Panacea or Poison: Can Propensity Score Modeling Replicate the Results from Randomized Control Trials? – PART I

Christopher M. Campbell, Ph.D.
Ryan M. Labrecque, Ph.D.
Portland State University

Supported by NIJ Award (#2016-R2-CX-0030)
Data Resources Program: Funding for Analysis of Existing Data

Presented at the 73rd Annual Meeting of the American Society of Criminology
on November 8, 2016 in Philadelphia, Pennsylvania

Propensity Score Modeling (PSM)

- Simulate effects of an RCT
  - Equal groups on likelihood of being treated (propensity score)
- Increasingly relied on in evaluation
- PSM shown to reduce bias
- Previous tests show mixed findings
  - *Possible* to achieve similar results as RCT, but
  - Can also yield *opposite effects* (Peikes, Moreno, & Orzol, 2008)
- Problems in use:
  - Used correctly only *28%* of time (Austin, 2008)
  - Blind acceptance / oversimplification calculating pscore

Current Study

Can PSM methods replicate the findings of RCTs?

Aim: Test multiple PSM techniques compared to an RCT

Measures of Balance

1. Statistical significance
   - Aim: less than 5% of covariates significantly different
2. Standardized Percent Bias (% Bias)
   - Difference in means or proportions between groups
   - Aim: Less than 20% is imperative, less than 10% is ideal
3. Area under the receiver operating characteristic (AUC)
   - Ability of pscore to predict treatment
   - Aim: Pre-match AUC >.7, post-match AUC = .5

Intensive Supervision for High-Risk Offenders (ICPSR 6358)

- Collected eligible RCTs from NACJD
- BJA funded study of intensive supervision program (ISP)
  - ISP: Stringent community monitoring
  - Primary intent: Effects of ISP on recidivism outcomes
- Method
  - 1987-1990, 14 sites, with 6 and 12 month follow-up
  - RAND randomly assigned to ISP or supervision as usual
- Result
  - ISP increase in violations and rearrests
- PSM testing: Houston, TX

Introducing Selection Bias

- Remove equality between groups / introduce selection bias
  - Identify unique characteristics of treatment cases
    - Coupled forward and backward stepwise logistic regression
    - Additive scale of characteristics
    - Select all cases above the scaled mean
  - Deemed biased if…
    - More than 10% of covariates were significantly different
    - Average percent bias of all covariates ≥ 15%.
    - Biased sample represents < 50% of original treatment group

Analytical Procedure

- Dr. Labrecque introduced bias and conducted imbalance check

- Dr. Campbell removed bias with PSM (logit) techniques in Stata15
  - Pscore covariates selected based on signif. Differences and %bias
  1. One-to-one greedy match with a caliper (1-1 w/c)
  2. One-to-many match, with caliper and replacement (1-k w/c)
  3. Stratification on propensity score (Strata)
  4. Inverse probability of treatment weighting (IPTW)
  - Post-assessment of balance treatment and comparison

psmatch2

pscore + manual calculation

- PSM outcome estimates then compared original RCT
  - Statistical significance, direction, magnitude

Testing PSM

- Primary outcomes chosen, but analyzed multiple
- Significance
  - Change in significance (< .05)
- Direction
  - Change in sign of Cohen's *d*
- Magnitude
  - Difference in *d* between RCT and PSM
    - 0 = PSM perfectly replicates RCT
    - Positive difference indicates PSM overestimates treatment
    - Negative difference indicates PSM underestimates treatment

Model Fit Summary Statistics

	RCT	Biased Sample	1-1 w/c	1-k w/c	Strat	IPTW
Sample size of treated	239					
% Covariates sig diff	3.0					
Mean % bias	7.1					
% Cov. bias over 20	3.0					
AUC	0.596					

Total covariates compared = 99

Covariates used in pscore calculation = 36

Model Fit Summary Statistics

	RCT	Biased Sample	1-1 w/c	1-k w/c	Strat	IPTW
Sample size of treated	239	72				
% Covariates sig diff	3.0	16.2				
Mean % bias	7.1	17.2				
% Cov. bias over 20	3.0	28.3				
AUC	0.596	0.875				

Total covariates compared = 99

Covariates used in pscore calculation = 36

Model Fit Summary Statistics

	RCT	Biased Sample	1-1 w/c	1-k w/c	Strat	IPTW
Sample size of treated	239	72	67	71		
% Covariates sig diff	3.0	16.2	4.1	7.1		
Mean % bias	7.1	17.2	12.3	15.5		
% Cov. bias over 20	3.0	28.3	21.4	29.3		
AUC	0.596	0.875	0.525	0.507		

Total covariates compared = 99

Covariates used in pscore calculation = 36

Model Fit Summary Statistics

	RCT	Biased Sample	1-1 w/c	1-k w/c	Strat	IPTW
Sample size of treated	239	72	67	71	56	118
% Covariates sig diff	3.0	16.2	4.1	7.1	2.0	8.2
Mean % bias	7.1	17.2	12.3	15.5	12.1	14.1
% Cov. bias over 20	3.0	28.3	21.4	29.3	18.2	25.5
AUC	0.596	0.875	0.525	0.507	0.864	0.658

Total covariates compared = 99

Covariates used in pscore calculation = 36

Results

	RCT		1-1 w/c		1-k w/c		Strat		IPTW	
Sample Size	C	T								
	219	239								
New tech. violation (%)	34.2	82.8								
Cohen's *d*	*1.20									
Diff in *d*	-									
Diff in %	-	-								
New arrest (%)	39.7	44.4								
Cohen's *d*	0.09									
Diff in *d*	-									
Diff in %	-	-								

*$p < .05$

Results

	RCT		1-1 w/c		1-k w/c		Strat		IPTW	
Sample Size	C	T	C	T						
	219	239	67	67						
New tech. violation (%)	34.2	82.8	31.3	86.6						
Cohen's *d*	*1.20		*1.46							
Diff in *d*	-		0.26							
Diff in %	-	-	-2.9	+3.8						
New arrest (%)	39.7	44.4	44.7	53.7						
Cohen's *d*	0.09		0.20							
Diff in *d*	-		0.11							
Diff in %	-	-	+5.0	+9.3						

*$p < .05$

Results

	RCT		1-1 w/c		1-k w/c		Strat		IPTW	
Sample Size	C	T	C	T						
	219	239	67	67						
New tech. violation (%)	34.2	82.8	31.3	86.6						
Cohen's *d*	*1.20		*1.46							
Diff in *d*	-		0.26							
Diff in %	-	-	-2.9	+3.8						
New arrest (%)	39.7	44.4	44.7	53.7						
Cohen's *d*	0.09		0.20							
Diff in *d*	-		0.11							
Diff in %	-	-	+5.0	+9.3						

*$p < .05$

Results

	RCT		1-1 w/c		1-k w/c		Strat		IPTW	
Sample Size	C	T	C	T	C	T				
	219	239	67	67	83	71				
New tech. violation (%)	34.2	82.8	31.3	86.6	34.7	83.1				
Cohen's *d*	*1.20		*1.46		*1.23					
Diff in *d*	-		0.26		0.03					
Diff in %	-	-	-2.9	+3.8	+0.5	+0.3				
New arrest (%)	39.7	44.4	44.7	53.7	38.5	52.1				
Cohen's *d*	0.09		0.20		0.30					
Diff in *d*	-		0.11		0.21					
Diff in %	-	-	+5.0	+9.3	+1.2	+7.7				

*$p < .05$

Results

	RCT		1-1 w/c		1-k w/c		Strat		IPTW	
Sample Size	C	T	C	T	C	T	C	T		
	219	239	67	67	83	71	105	56		
New tech. violation (%)	34.2	82.8	31.3	86.6	34.7	83.1	32.4	81.2		
Cohen's *d*	*1.20		*1.46		*1.23		*1.21			
Diff in *d*	-		0.26		0.03		0.04			
Diff in %	-	-	-2.9	+3.8	+0.5	+0.3	-1.8	-1.6		
New arrest (%)	39.7	44.4	44.7	53.7	38.5	52.1	40.5	47.1		
Cohen's *d*	0.09		0.20		0.30		0.14			
Diff in *d*	-		0.11		0.21		0.05			
Diff in %	-	-	+5.0	+9.3	+1.2	+7.7	-3.9	+2.7		

*$p < .05$

Results

	RCT		1-1 w/c		1-k w/c		Strat		IPTW	
Sample Size	C	T	C	T	C	T	C	T	C	T
	219	239	67	67	83	71	105	56	164	118
New tech. violation (%)	34.2	82.8	31.3	86.6	34.7	83.1	32.4	81.2	33.4	82.5
Cohen's *d*	*1.20		*1.46		*1.23		*1.21		*1.23	
Diff in *d*	-		0.26		0.03		0.04		0.03	
Diff in %	-	-	-2.9	+3.8	+0.5	+0.3	-1.8	-1.6	-0.8	-0.3
New arrest (%)	39.7	44.4	44.7	53.7	38.5	52.1	40.5	47.1	39.9	46.5
Cohen's *d*	0.09		0.20		0.30		0.14		0.15	
Diff in *d*	-		0.11		0.21		0.05		0.06	
Diff in %	-	-	+5.0	+9.3	+1.2	+7.7	-3.9	+2.7	+0.2	+2.1

*$p < .05$

Result of Test in Houston ISP Study

- All PSM techniques achieved similar results as RCT
  - Same significance, direction, and range in d = .03 to .26
- 1-1 w/c was furthest from RCT
  - Overestimated treatment between 3.9 and 9.3%
- Propensity score stratification was most conservative estimate and closest to RCT (+/- 2 to 4%)

Conclusion

- PSM techniques strikingly similar to RCT
- PSM overestimates treatment effect

- Next Steps
  - Tests for differences
    - Propensity score density and distance
    - Noninferiority / equivalence tests
    - Sensitivity analyses
  - Test multiple studies
    - Meta-analyze
  - Identify *when* differences occur and *why*
  - Test accuracy by technique used in relation to
    - Outcome base rate
    - Sample size
    - Covariates used in propensity score

Contact Information

Christopher M. Campbell, Ph.D.
Assistant Professor
Department of Criminology and Criminal Justice
Portland State University

Phone: 503 725-9896
E-mail: ccampbell@pdx.edu