Evaluating UNC's Chancellor's Science Scholars Program with Balancing Weights

Noah Greifer[1], A. T. Panter[2], & Viji Sathy[3]

[1]*noah@unc.edu,* [2]*panter@unc.edu,* [3]*viji_sathy@unc.edu*
The University of North Carolina at Chapel Hill
Department of Psychology & Neuroscience

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Abstract

In studies lacking random assignment to condition, bias can occur in effect estimates due to confounding by pre-treatment variables. We examine three methods to weight units to arrive at covariate balance: propensity score weights estimated with logistic regression, Imai and Ratkovic's Covariate Balancing Propensity Score weights, and Zubizarreta's stable balancing weights. These methods are applied to the evaluation of the Chancellor's Science Scholars (CSS) program at UNC, an academic enrichment program aimed at increasing minority achievement in science, technology, engineering,. And math (STEM). After weighting, balance is achieved on most variables of interest, though logistic regression performs the worst. All methods yield statistically significant effects of CSS participation on science GPA and overall GPA after the first year.

Introduction

- Need to eliminate confounding in observational studies
- Need to report balance after preprocessing
- Preprocessing methods have different strengths and weaknesses
  - We consider logistic regression PS weights, CBPS (Imai & Ratkovic, 2014), and stabilized balancing weights (SBW; Zubizarreta, 2015)
- Need to assess balance after preprocessing
- These methods applied to evaluation of UNC's Chancellor's Science Scholar program
  - Experimental academic enrichment program to improve STEM achievement for underrepresented students
  - Students admitted through intensive selection process leading to confounding for academic outcomes

Methods

- ***Participants:*** CSS participants (n = 90) and science-interested UNC undergraduates (n = 4057) across three cohorts.
- ***Covariates:*** High school academic achievement, high school environment, and demographic variables listed in figure 1. Students had missing data for high school GPA (9.3%) and high school type (1.8%), two prognostic variables. Both were considered missing at random.
- ***Outcomes:*** Overall GPA and Science GPA after the first year.
- ***Analysis.*** All analyses took place in R (R Core Team, 2017).
  - Multiple imputation with chained equations using the *mice* package (van Buuren & Groothuis-Oudshoorn, 2011) to generate 15 imputed data sets.
  - Three balancing weights:
    - PS weights from logistic regression with model building
    - CBPS were estimated using the *CBPS* package (Fong et al., 2016).
    - Stabilized balance weights were estimated using the *sbw* package (Zubizarreta & Allouah, 2016).
  - Balance assessed within in each cohort after applying each type of weight (figure 1) using the *cobalt* package (Greifer, 2017).
  - We used the weights to estimate the average treatment effect on the treated (ATT), controlling for cohort year.
    - Standard errors were estimating using the sandwich variance estimator implemented in the *survey* package (Lumley, 2016).
    - Estimates were combined across imputations using Rubin's rules (Barnard & Rubin, 1999).

Covariate Balance

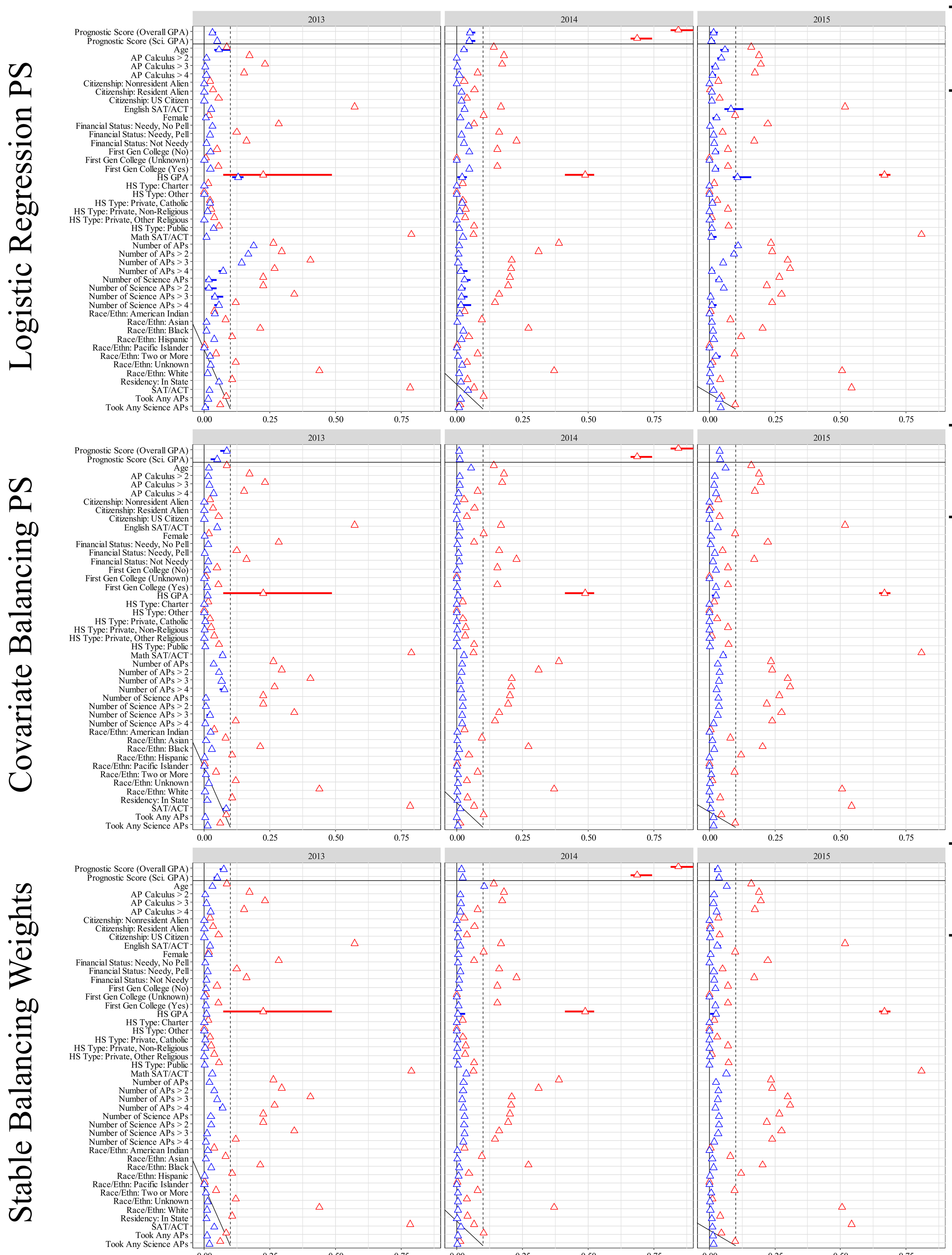

Figure 1. Covariate balance across imputations for each of the three cohorts and each of the three balancing methods. Red triangles represent the average unweighted mean differences across imputations, and blue triangles represent the average weighted mean differences across imputations. Bars represent the range of mean differences across imputations, but are too short to plot for most covariates. For continuous variables, including the prognostic scores, the mean differences are standardized by the standard deviation of the original treated group. For binary variables, the mean differences are the raw differences in proportion. All mean differences are presented in absolute value. A threshold at 0.1 is denoted by the dotted lines; mean difference below this threshold are indicative of good balance.

Estimates

	Science GPA	Overall GPA
Estimate	**0.167**	**0.109**
Standard Error	0.066	0.051
P-value	*0.016*	*0.040*
95% Conf. Interval	(0.037, 0.296)	(0.009, 0.209)

	Science GPA	Overall GPA
Estimate	**0.167**	**0.115**
Standard Error	0.066	0.051
P-value	*0.016*	*0.032*
95% Conf. Interval	(0.037, 0.297)	(0.015, 0.216)

	Science GPA	Overall GPA
Estimate	**0.164**	**0.122**
Standard Error	0.066	0.051
P-value	*0.018*	*0.023*
95% Conf. Interval	(0.035, 0.293)	(0.021, 0.222)

Tables 1, 2, and 3. The effect of CSS participation on Science GPA and Overall GPA after the first year. These estimates account for the main effect of cohort. GPAs are on a 4.0 scale. Standard errors were calculated using the sandwich variance estimator as implemented in *survey* for R.

Results

- The effect estimates, p-values, and confidence intervals are listed in tables 1, 2, and 3.
  - The effect on science GPA was estimated to be around 0.17. The effect on overall GPA was estimated to be around 0.12.
- The weighting methods balanced covariates approximately equally well.
  - CBPS balanced all covariates at a threshold of 0.1.
  - The logistic regression PS weights and SBW balanced nearly all covariates.
  - Prognostic score for both outcomes were balanced for all methods; this is an indicator of an effect estimate with little bias (Stuart, Lee, & Leacy, 2013).
- The standard errors and 95% confidence interval widths were approximately the same for all three methods.
  - CBPS consistently showed somewhat larger standard errors and wider confidence interval widths.

Conclusions

- CSS appears to be a promising program with small but present effects on GPA after the first year.
- Contrary to expectations, the three balancing methods performed approximately equally.
  - CBPS least precise due to extreme constraints
  - SBW most precise for the effect on science GPA
  - Logistic regression most for the effect on overall GPA
- Recommendations:
  - With larger samples, use CBPS for bias elimination
  - With smaller samples, use SBW with relaxed constraints
  - Consider prognostic score in balance evaluation

References

Barnard, J., & Rubin, D. B. (1999). Miscellanea. Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*(4), 948–955.

Fong, C., Ratkovic, M., Hazlett, C., Yang, X., & Imai, K. (2016). CBPS: Covariate Balancing Propensity Score (Version 0.13).

Greifer, N. (2017). cobalt: Covariate Balance Tables and Plots (Version 2.0.0).

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243–263.

Lumley, T. (2016). survey: analysis of complex survey samples (Version 3.31-5).

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8), S84.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Zubizarreta, J. R. (2015). Stable Weights that Balance Covariates for Estimation With Incomplete Outcome Data. *Journal of the American Statistical Association*, *110*(511), 910–922.

Zubizarreta, J. R., & Allouah, A. (2016). sbw: Stable weights that balance covariates (Version 0.0.2).