AAIM Perspectives

AAIM is the largest academically focused specialty organization representing departments of internal medicine at medical schools and teaching hospitals in the United States and Canada. As a consortium of five organizations, AAIM represents department chairs and chiefs; clerkship, residency, and fellowship program directors; division chiefs; and academic and business administrators as well as other faculty and staff in departments of internal medicine and their divisions.

Challenges of Assessing Therapeutic or Diagnostic Outcomes with Observational Data

William S. Weintraub, MD,[a] Robert W. Yeh, MD, MSc, MBA[b]

[a]*MedStar Washington Hospital Center, Northwest DC;* [b]*Beth Israel Deaconess Medical Center, Boston, Mass.*

KEYWORDS: Epidemiology; Observational data; Outcomes research; Selection bias

Medical care is replete with decisions on diagnostic approaches and therapeutic strategies. To establish the evidence that can help clinicians decide on the correct choice, patients are sampled from the larger population under consideration, studied in detail, and results in these patients are then applied more broadly. Two approaches are observational studies and randomized trials. Publications resulting from observational studies will often offer considerable detail on their methods and then in the limitations section offer a statement to the effect that "While residual selection bias cannot be excluded, we controlled for all measured differences between the treatment groups." What does it mean; how important is it, in general and in any one study; and are randomized trials necessary to overcome this particular problem?

In randomized trials, patients drawn from the larger population are assigned to one therapy or the other by a process of random selection. For a randomized trial to proceed properly, there should be equipoise between the study arms, that is, patients and clinicians do not favor one choice over the other. The randomized trial has been the gold standard approach for the singular reason that it can, at least in principle, eliminate the bias that might be created by the nonrandom selection of treatment by clinicians and patients, that is, "treatment selection bias." Randomization results in the subjects in each arm having similar characteristics, both characteristics that are measured and that are unmeasured. Another type of selection bias can occur in randomized trials, in which the patients selected for a trial are not representative of patients being considered for a therapy, that is, the trial results cannot be generalized to the broader patient population. In all studies randomized or not, selection of appropriate patient populations to resolve clinical questions is crucial. Randomized trials are also expensive to mount, can become outdated, and then may not be repeated due to either lack of resources or lack of equipoise.

Due to the limitations of randomized trials, investigators have considered various nonrandomized approaches. While many variations exist, they generally fall into 2 categories, case-control and cohort studies. In case-control studies, patients with and without an outcome of interest are compared for a prior exposure. In a cohort study, patients are followed from a point of inception, and then individuals with and without exposure can be compared for the incidence of an outcome

Funding: Funded in part by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number U54-GM104941 (Principal Investigator: Binder-Macleod).

Conflicts of Interest: There are no conflicts of interest.

Authorship: Both authors had access to the data and a role in writing the manuscript.

Requests for reprints should be addressed to William S. Weintraub, MD, MedStar Heart & Vascular Institute Suite 4B1, MedStar Washington Hospital Center, 110 Irving Street Northwest, DC 20010.

E-mail address: william.s.weintraub@medstar.net

0002-9343/$ - see front matter © 2017 Published by Elsevier Inc.
https://doi.org/10.1016/j.amjmed.2017.10.027

of interest. In both case control and cohort studies, there can be difficulties with exposure due to variable adherence or crossover from one form of therapy to the other.

The major limitation with an observational study compared with a randomized trial is the influence of "treatment selection bias."[1] Treatment selection bias can occur when the therapeutic selection is influenced by patient characteristics, including severity and acuteness of illness and comorbidity. When such variables also are associated with better or worse outcomes, they are called confounders. If unaccounted for, confounders will bias the result of observational studies, potentially resulting in erroneous conclusions. The literature is replete with observational studies that showed different results from randomized trials, likely due to confounding.[2,3] Statistical methods can be used to reduce the bias due to confounding where the confounding variables are measured. However, unmeasured confounders may make it difficult to assess whether a difference (or lack thereof) between the treatment arms is due to therapeutic effect or to residual confounding. Missing or misclassification of covariates (patient characteristics) is a particular problem in observational studies; in contrast to randomized trials, these data are necessary to correct for potential confounding. In all studies, randomized or not, correct classification of outcomes and as complete follow-up as possible are crucial.

PERSPECTIVES VIEWPOINTS

- Observational assessments of therapeutic or diagnostic options are subject to treatment selection bias.
- Statistical methods can reduce treatment selection bias but cannot account for unmeasured confounders.
- Simulation modeling can help assess whether there is residual selection bias.
- Sample size cannot overcome bias.
- Data are most believable when randomized clinical trial data and observational point in the same direction.

STATISTICAL METHODS TO REDUCE TREATMENT SELECTION BIAS

Various statistical approaches have been proposed to reduce treatment selection bias. The most common approaches employ variations of multivariable analysis. The most straightforward is to use logistic regression or Cox modeling, in which the treatment is included as a covariate along with measured potential confounders. Another series of approaches use propensity analysis.[4] This approach starts with identifying a set of variables that are related to the propensity to choose one form of therapy over the other. Using a logistic regression model, the probability of receiving one form of therapy vs the other (ie, the propensity score) may be calculated for each patient. Propensity scores can be used in several different ways, including creating matched groups, stratifying outcomes, or by reweighting of samples based on the inverse probability of receiving the treatment received.[5] All of these methods for overcoming treatment selection bias are limited to accounting for the bias induced by measured variables. None can account for unmeasured confounding variables. A recent study by Elze et al[6] found little, if any, advantage to propensity score approaches over multivariable analysis in the ability to reduce treatment selection bias.

An alternative approach to propensity analysis and other multivariate techniques is to use an instrumental variable.[7] An instrumental variable should cleanly separate the groups, with the distribution of variables the same in the resulting groups, except for the treatment variable, which should be quite different between the groups. The instrumental variable should not be associated with the outcome, apart from through its association with the treatment variable. The best instrumental variable is randomization, which cleanly separates the groups, but does not by itself predict outcome. An increasingly popular instrumental variable is Mendelian randomization in a genetic study.[8] For instance, the Mendelian randomization to familial hypercholesterolemia results in much higher risk of cardiovascular disease than individuals not randomized to familial hypercholesterolemia.[9] Attempts to develop other instrumental variables, such as geographic location or practice styles of physicians, have been used but may not fulfill the necessary assumptions for validity, because of differences that remain between the groups in critical variables, including unmeasured confounders.[10]

The effect of an unmeasured or a group of confounders may be evaluated by sensitivity analysis. In the method of Lin et al,[11] the observed difference in outcome is considered with the potential prevalence of the confounder in each arm. For any such pair of prevalences of the confounder in the 2 arms, the strength of the confounder, generally expressed as a hazard ratio that would explain the difference between the treatment arms, may then be calculated. This method cannot find the potential unmeasured confounder, but can provide insight into whether such a confounder is likely.

EXAMPLES OF OBSERVATIONAL STUDIES

An observational study is likely to be found to be credible where investigators, clinicians, and other stakeholders feel that treatment selection bias is likely to be minimized. An example is closure devices for vascular