

Commentary on Yu et al.: Opportunities and Challenges for Matching Methods in Large Databases

Elizabeth A. Stuart and Benjamin Ackerman

In their paper titled “Matching Methods for Observational Studies Derived from Large Administrative Databases,” authors Ruoqi Yu, Jeffrey Silber and Paul Rosenbaum [13] discuss matching methods in the age of “big data.” Matching methods, such as Mahalanobis distance matching, exact matching and propensity score matching, are well established design strategies for reducing bias due to observed characteristics in nonexperimental studies. Fundamentally, matching methods aim to create matched groups of units (often individuals) who are similar to one another on a (sometimes large) set of covariates. Exact matching aims to do that matching on each variable simultaneously (e.g., find for each treated individual a control individual with the same values of age, gender, baseline health status, etc.). Other matching methods summarize the covariates into a low-dimensional summary (e.g., the probability of receiving treatment, known as the propensity score [9]) and match units on that summary. The simplest form of matching is 1:1 matching, where each treated subject is matched to 1 comparison subject similar on the variables used in the matching. Many variations on 1:1 matching exist, including variable ratio matching, full matching and methods that allow controls to be used as a match more than once (“matching with replacement”). Recent advances in matching include “fine balance,” which aims to exactly match the marginal distributions of the covariates rather than requiring each pair match exactly [8]. See Stuart (2010) [11] for the background and details of some of those approaches. (Note that while Yu et al. [13] use the term “control” to refer to the reference group, we prefer the term “comparison” group in the nonexperimental context, to distinguish it from the control group in a randomized controlled trial. For current purposes, readers can consider the “control” and “comparison” terms equivalent.)

Weighting, in particular inverse probability of treatment weighting [1], is another common strategy for handling

observed covariates in nonexperimental studies. Some researchers view it as preferable to matching because it does not “throw away data” as matching can appear to do (e.g., a study with 1000 treated and 2000 comparison subjects may end up using only 1000 of the comparison subjects in a 1:1 matched design). However, there are a number of benefits to considering matching. First, it is often an attractive approach for nonexperimental study design because of its strong design aspects: it is straightforward, for example, to show the similarity of the matched groups and for even nontechnical readers to intuitively see that the individuals that remain in the analysis “look similar” to one another, at least on the observed covariates [4]. In contrast, with weighting it can be challenging for readers to interpret what it means when, for example, a subject receives a “weight” of 1.3; what is 1.3 people? Finally, it is important to note that the apparent change in sample size does not always lead to reduced power; with weighting the key metric would be the effective sample size, not the number of subjects that remain in the analysis. The total sample size in this case can thus sometimes be a not fully informative metric. For these reasons, matching should be considered as a key tool and often quite appropriate strategy in the nonexperimental study design toolbox and we are happy to see this paper, which helps provide a method for implementing matching in large-scale data.

While the increasing availability of large administrative data (such as electronic health records (EHR), medical claims data, large-scale educational datasets or social media data) allows for unprecedented inference and research, its high dimensional nature poses computational challenges that make certain existing matching methods either infeasible to implement or unfavorable given the context and time needed to complete. In fact, such data often lead to a conundrum: large numbers of covariates observed, such that being able to match on many or all of them can help satisfy the key assumption of unconfounded treatment assignment [9], but computationally it may be challenging to actually do that matching.

These large and comprehensive datasets thus provide opportunity for significant bias reduction due to observed covariates, but there is still the issue of how to actually deal with all of those covariates. The approach proposed by Yu et al. (2020) [13] provides a potential solution to

Elizabeth A. Stuart is Professor of Mental Health, Biostatistics, and Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, 21205, USA (e-mail: estuart@jhu.edu). Benjamin Ackerman is a Doctoral Candidate in the Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, 21205, USA.