The central role of the propensity score in epidemiology

Brian K. Lee

Observational Studies, Volume 9, Issue 1, 2023, pp. 55-57 (Article)

Published by University of Pennsylvania Press

DOI: https://doi.org/10.1353/obs.2023.0004

For additional information about this article
https://muse.jhu.edu/article/877886
The central role of the propensity score in epidemiology

Brian K. Lee

Department of Epidemiology and Biostatistics
Drexel University Dornsife School of Public Health
Philadelphia, PA 19104

bkl29@drexel.edu

Abstract

In this commentary, I provide a personal perspective on how the propensity score has become important to epidemiology.

Keywords: Propensity Score, Epidemiology

1. Commentary

I first learned about the propensity score as a first-year epidemiology graduate student in 2005, twenty-two years after Rosenbaum and Rubin’s paper was published in Biometrika.\(^1\) As a technically proficient but otherwise not supremely insightful student, I encountered the propensity score in a biostatistics problem set, where I was directed to apply propensity score stratification to calculate a particular estimand of interest. Step-by-step directions were provided on what to do but without any greater context provided as to why or how this technique would be useful. So, my first encounter with the propensity score was essentially a statistical programming exercise and I was left feeling unsatisfied. Technically speaking, I knew how to conduct a propensity score analysis, but I did not know why I would want to conduct a propensity score analysis: it was like I followed a recipe to bake a cake without getting to eat the cake and understanding the deliciousness of the final product. I remember thinking to myself: “that was a lot of work to arrive at basically the same answer I would have gotten using linear regression.” I am not faulting my excellent course instructors in any way, but it seems that even at twenty-plus years of age, the propensity score was not yet fully mature, at least pedagogically speaking.

Looking back 40 years to when Rosenbaum and Rubin’s paper was published in 1983, the authors were remarkably prescient and perhaps even a bit overconfident, in giving their paper the title: “the central role of the propensity score in observational studies for causal effects.” Back in the 2000’s, the propensity score certainly did not occupy a central role in the methodological toolkit for observational studies: from 1983 to 2003, a PubMed search for the term ‘propensity score’ returns about 250 results. Uptake was sparse, applications in the literature were infrequent, and propensity score methodology as well as techniques for teaching it were not optimized even two decades after introduction. But from 1983 to August 2022, the same PubMed search for ‘propensity score’ returns over 38,000 results. In the 3\(^{rd}\) and 4\(^{th}\) decades of its existence, the propensity score matured to the point where it has become second to only outcome regression as the most commonly used parametric method for causal inference.\(^2\) Contextual factors during this time period likely contributed to this
growth: the debate regarding randomized trials vs. observational studies; the explosion of interest in causal inference in medicine and public health; and more advanced computational hardware and software for performing propensity score analyses. Beyond these factors, however, I would like to suggest two additional developments seeded by Rosenbaum and Rubin years ago that may have helped the propensity score come to fruition in some small way.

The first development is that the propensity score has transcended the default state of ‘niche topic’ that many statistical innovations seem to be relegated to. A large part of this may be due to epidemiologists who, perhaps even more so than their statistician colleagues (at least in my joint epidemiology-biostatistics department), have widely embraced the use of propensity scores. Although writing for *Biometrika*, a journal that is not standard reading in epidemiology circles, Rosenbaum and Rubin still made the material comprehensible for epidemiologists. In particular, one item of discussion in their article, at least for me, stands out as a reason why epidemiologists have taken to propensity scores so readily: matching. It was not until 2008 when I took a causal inference class led by Elizabeth Stuart that I began to understand the potential of the propensity score. In that class, the propensity score was introduced in a manner inspired by how Rosenbaum and Rubin did so: within the broader theme of comparing randomized experiments to non-randomized studies. However, Stuart added a slight but significant wrinkle: instead of discussing matching *after* the propensity score (as Rosenbaum and Rubin did, as a means of adjustment for the propensity score), the course first introduced matching, even before mentioning the term ‘propensity score.’ Matching is familiar to even the most novice of epidemiologists, usually in the context of a matched case-control study where cases and controls are selected with similar covariate values. In addition, matching fits naturally with basic epidemiology curriculum that presents causal inference using a counterfactual theory framework. While matching is a trivial task for one or two covariates, the curse of dimensionality prevents exact matching on more than a handful of covariates from being a viable strategy in most epidemiological studies. It was in this context that the lightbulb in my head went off and the propensity score finally made sense, as a means of enabling matching to ensure that exposed and unexposed individuals were as similar as possible on observed covariates. Propensity score matching became a straightforward way for epidemiologists to bridge the gap from counterfactual theory steeped in discussion of impossible scenarios (e.g., time machines, parallel universes, and study investigators imbued with omniscience) to actual, practical application. Moreover, propensity scores serve as a natural starting point for discussion of additional advanced topics, such as doubly robust methods, inverse probability weighting, and methods for complex longitudinal data such marginal structural models.

The second development is that machine learning and artificial intelligence has become a hot topic across many fields of science including epidemiology. Traditionally, many risk factor epidemiologists have loathed ‘black box epidemiology’, where statistical associations are disconnected from underlying theory. And yet big data has become increasingly common, luckily also accompanied by a similar increase in access to computational hardware and software that can handle such data. It is in such high-dimensional data settings that machine learning has demonstrated its epidemiological value, including in genome-wide association studies, neuroimaging data, and risk prediction and forecasting. Propensity score
modeling, with its two stage modeling approach, is of course a natural fit for the incorpo-
ration of machine learning methods in the exposure model, where the pursuit of covariate
balance can tolerate ‘black-box methods’, even while theoretical considerations of etiology
are retained in the outcome model. In their article 40 years ago, Rosenbaum and Rubin
estimated the propensity score using a logit stepwise model which, while not a particularly
advanced machine learning algorithm, is indisputably an automated algorithm devoid of
etiological theory. Building on Rosenbaum and Rubin’s idea of using automated model
selection, I, along with Elizabeth Stuart and Justin Lessler, demonstrated that certain en-
semble learning algorithms for propensity score estimation were superior to standard logistic
regression models that were standard practice at the time.7,8 A slew of similar automated
selection algorithms have been incorporated into contemporary causal inference methods,
such as usage of high-dimensional propensity score methods, SuperLearner, and more.9

In conclusion, the current central importance of the propensity score to the epidemi-
ologist’s toolbox is undisputed. I highlight two aspects of the propensity score seeded by
Rosenbaum and Rubin 40 years ago that may have contributed to its widespread use to-
day. While propensity score methodology is still continually being refined and optimized, in
their seminal paper, Rosenbaum and Rubin laid out a fundamentally important discussion
of theory and application that remains to this day a bedrock for further development.

References

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational
3. Hoover RN. The sound and the fury: was it all worth it? Epidemiology. Nov
2008;19(6):780-2; discussion 789-93. doi:10.1097/EDE.0b013e318188e21d
4. Maldonado G. The role of counterfactual theory in causal reasoning. Annals of
5. Greenland S, Gago-Dominguez M, Castelao JE. The value of risk-factor ("black-box")
7. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting.
8. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine