

## JAMA Guide to Statistics and Methods

# Overlap Weighting

## A Propensity Score Method That Mimics Attributes of a Randomized Clinical Trial

Laine E. Thomas, PhD; Fan Li, PhD; Michael J. Pencina, PhD

**Evidence** obtained from clinical practice settings that compares alternative treatments is an important source of information about populations and end points for which randomized clinical trials are unavailable or infeasible.<sup>1</sup> Unlike clinical trials, which strive to ensure patient characteristics are comparable across treatment groups through randomization, observational studies must attempt to adjust for differences (ie, confounding). This is frequently addressed with a propensity score (PS) that summarizes differences in patient characteristics between treatment groups. The PS is the probability that each individual will be assigned to receive the treatment of interest given their measured covariates.<sup>2</sup> Matching or weighting on the PS is used to adjust comparisons between the 2 groups being compared.<sup>2,3</sup>

In an article in *JAMA Cardiology*, Mehta et al evaluated the association between angiotensin-converting enzyme inhibitors (ACEIs), angiotensin II receptor blockers (ARBs), or both with testing positive for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (COVID-19), in 18 472 patients who were tested in the Cleveland Clinic Health System between March 8, 2020, and April 12, 2020.<sup>4</sup> Overlap weighting<sup>5,6</sup> based on the PS was used to adjust for confounding in the comparison of 2285 patients who had been treated with ACEIs/ARBs with 16 187 patients who did not receive ACEIs/ARBs. After adjustment, there was no significant association between ACEI/ARB use and testing positive for SARS-CoV-2.

### Use of the Method

#### Why Is Overlap Weighting Used?

Overlap weighting is a PS method that attempts to mimic important attributes of randomized clinical trials: a clinically relevant target population, covariate balance, and precision. The *target population* is the group of patients for whom the conclusions are drawn.<sup>3</sup> *Balance* refers to the similarity of patient characteristics across treatment, which is an important condition to avoid bias. *Precision* denotes the certainty about the estimate of association between the treatment and the outcome of interest; more precise estimates have narrower CIs and greater statistical power. Although classic PS methods of inverse probability of treatment weighting (IPTW) and matching can adjust for differences in measured characteristics,<sup>2,3</sup> these methods have potential limitations with respect to target population, balance, and precision.<sup>5</sup>

Conventional IPTW assigns a weight of  $1/PS$  for treated and  $1/(1 - PS)$  for untreated patients, allowing individuals with underrepresented characteristics to count more in the analysis.<sup>3</sup> Matching operates differently by taking each treated study participant and finding the closest PS match among controls, usually within a bound. In observational data, in which the initial differences in treatment groups

may be large, these methods can modify the target population,<sup>3</sup> fail to achieve good balance, or substantially worsen precision.<sup>5</sup>

Overlap weighting overcomes these limitations by assigning weights to each patient that are proportional to the probability of that patient belonging to the opposite treatment group.<sup>6</sup> Specifically, treated patients are weighted by the probability of not receiving treatment ( $1 - PS$ ) and untreated patients are weighted by the probability of receiving the treatment ( $PS$ ). These weights are smaller for extreme PS values so that outliers who are nearly always treated ( $PS$  near 1) or never treated ( $PS$  near 0) do not dominate results and worsen precision, as occurs with IPTW. These outliers contribute relatively less to the result, while patients whose characteristics are compatible with either treatment contribute relatively more (Figure, A and B). The resulting target population mimics the characteristics of a pragmatic randomized trial that is highly inclusive, excluding no study participants from the available sample but emphasizing the comparison of patients at clinical equipoise. Moreover, overlap weighting has desirable statistical properties. It leads to exact balance on the mean of every measured covariate when the PS is estimated by a logistic regression and is proven to optimize precision of the estimated association between treatment and outcomes among a large class of PS weighting methods, including IPTW and an analogue to matching.<sup>6</sup> Overlap weighting can be as efficient as randomization if no adjustment was needed.<sup>5</sup>

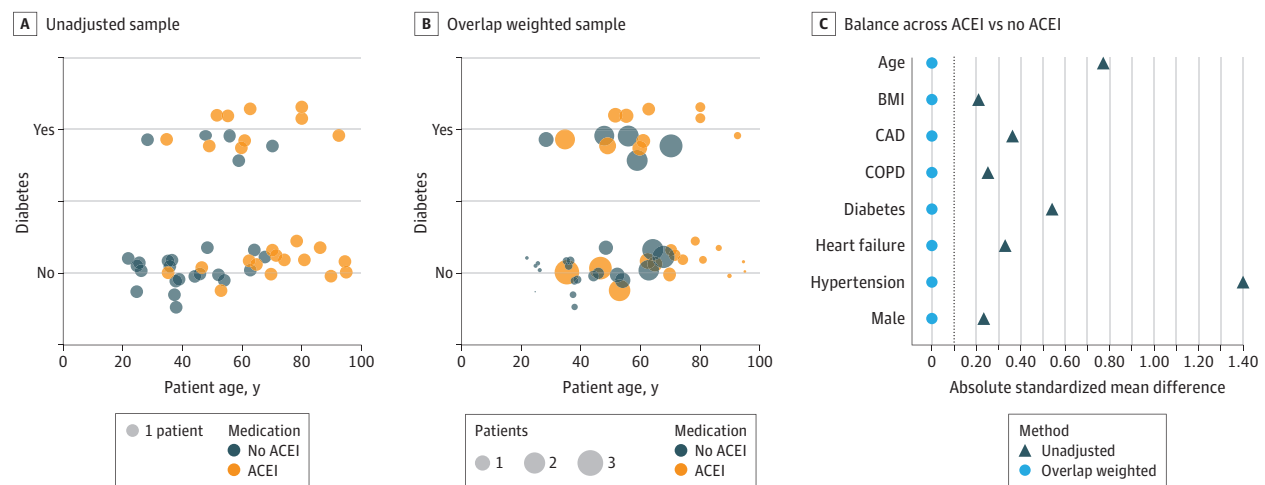
#### What Are the Limitations of Overlap Weighting?

Like all PS methods, overlap weighting cannot adjust for patient characteristics that are not measured and included in the model for the PS. It is important to identify confounding variables from the literature, attempt to include them in the analysis, and recognize potential bias due to unmeasured factors. For applications in which the initial imbalances in patient characteristics between treatment groups are modest, overlap weighting yields similar results to IPTW. The advantages of overlap weighting are greatest when comparator groups are initially very different.

#### Why Did the Authors Use Overlap Weighting in This Study?

Mehta et al<sup>4</sup> used overlap weighting to achieve good balance and minimize variance of the estimated association between ACEI/ARB treatment and test results positive for SARS-CoV-2. Both goals were achieved. Balance was demonstrated by reporting the overlap weighted covariate means (or proportions) for the group that received ACEIs/ARBs and the group that did not receive ACEIs/ARBs. There was no difference between groups after weighting (Figure, C). The list of covariates included risk factors related to receiving ACEI/ARB treatment and associated with testing positive for COVID-19. The adjusted treatment comparisons were estimated with narrow CIs, providing strong evidence for the null result.

Figure. Effect of Overlap Weighting on the Relative Contribution of 50 Simulated Patients With Different Ages and Diabetes Status



Simulated according to the distribution of the same variables in the study by Mehta et al.<sup>4</sup> The bubble size reflects the relative contribution of each patient to analysis. A, Each patient represents only themselves. Patients receiving angiotensin-converting enzyme inhibitors (ACEIs) are older and more likely to have diabetes. B, After overlap weighting some patients represent up to 3 other patients, while others represent less than 1 other patient. C, The absolute

standardized mean difference is the absolute value of the difference in mean between treatment groups divided by the SD. An absolute mean standardized difference less than or equal to 0.10 indicates good balance. BMI indicates body mass index; CAD, coronary artery disease; COPD, chronic obstructive pulmonary disease.

### How Should the Results of Overlap Weighting Be Interpreted in This Study?

The primary results of the study by Mehta et al<sup>4</sup> can be interpreted just like other PS methods. That is, after adjustment for differences in cardiovascular risk factors, 9.1% of patients who were treated with ACEIs/ARBs tested positive for SARS-CoV-2 compared with 9.4% of patients who were not treated with ACEIs/ARBs (odds ratio, 0.97 [95% CI, 0.81-1.15]). These estimates are measures of association between ACEI/ARB status and test positivity, with respect to a population of patients at equipoise either to receive treatment with ACEIs/ARBs or not and for whom all measured covariates are made similar across treatments through overlap weighting. Bias due to unmeasured differences between patients who received ACEI/ARB treatment vs those who did not cannot be ruled out.

### Caveats to Consider When Assessing the Results of an Overlap-Weighted Analysis

Overlap weighting creates exact balance on the mean of every measured covariate when the PS is estimated by logistic regression (Figure, C). This is particularly important for reducing bias<sup>7</sup>; however, balance on the mean may not result in complete adjustment for confounding on that variable. In addition, the baseline characteristics table of the overlap weighted sample should be presented (Table 2 and 3 in Mehta et al<sup>4</sup>). This table can include covariate means, medians, interquartile ranges, or any other statistics that are useful to understanding the population. This approach will help to demonstrate which randomized clinical trial is best emulated by the overlap weighted analysis with respect to target population, balance, and precision.

#### ARTICLE INFORMATION

**Author Affiliations:** Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina (Thomas, Pencina); Duke Clinical Research Institute, Durham, North Carolina (Thomas, Li, Pencina); Department of Statistical Science, Duke University, Durham, North Carolina (Li).

**Corresponding Author:** Michael J. Pencina, PhD, Department of Biostatistics and Bioinformatics, Duke University, 2400 Pratt St, Durham, NC 27705 ([michael.pencina@duke.edu](mailto:michael.pencina@duke.edu)).

**Section Editors:** Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, JAMA.

**Published Online:** May 5, 2020.  
doi:10.1001/jama.2020.7819

**Conflict of Interest Disclosures:** Dr Pencina reported receiving grants from Regeneron/Sanofi and Amgen and personal fees from Merck and Boehringer Ingelheim outside the submitted work. No other disclosures were reported.

#### REFERENCES

- Basch E, Schrag D. The evolving uses of "real-world" data. *JAMA*. 2019;321(14):1359-1360. doi:10.1001/jama.2019.4064
- Haukoos JS, Lewis RJ. The propensity score. *JAMA*. 2015;314(15):1637-1638. doi:10.1001/jama.2015.13480
- Thomas L, Li F, Pencina M. Using propensity score methods to create target populations in observational clinical research. *JAMA*. 2020;323(5):466-467. doi:10.1001/jama.2019.21558
- Mehta N, Kalra A, Nowacki AS, et al. Association of use of angiotensin-converting enzyme inhibitors

and angiotensin II receptor blockers with testing positive for coronavirus disease 2019 (COVID-19). *JAMA Cardiol*. Published online May 5, 2020. doi:10.1001/jamacardio.2020.1855

5. Li F, Thomas LE, Li F. Addressing extreme propensity scores via the overlap weights. *Am J Epidemiol*. 2019;188(1):250-257.

6. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. *J Am Stat Assoc*. 2018;113(521):390-400. doi:10.1080/01621459.2016.1260466

7. Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. *J Am Stat Assoc*. 2015;110(511):910-922. doi:10.1080/01621459.2015.1023805