Collection of Biostatistics Research Archive
COBRA Preprint Series

Augmentation of Propensity Scores for Medical Records-based Research

Mikel Aickin*

*mikel@mikelaickin.us

Augmentation of Propensity Scores for Medical Records-based Research

Mikel Aickin

Abstract

Therapeutic research based on electronic medical records suffers from the possibility of various kinds of confounding. Over the past 30 years, propensity scores have increasingly been used to try to reduce this possibility. In this article a gap is identified in the propensity score methodology, and it is proposed to augment traditional treatment-propensity scores with outcome-propensity scores, thereby removing all other aspects of common causes from the analysis of treatment effects.

Augmentation of Propensity Scores for Medical Records-based Research

Mikel Aickin
June 2013

Abstract

Therapeutic research based on electronic medical records suffers from the possibility of various kinds of confounding. Over the past 30 years, propensity scores have increasingly been used to try to reduce this possibility. In this article a gap is identified in the propensity score methodology, and it is proposed to augment traditional treatment-propensity scores with outcome-propensity scores, thereby removing all other aspects of common causes from the analysis of treatment effects.

Keywords: confounding, electronic medical records, complete-propensity scores

1. Introduction

Although an enormous store of data resides in electronic medical records, a chief barrier to its use for research is the belief that therapeutic harms and benefits are irretrievably confounded. If a reasonably rich set of the potential common causes (confounders) can be identified in the medical record, then the use of propensity score analysis can greatly reduce the size and complexity of the confounding problem. The purpose of this note is to give a preliminary indication of a gap in the theory as it is currently used, and to propose a remedy through the introduction of outcome-propensity scores, to be used in addition to the conventional treatment-propensity scores. The theoretical arguments and issues of practical implementation are sketched out here.

2. Conventional Propensity

The issues around propensity scores are most easily understood in terms of the counterfactual causal model (Aickin 2012). This says that outcomes are a function of treatments and common causes,

$$y = \eta(x,z,w)$$

Here y is an outcome measure, x is a treatment, and (z,w) are common causes. η is a population level function that gives the outcome for any one of several methods of determining the treatment. It is assumed that z (which will generally be multivariate) is a minimal set of common causes, in the sense that conditional on z, w no longer produces counfounding of treatment effects.

Thus the model allows z (but not w) to affect the choice of treatment, which is part of what creates the potential confounding of treatment effects. The treatment-propensity score is $pr(x=\xi|z)$, the probability of treatment ξ given the (minimal set of) common causes. The seminal paper of Rosenbaum and Rubin (1983) showed that

$$[x=\xi] \perp z \mid pr(x=\xi|z)$$

meaning that whether treatment ξ happens is independent of all the common causes, given the treatment-propensity score $pr(x=\xi|z)$. This has the immediate consequence that

$$pr(y|x=\xi, pr(x=\xi|z)) = pr(\eta(x,z,w)|x= \xi, pr(x=\xi|z)) = pr(\eta(x,z,w)|pr(x=\xi|z))$$

The extreme left term can be estimated from a medical study, and the extreme right term gives the distribution of the causal outcome, which may be compared between treatments to derive causal effects of treatments. In practice, when y is binary then it is replaced by y=1 above. In general, it can be replaced by y$\in$ B, or by the conditional expectation of y, and the argument remains valid.

Rosenbaum and Rubin's propensity score has been used increasingly in nonintervention medical research (Austin 2008), although arguably not in the way that they developed it. One substantial gap concerns whether the analysis should commence from $pr(y|x=\xi, pr(x=\xi|z))$ or from $pr(y|x=\xi, z, pr(x=\xi|z))$, since the additional z cannot generally be removed from the latter expression. This has led to further research on the wisdom of including z in this way (Rubin & Thomas 2000, for example)

3. Complete Propensity

The purpose of conditioning on a function of the common causes is to reduce, or perhaps even

eliminate confounding. Of course confounding is a relationship between outcome and treatment, but the "propensity score" (here treatment-propensity) only removes part of the problem. The outcome-propensity scores are $pr(y|x=\xi,z)$ for various treatments ξ. Its important property is

$$[y\in B],[x=\xi] \perp z \mid pr(y\in B|x=\xi,z), pr(x=\xi|z).$$

Taken together, the treatment- and outcome-propensities form the complete-propensity scores. Conditioning on them removes any issue of residual effects of common causes, since

$$pr(y|x=\xi,z,pr(y|x=\xi,z), pr(x=\xi|z)) = pr(y|x=\xi,pr(y|x=\xi,z), pr(x=\xi|z)) =$$
$$= pr(\eta(x,z,w)|pr(y|x=\xi,z), pr(x=\xi|z))$$

Thus by conditioning on both treatment- and outcome-propensities, the remainder of the common causes can be ignored, and the result is still a causal analysis by the counterfactual model.

4. Medical Records Applications

Perhaps the greatest field for the employment of propensity scores is in therapeutic research based on electronic medical records (Rubin 1997). The advantages of medical records are the very large numbers of patients involved, the fact that the patients are exactly those for whom therapeutic information is desired (unlike the hightly selected samples of randomized controlled trials), and the fact that the data are already collected. The drawbacks are that medical records seldom contain research-quality data, especialliy on outcomes, and there is always the potential that common causes have created (or destroyed) the actual relationships between treatments and outcomes.

This latter point is the one that potentially can be addressed by propensity scores. The procedure would be as follows. Identify the pool of all patients suffering at some point in time from the condition of interest. Extract their data, including whatever the particular medical record system has to offer, and certainly including treatments received. Attempt to identify the potential common causes. Estimate the treatment-propensities (probabilities of various treatments given the full battery of common causes). Estimate the outcome-propensities (probabilities of varous outcomes given each treatment in turn, and conditional on the common causes). This gives a battery of propensity scores for use at the next step.

Form groups of patients who are the same (or more practically, highly similar) with regard to their battery of complete-propensities. This can be done in a number of ways, of which the formation of matched comparison groups (MCGs) is perhaps the simplest (Aickin 2012). This step corresponds to conditioning on the complete-propensity scores. Some MCGs will be too small, or not contain patients who had the different treatments to be compared. These MCGs are uniformative, and are omitted from the analysis. For each of the remaining MCGs compute a within-group treatment effect estimate. There could be seveeral of these, if there are multiple treatments to be compared. Cumulate these MCG-level effects to the level of the entire patient sample.

It is clear that there may be several ways to accomplish some of these steps, and the researchers should be encouraged to explore as many of them as possible, in order to support the robustness of the results that are ultimately presented.

5. Discussion

The main reason cited for reluctance to accept the results of nonintervention studies, whether in medicine or elsewhere, is the potential for confounding due to common causes. While the same

threat exists in randomized intervention studies, it is felt to be reduced (or sometimes, falsely, eliminated) by the randomization itself. Especially in medical records-based studies the collection of potential common causes can be very large, and the process of weeding through them to pick a small group for conditioning is frustrated by the very large number of implied analysis, and the frequent ambiguities between the benefits and drawbacks of different choices. Thus the Rosenbaum-Rubin propensity score approach, reducing the conditioning variables to a few (often only one), promised to very substantially shrink the magnitude of the problem. Their development left an important gap, however, regarding the potential re-introduction of aspects of the common causes into the analysis. This seems perhaps confusing, since the point of propensity score conditioning is to remove common cause influences, but it is a real problem, because the Rosenbaum-Rubin approach does not remove the resdual effects of common causes on the relationship between outcome and treatment. By adding the outcome-propensities to their treatment-propensities, this separation of common causes from treatment effects is complete. Although it requires additional analytic work, this is not very much more than required by the development and employment of the original propensity score.

References

Aickin M. Analysis of nonintervention studies: technical supplement. The Permanente Journal 2012;16(4);e100-120 (http://dx.doi.org/TPP/12-046)

Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Statistics in Medicine 2008;27:2037-2049

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55

Rubin DB. Estimating causal effects from large data sets using propensity scores. Annals of Internal Medicine 1997;127:757-763

Rubin DB, Thomas N. Combining propensity score matching with additional adjustments fo prognostic covariates. Journal of the American Statistical Association 2009;95:573-565