Extending Propensity Scores: Polychotomous Outcomes, Sample Selection, and Incorrect Model Specification

Michael A. Posner
Boston University School of Public Health
mposner@bu.edu

Case Western Reserve University
Center for Health Care Research and Policy
October 22, 2004

1

Outline

- Observational studies - the good and the bad
- Propensity scores - an overview
- Extending propensity score methods
  – Polychotomous outcomes
  – Weighted regression instead of sub-sampling
- The errors of ignoring selection bias
  – Propensity score simulations
- Future work

2

Observational Studies

- Observational Study: non-random assignment to treatment
- Why use observational studies?
  – Randomized trials impossible (ethics, design)
  – Randomized trials produce biased results
  – Prevalence of administrative data
- …and why not *always* use them?
  – Selection Bias
  – RCTs are the "gold standard"
  – RCTs deal with unobserved covariates

3

An Example…

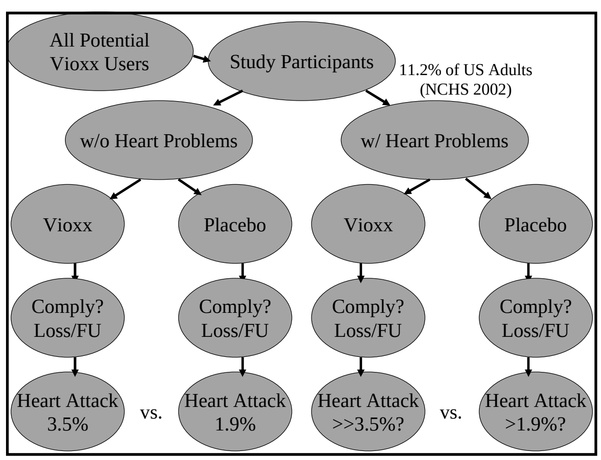
- "Parachute use to prevent death and major trauma related to gravitational challenge: a systematic review of RCTs" BMJ, December 2003.
- Objective: To determine if parachutes are effective in preventing trauma and death
- Results: No RCTs found on parachute use. Studies of free fall do not show 100% mortality.
- Conclusion: "The effectiveness of parachutes has not been subjected to rigorous evaluation using RCTs. [Some] criticize the adoption of interventions evaluated using only observational data. Everyone might benefit if the radical protagonists…organized and participated in a double blind, randomized, placebo-controlled, crossover trial of the parachute."

4

RCT Bias

- Psychiatry
  – Excludes patients with multiple symptoms
- Vioxx - New York Times article (Oct 2004)
  – "Good Riddance to a Bad Drug" by Eric J. Topol, Chair, Cardiovascular Medicine, Cleveland Clinic
  – Merck's new study heart attack/stroke rate:
    - 3.5% on Vioxx
    - 1.9% on Placebo
  – Underestimate since Merck's trial excluded anyone with known heart disease

5



The Problem with Observational Studies: Selection Bias

- What is selection bias?
  - Those who are in the “treatment” group are not the same as those who get the “control” group
- Uneven baseline covariate distributions
- Need to address this before making valid inferences

7

Simple Illustration

- Mammography for women over 70 years old
  - outcome = early detection of breast cancer
  - treatment = mammography
  - covariate = income
- Assume (hypothetically) that rich women typically get mammography, and poor women typically do not
  - Effect of mammography/income confounded
- Does covariate regression address this?

8

Propensity Score Motivation

- Goal - Identify groups that have similar baseline characteristics, then compare outcome of interest by groups
- Cochran (1968) - Remove bias by stratifying on confounders (90% reduction with 5 strata)
- Rosenbaum and Rubin (1983) - Combine multiple confounders into one measure (propensity score) and use it to stratify

9

Propensity Score Method

- Calculate the conditional probability, or “propensity,” of being in the treatment group (given your covariates)
  - propensity = P(treated | covariates)
- Create groups that are similar to each other in their propensity score by sub-sampling
  - Split into percentiles
  - Matching (nearest, logit, or Mahalanobis)
  - Matching with calipers
  - Greedy algorithm (Parsons)
- Analyze using typical statistical techniques

10

(Some) Applications of PS

- Myocardial Infarction/Coronary Artery Disease
  - Numerous articles (Am J Med, NEJM, Ann Thoracic Surg, JAMA, etc.)
- Psychiatry
  - Lavori, et al - Stats Med 1988, Neuro 1992, J Psy Res 1998
- Mammography
  - Posner, et al - HSORM 2002
- Domestic Abuse
  - Berk, Newton - Am Soc Rev 1985
- Income/Tax Statistics
  - Czajka, et al - J Bus & Econ Stat, 1992
- Gender Bias
  - Zanutto - ASA Proceedings, 2002

11

Recent Research

- Comparing observational studies and RCTs
  - Similar results, when biases are correctly addressed
  - Literature Reviews/Meta-analyses
    - Benson, Hartz. NEJM, 2000
    - Concato, et al. NEJM, 2000
    - Ioannidis, et al. JAMA, 2001

12

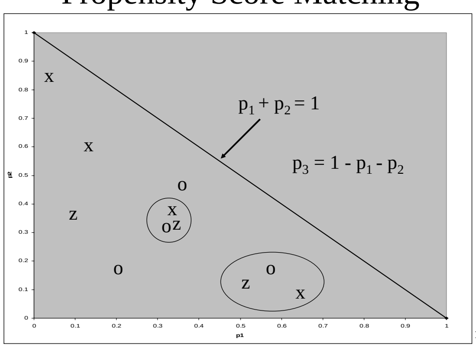
Current Research

- Understanding Selection Bias
  - Hernán, Hernández-Diaz, Robins (Epidemiology, 2004)
- Polychotomous Outcomes (Dose-Response)
  - Imbens (Biometrika, 2000)
    - Mental Health Services - Foster (Medical Care, 2003)
  - Imai and Van Dyk (JASA, 2004)
- Matching to Multiple Controls
  - Stuart and Rubin (in progress)
- Missing Data
    - D'Agostino, Rubin, Others via EM, ECM algorithms
- Drop-out prior to completion of trial

Polychotomous Treatments

- Multiple treatment groups
  - Drug A vs. B vs. C
  - Dose-response
- Propensity score analysis of polychotomous treatments
  - Pairwise Logistic Regression
    - k-1 logistic regression equations with one group as the constant reference group
  - Multinomial Regression

Polychotomous Multinomial Propensity Score Matching



Polychotomous Propensity Scores Example: Respite Analysis

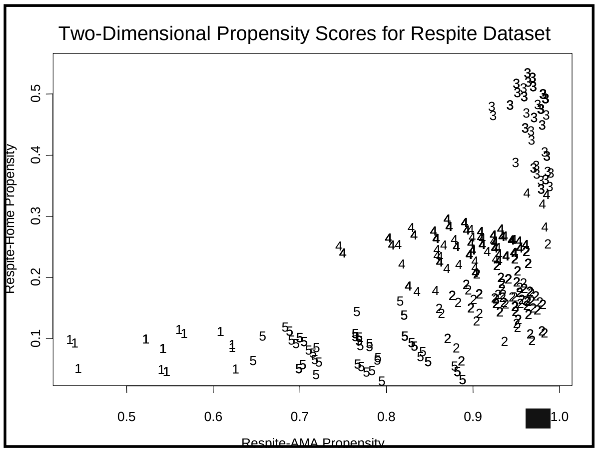
- Ongoing Research
  - w/ Boston Health Care for the Homeless
- Respite unit takes homeless patients not ready to return to streets after hospitalization
  - Treatment groups: respite, home, AMA
- Outcome = 90-day hospital readmission
- Assignment to respite likely associated with patient characteristics, thus we need to consider dealing with selection bias

Selection Bias Susceptibility

- All variables associated with exposure (respite vs. AMA vs. home) [X^2 $p < .05$]
  - Age (young, middle, old)
  - Race (White, Black, Other)
  - Alcohol Use (yes/no)
  - Drug Use (yes/no)
  - Comorbidity (DCG RR <.5, .5-1.5, >1.5)
  - Length of Stay (short, medium, long)
- Many ways of showing or testing covariate balance (Love, JSM 2002)

Methods: Two Logistic Regressions

- Calculate the logistic regression equation for the propensity of being in respite:
  - 1. P(Respite | Respite or AMA)
  - 2. P(Respite | Respite or Home)
- Apply these equations to produce two propensity scores for each observation
- Use this two-dimensional vector to quasi-randomize using cluster analysis



Results: Two Logistic Regressions

Weighted Regression

- Weighted by the ratio of the largest sample size in that cluster, divided by the sample size in that treatment group
- Weights are then standardized to equal total sample size (default in SAS)
- Each treatment group is given equal weight in the final analysis

Example of Weight Calculation

	cluster 1			
	Home	AMA	Respite	
sample size	15	3	1	←Total=19 (samp size)
total weight	15	15	15	←Maximum Samp Size
weight/obs	1	5	15	←Weight/sample size
reduced wt	0.422	2.111	6.333	←19/45 * weight/obs
weighted total	6.333	6.333	6.333	←Total=19, equal wts

Similar calculations can be done in other clusters

Results: Comparing Methods

		OR*	
		0.61	
		1.09	
		0.33	
		0.60	
		0.56	
		1.47	

*Odds Ratio uses “home” as reference group

Sub-Sampling Concerns

- Sub-sampling results are unstable and very sensitive to randomization (small n/group)

		OR	
		0.33	
		0.60	
		0.58	
		1.49	
		invalid	results
		invalid	results
		1.41	
		2.93	
		invalid	results
		invalid	results
		0.68	
		1.60	

Which Method to Use?

- Sample size potentially too small with sub-sampling
- Randomization can produce inconsistent results

- Weighted method produces consistent results with more precision

25

Propensity Score Simulations Goals

- Understand conditions for which propensity scores are needed
- Assess what goes wrong when models are incorrectly specified
- Demonstrate that stratification helps
- Understand what regression does and doesn't do

26

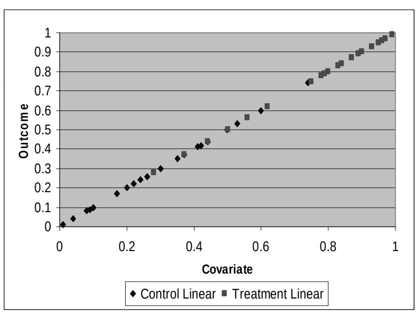
Correctly Modeled - Linear

0.75
Avg Outcome - Trt

Difference = 0.45

0.30
Avg Outcome - Ctl

Covariate Adjustment will deal with this difference



Out = β_0+ β_1 Cov + β_2 Trt

Out = 0 + 1 Cov + 0 Trt

27

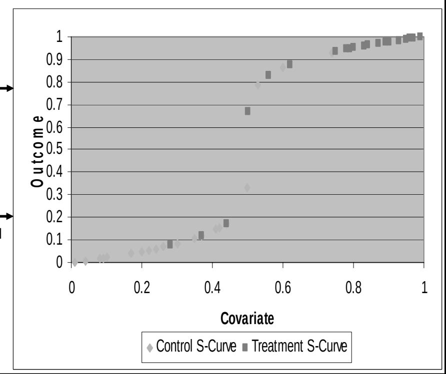
Correctly Modeled - S-Shaped

0.79
Avg Outcome - Trt

Difference = 0.59

0.20
Avg Outcome - Ctl

Covariate Adjustment will deal with this difference



Incorrect Model Specification

- What does it mean to be incorrectly specified?
  - Regression Model: ED ~ Mam + Income + ε
  - Examples of incorrect model specification:
    - ED ~ Mam + $(\text{Income})^2$ + ε
    - ED ~ Mam + $(\text{Income})^4$ + ε
    - ED ~ Mam + Income + Education + ε
  - Generally: ED ~ Mam + f(X)β + g(Z)γ + ε

    X are associated with Mam, Z are not

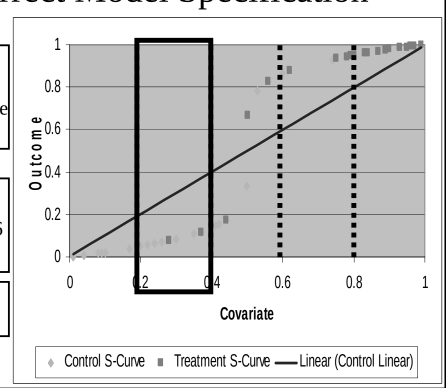
29

Incorrect Model Specification

Residual: S-Curve value minus linear value for each point

Average Resid
Treatment = 0.06
Control = -0.10

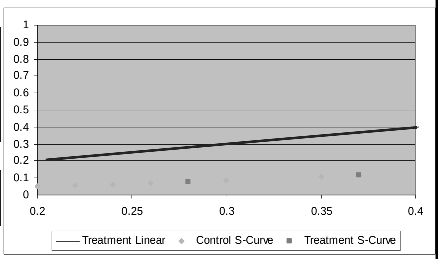
Bias = 0.16



Stratification Approach

-0.23	-0.20
Avg Resid Trt	Avg Resid Ctl

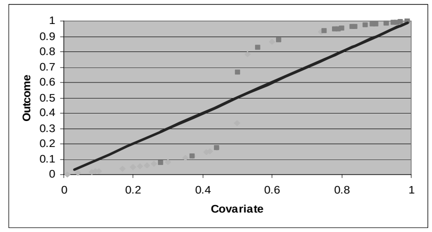
Bias = 0.03



31

Bias from Incorrectly Modeling with Stratification

Resid Trt	n/a	-.23	.05	.19	.08
Resid Ctl	-.06	-.20	-.14	.23	n/a
Bias	small	.03	.19	.04	small



32

Why Stratification Works

Bias = 0.10 < 0.16

33

Regression with Covariates

- What it does:
  - Adjusts for correctly specified model
- What it doesn't do:
  - Adjusts for incorrectly specified model
  - Deal with unobserved covariates

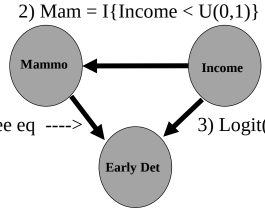
34

Simulations

2) Mam = I{Income < U(0,1)}

1) Income ~ U(0,1)

3) see eq ---->



3) Logit(p_{ed}) = Mam + Income + U(-1,1)

4) ED = I{p_{ed} < U(0,1)}

35

Simulations: Quasi-Randomization

- Examining Mean Income

	n	mam	no mam	t
	1,000,000	0.67	0.33	-708
	520,024	0.51	0.49	-31

- Susceptible to selection bias

36

Simulations: Comparing Bias

Underlying (true) model is ED ~ Mam + Inc

Regression Model		OR_{mam}
ED ~ Mam + Inc	regr	2.65
	w/PS	2.66
ED ~ Mam	regr	3.61
	w/PS	2.69

Adjustment is necessary when incorrect modeling might be present

37

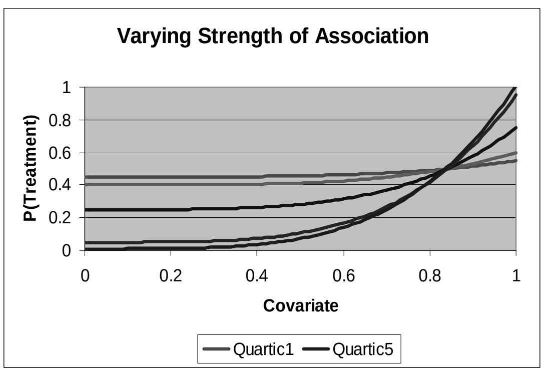
Simulations: Incorrectly Specified Model

	No PS	PS
	2.65	2.66
	2.81	2.70
	3.17	2.78

Regression model always ED ~ Mammography + Income

The more incorrect the model, the more the bias

38



39

Simulations: Strength of Propensity Score Model

	No PS	c (PS)	w/ PS
Quartic 1	2.73	0.55	2.70
Quartic 2	2.79	0.65	2.72
Quartic 3	2.95	0.78	2.76
Quartic 4	3.03	0.84	2.78
Quartic 5	3.17	0.92	2.78

Conclusion 1: The stronger the predictive ability of the propensity model, the more susceptible to selection bias

Conclusion 2: Using the c-statistic from the propensity score model may be better than looking at the bivariate distributions to assess the need for propensity score methods

40

Future Research

- Using other measures besides c-statistic
- Varying the strength of the relationship between exposure(s) and outcome
- Including covariates into the simulation models (like education) and varying parameters to determine their effects on biases and outcomes

41

Summary/Conclusions

- Methods for dealing with bias in observational studies are important
- We have seen two extensions to propensity score methods
  – polychotmous treatments
  – weighted techniques
- When considering propensity score methods, we should…
  – Examine the predictive ability of assignment to treatment
  – (Perhaps) examine the relationship between exposure and outcome

42

Thank You!

- Arlene Ash, PhD – Thesis Advisor
  - Director, Health Care Research Unit, BMC
- L. Adrienne Cupples, PhD
  - Chair, Dept of Biostatistics, BUSPH

43