Prognostic score-based balance measures for propensity score methods in comparative effectiveness research (Stuart, Lee, Leacy, 2013)

Presentation by Adam Sadowski

University of Waterloo

December 13, 2017

Overview

Motivation

Multiple **propensity score (PS) methods** to produce causal treatment effect estimates:

- Inverse Probability Weighting (IPW)
- Subclassification/Stratification
- Matching

Multiple **balance measures (BM's)** act as diagnostics for how well these methods balance treatment groups on observed covariates:

- Absolute Standardized Mean Difference (ASMD)
- Kolmogorov-Smirnov (KS) test statistic
- SMD of Propensity Scores
- SMD of Prognostic Scores

Motivation

Theory behind the PS and BM approach:
More Balance $\rightarrow$ *Less Bias in Causal Treatment Effect Estimate* (B_{CTE})
But the authors explain (and will show) how this is not always true:

- PS methods balance on propensity scores, which are determined more-so by the covariates strongly associated with treatment assignment.
- There may be covariates that are weakly associated with treatment assignment but quite strongly associated with outcome.
- Therefore it's possible we can underestimate B_{CTE} due to small imbalances in the latter covariates

Motivation

- We need a BM that performs well notwithstanding; and best among all others to simplify the choice of BM.

How?

- Since (most) BM's assess absolute differences of covariates between groups, they should positively correlate with $|B_{CTE}|$.
- The best BM will have the highest correlations with $|B_{CTE}|$ under a variety of settings (e.g. different PS methods and models).

Definitions

Prognostic Scores

Predicted outcomes (prognoses) under the control condition. Estimated by modelling outcome in the control condition, and applying the model to all subjects to get their prognoses under the control condition. **Advantage:** Balance on these scores (low SMD's) reflect balance on the most predictive and potentially bias-causing covariates.

Propensity Score: $ps(x)$

Probability of getting treated given observed covariates. Used to weight, group, or match subjects e.g., so that randomization of subjects is simulated (covariates become balanced among treatment and control). Covariate confounding of the treatment effect is thus reduced.

Structure of SMD Balance Measures

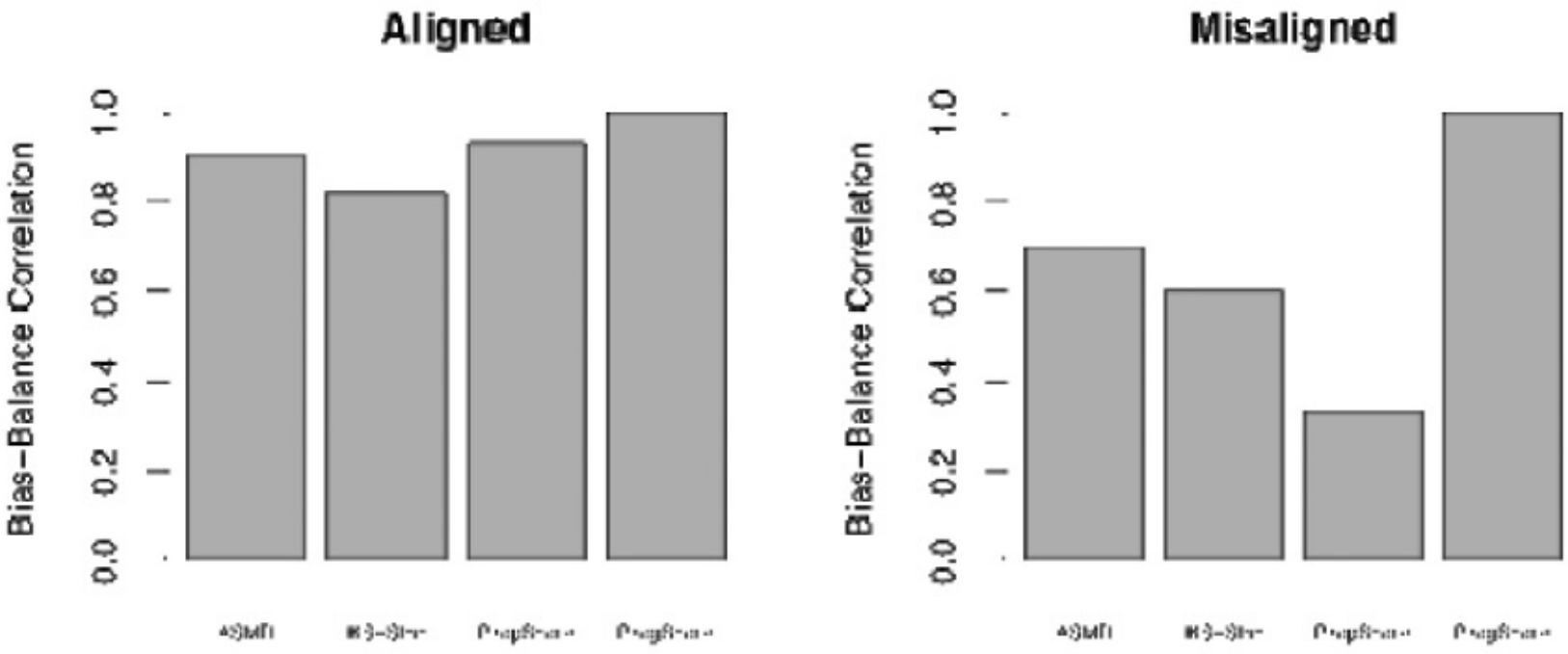
$$\frac{|\bar{x}_{j,1} - \bar{x}_{j,0}|}{sd(x_{j,1})} \text{ for } j = 1, ..., p \text{ where } p \text{ is the number of covariates.}$$

Smaller Simulation

- **Purpose:** Demonstrate how traditional BM's may mislead; how they do not always strongly correlate with B_{CTE}.
- **Method:**
  - IPW
  - 2 Normal Covariates
    - Aligned Setting: One covariate is both predictive of treatment assignment and outcome.
    - Misaligned Setting: One covariate is predictive of assignment but not associated with outcome, and the other covariate is vice-versa.
  - 4 BM's
    - average ASMD
    - average KS
    - SMD of Propensity Scores
    - SMD of Prognostic Scores

Smaller Simulation

- **Results:**



(1) ASMD, (2) KS-Stat, (3) Prop-Score, (4) Prog-Score

Larger Simulation

Details

- 9 BM's
  - Mean, med., max. SMD; SMD's $< .1, .25$; mean KS and t-stat., SMD in ps, SMD in prog. score
- 3 PS Methods
- 4 Outcome Models
  - Treatment: Binary
  - Outcomes: Continuous
  - Treatment Effect: 3.0
  - 2 Model Settings
    - Cntns: covariates are continuous
    - Mixed: covariates are binary and continuous

Sample Size: 1000, **Datasets:** 2000

Cntns: 4 confounders, 2 treatment predictors, 2 outcome predictors, 1 neutral (9 covariates total)
Mixed: 4 of the above were Bernoulli(0.5)

Larger Simulation

Table: Outcome Models and Contained Terms

Model	Main Effects	Interactions	Squared
A	6	0	0
B	6	4	0
C	6	0	2
D	6	4	2

Larger Simulation

Table: Prognostic Score Models: 6 of 9 covariates are related to outcome.

Model	**Main Effects**	**Omit. Cov./Conf.**
1	9	0
2	6	0
3	5	Ct
4	5	WC
5	5	CovO
6	5	SC
7	6	Plus CovT

Larger Simulation

Bias Calculation for Each Dataset

- 40 PS models with varied misspecification
  - varied covariate balance
- 40 sets of weights determined by the given propensity scores
- 40 values for each balance measure and treatment effect (R *survey*)
- Pearson's r between each balance measure and $|B_{CTE}|$

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}}$$

  - x_i: a specified balance measure (under specified outcome and covariate model settings) reflecting balance on covariates in relation to PS model $i = 1, ..., 40$.

(Reported correlations are mean r's of 2000 datasets)

Larger Simulation

Table 1

Mean Pearson correlations between balance measure and bias for weighting by the odds, N=1000

	Absolute Bias Outcome A		Absolute Bias Outcome B		Absolute Bias Outcome C		Absolute Bias Outcome D	
	Cntns	**Mixed**	**Cntns**	**Mixed**	**Cntns**	**Mixed**	**Cntns**	**Mixed**
Mean SMD	0.87	0.71	0.86	0.70	0.85	0.69	0.83	0.68
Median SMD	0.57	0.46	0.56	0.45	0.55	0.44	0.53	0.43
Max SMD	0.73	0.63	0.72	0.63	0.72	0.62	0.71	0.62
%(SMD) <0.25	0.52	0.43	0.51	0.42	0.50	0.40	0.49	0.40
%(SMD) <0.1	0.79	0.68	0.78	0.67	0.77	0.66	0.76	0.65
Mean K-S Statistic	0.69	0.58	0.67	0.57	0.65	0.55	0.63	0.54
Mean T-Test Statistic	0.89	0.72	0.87	0.71	0.88	0.72	0.86	0.71
Propensity Score SMD	–0.31	–0.39	–0.30	–0.39	–0.29	–0.38	–0.27	–0.37
Prognostic Score 1 SMD	1.00	1.00	0.98	0.99	0.96	0.98	0.94	0.97
Prognostic Score 2 SMD	1.00	1.00	0.98	0.99	0.97	0.98	0.95	0.98
Prognostic Score 3 SMD	0.98	0.99	0.96	0.98	0.95	0.97	0.93	0.97
Prognostic Score 4 SMD	0.99	1.00	0.97	0.99	0.95	0.98	0.93	0.97
Prognostic Score 5 SMD	1.00	1.00	0.98	0.99	0.96	0.98	0.95	0.97
Prognostic Score 6 SMD	0.80	0.79	0.78	0.75	0.58	0.55	0.54	0.45
Prognostic Score 7 SMD	1.00	1.00	0.98	0.99	0.96	0.98	0.94	0.97

Mean SMD: Mean Standardized Mean Difference (SMD); **Median SMD:** Median SMD; **Maximum SMD:** Maximum SMD; **%(SMD)<0.25:** Percentage of SMDs less than 0.25; **%(SMD)<0.1:** Percentage of SMDs less than 0.1; **Maximum K-S statistic:** Maximum Kolmogorov-Smirnov statistic; **Mean K-S statistic:** Mean Kolmogorov-Smirnov statistic; **Propensity Score SMD:** SMD in Propensity Scores; **Prognostic Score 1–7 SMD:** SMD in Prognostic Scores for Prognostic Score Models 1–7; **Cntns:** all covariates are continuous; **Mixed:** a mixture of continuous and binary covariates.

Results of Larger Simulation

- As long as most variables related to outcome are in the prog. score model, we have high correlations notwithstanding misspecification.
- Prog. Score 6 still outperforms most measures for Models A and B.
- PS Method did not affect results.
- Prog. Score has not only the highest correlations, but is the only balance measure that does not suffer in the Mixed setting.
- It can serve as a proxy for bias reduction on observed covariates.

Final Remarks

- **Model Specification:**
  - PS models should have all confounders and variables related to outcome.
  - Prognostic score models should have all variables related to outcome.
- **Limitations:**
  1. *What if you don't have access to outcome data?* 1 Reason: to avoid controversy whether a design was used to obtain a certain result. Solution: cross-validation
  2. *What if there exist interactions between treatment status and outcome?* The prog. score model can be substantially different depending on treatment. Solution: devise 2 outcome models, one for each group, and get prognoses of all subjects under each condition; then assess balance for prognoses under control and prognoses under treatment.