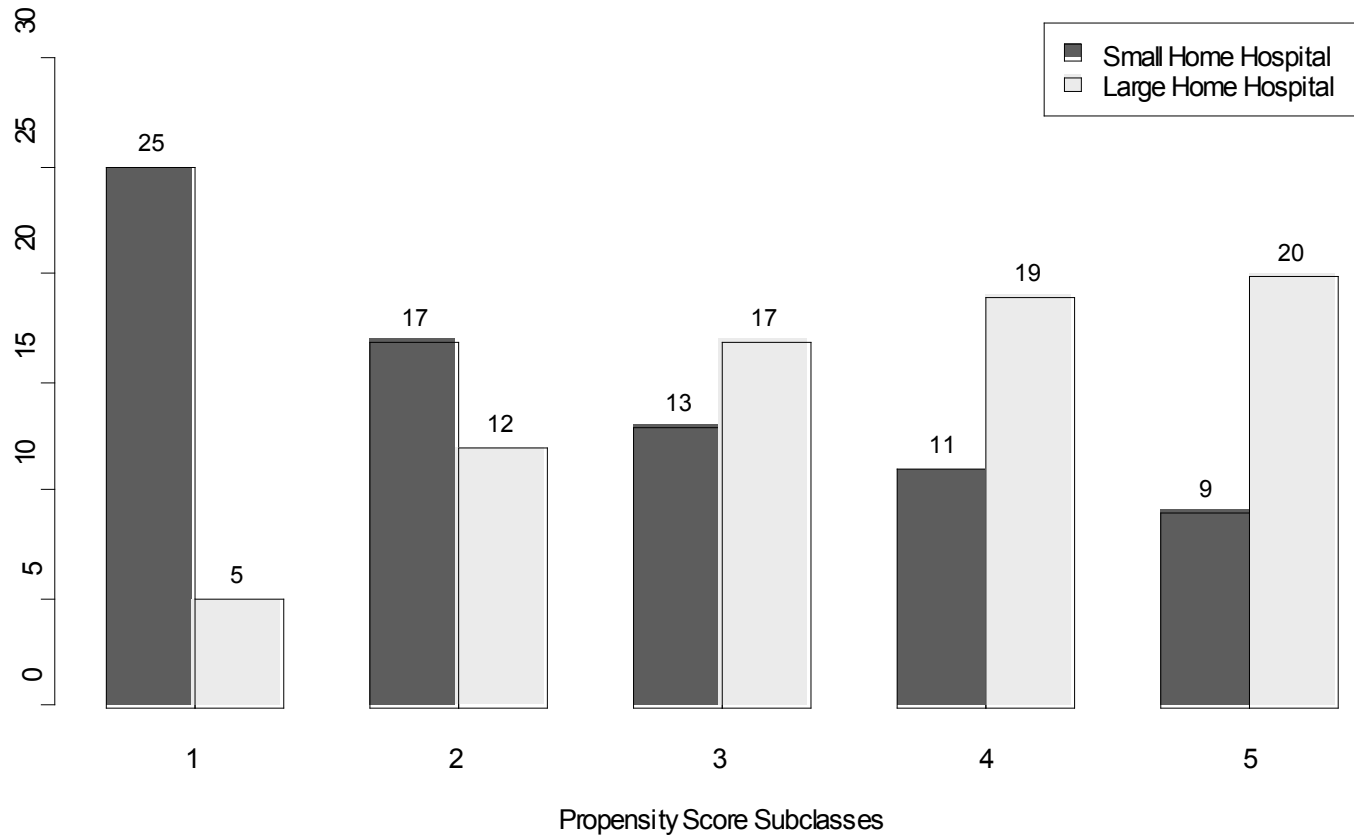
## Using a More Complex Template: Randomized Block Experiment with Noncompliance

- Causal effect of Large versus Small treating hospitals on cardia cancer survival
- Dataset from Karolinska Institute, Stockholm
- Medical researchers accept unconfounded assignment of “home (diagnosing) hospital” type, but NOT treating hospital type because of self-selected transfers
- Consider transfers between hospital types as a form of noncompliance with assignment

# Using a More Complex Template: Randomized Block Experiment with Noncompliance

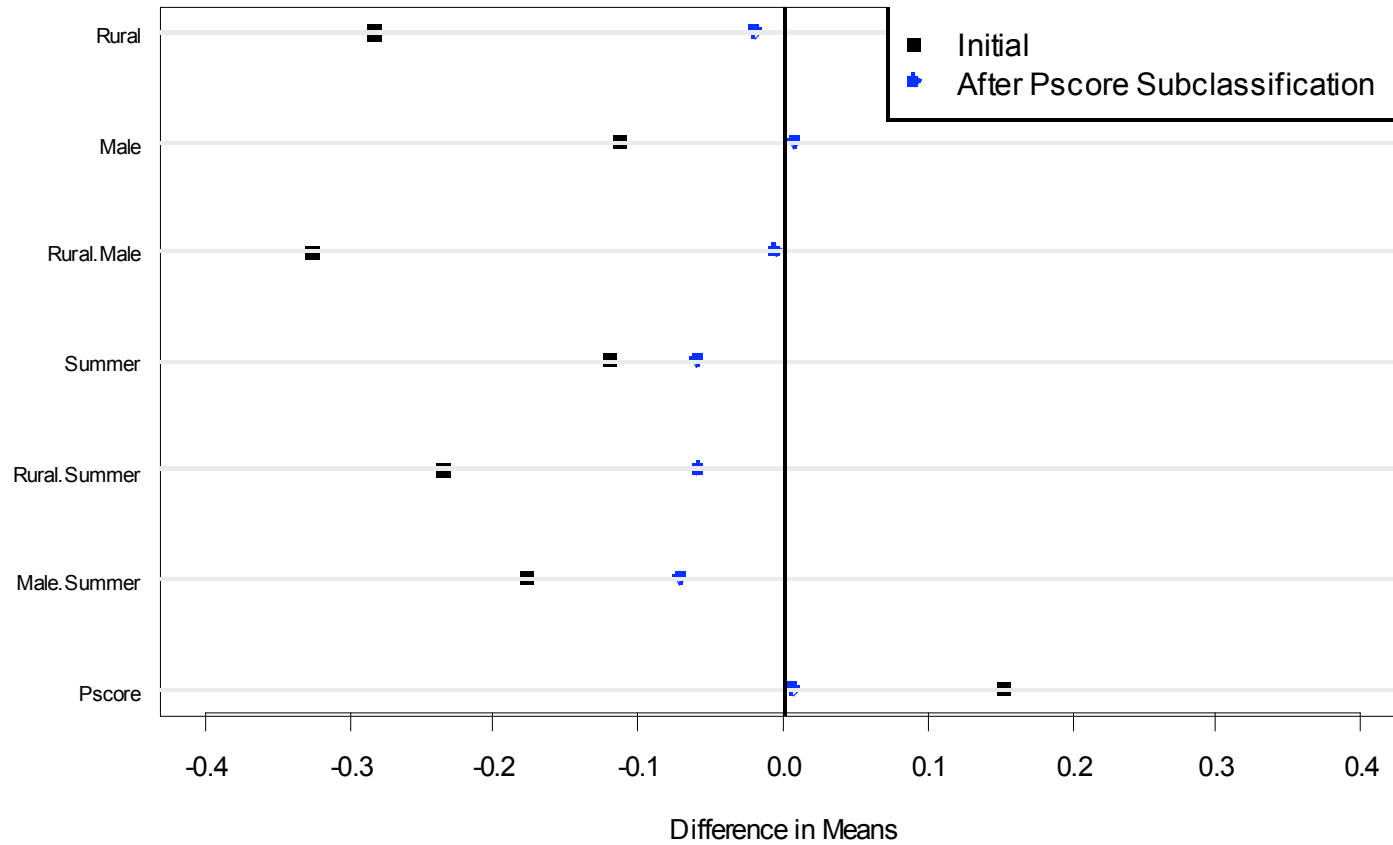
- Design has two distinct phases
- Phase 1, no outcome data available:
  - Propensity score analysis to approximate randomized block experiment for home hospital type
  - Ensure subclassification can create balance on covariates for large and small home hospital types
- Phase 2, uses intermediate outcome data on transfers:
  - Outline of the analysis for estimating causal effect of treating hospital type
  - Ensure within each subclass that there appear to be compliers who are treated in both and large and small treating hospitals

**Figure 5.1: Cardia Cancer, Number of People, Subclassified by Propensity Score**



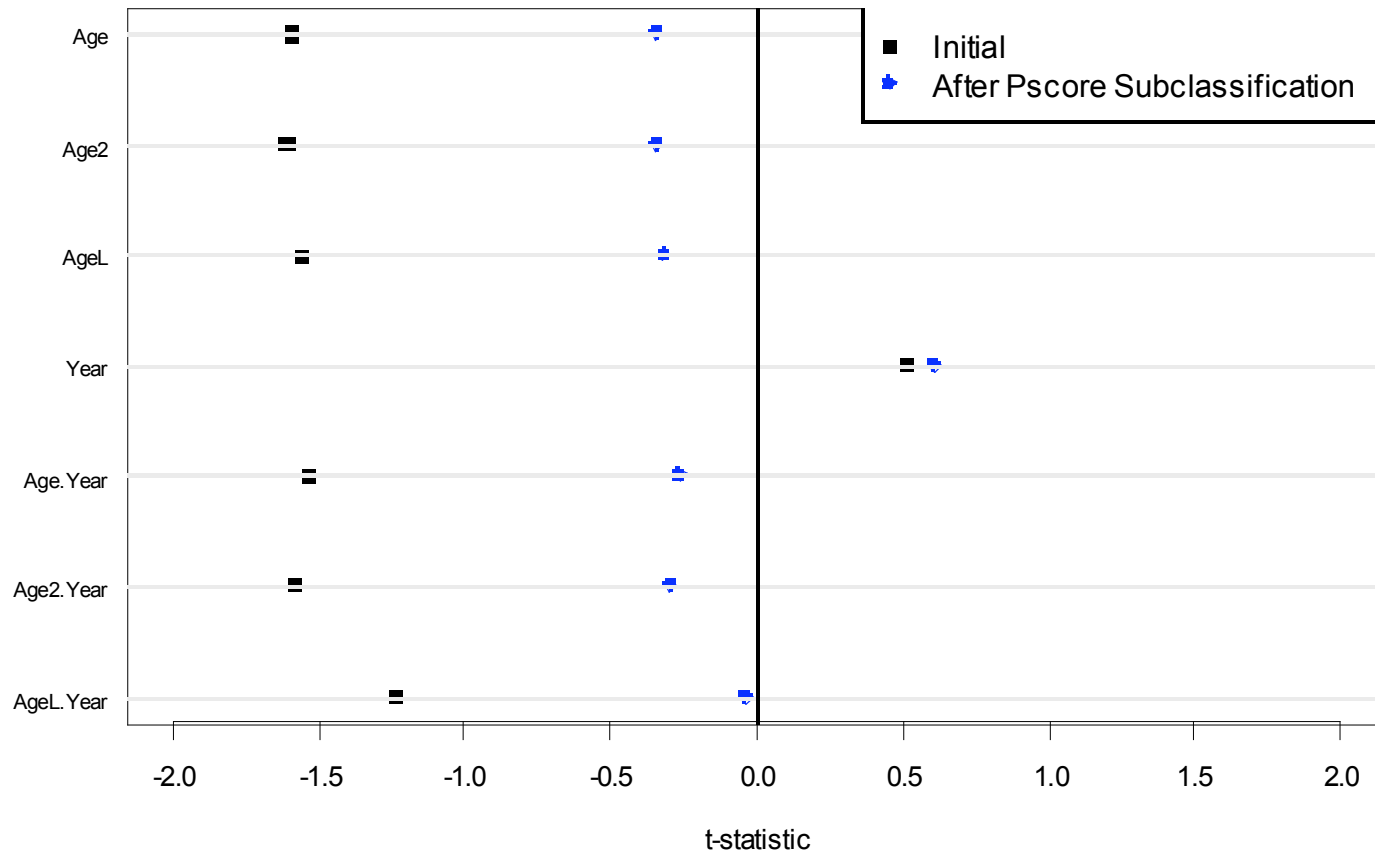
Reference Rubin, D.B. For Objective Causal Inference, Design Trumps Analysis. Annals of Applied Statistics, 2008.

**Figure 5.2: Cardia Cancer, Difference in Means for Binary Covariates and Pscore**



Reference Rubin, D.B. For Objective Causal Inference, Design Trumps Analysis. Annals of Applied Statistics, 2008.

**Figure 5.3: Cardia Cancer, t-statistics for Continuous Covariates**



Reference Rubin, D.B. For Objective Causal Inference, Design Trumps Analysis. Annals of Applied Statistics, 2008.

## Outline of Analysis within Each Subclass

- Critical that we anticipate compliers in both large and small treating hospitals within each subclass
- Monotonicity assumption = no defiers; medically very plausible

$$ITT = \pi_{LS} ITT_{LS} + \pi_{SS} ITT_{SS} + \pi_{LL} ITT_{LL}$$

$$CACE \equiv ITT_{LS} = \frac{1}{N_{LS}} \sum_{i \in LS} (Y_i(L) - Y_i(S)),$$

$$ITT = \pi_{LS} ITT_{LS}, \quad ITT_{LS} = ITT / \pi_{LS}.$$

## Method of Moments Estimates of the Number of Compliers Treated in Large and Small Hospital Types Under Monotonicity

<b>“Assigned”/Randomized Home Hospital Type</b>	<b>Treating Hospital Type</b>	<b>Approximate N in LS Principal Stratum Subclass</b>				
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<i>h</i>	<b>T</b>					
<i>ℓ</i>	<b>L</b>	3	3	11	9	5
<i>s</i>	<b>S</b>	14	5	8	5	2

Table 5.3: Cardia Cancer: Observed Counts in Observed Groups and Approximate Counts in Principal Strata Under Monotonicity Assumption – Subclass 3

(1)		(2)	(3)	(4)	(5)	(6)	
“Assigned”/Randomized Home Hospital Type		Treating Hospital Type $T$	#	Underlying Principal Strata:	Approximate Proportion in Population in Principal Strata	Approximate N in $LS$ Principal Stratum	
$h$	#			$h =$ $\ell$ $s$			
(1)	$\ell$	17	L	17	L L	38%	11
(2)			S	0	L S	62%	
			S	0	S S	0%	
(3)	$s$	13	L	5	L L	38%	
(4)			S	8	S S	0%	
					L S	62%	8

Reference Rubin, D.B. For Objective Causal Inference, Design Trumps Analysis. Annals of Applied Statistics, 2008.

# Conclusions: Objective Observational Study Design-Influenced by Cochran's Wisdom

- Should approximate a randomized experiment
  - No ultimate outcome data used or examined – “prospective”
  - Carefully consider decision-makers and the covariates used to make treatment assignments
  - If dataset is missing key covariates, usually do NOT continue
  - Use propensity score estimation to help create subclasses or matched pairs that achieve “balance” on covariates
  - Balance means treated and control subjects have distributions of covariates that are at least as similar as if they had been randomized into treatment and control
  - Analysis takes place within each subclass/matched pair, and then answers are combined across subclasses/pairs