Special Series on Research Methodology

Propensity Scores: A Practical Method for Assessing Treatment Effects in Pain and Symptom Management Research

Melissa M. Garrido, PhD

Geriatric Research, Education and Clinical Center, James J. Peters VA Medical Center, Bronx, New York; and Brookdale Department of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Abstract

When conducting research on pain and symptom management interventions for seriously ill individuals, randomized controlled trials are not always feasible or ethical to conduct. Secondary analyses of observational data sets that include information on treatments experienced and outcomes for individuals who did and did not receive a given treatment can be conducted, but confounding because of selection bias can obscure the treatment effect in which one is interested. Propensity scores provide a way to adjust for observable characteristics that differ between treatment and comparison groups. This article provides conceptual guidance in addition to an empirical example to illustrate two areas of propensity score analysis that often lead to confusion in practice: covariate selection and interpretation of resultant treatment effects. J Pain Symptom Manage 2014;48:711–718. *Published by Elsevier Inc. on behalf of American Academy of Hospice and Palliative Medicine.*

Key Words

Propensity score, treatment effect, causality, randomized controlled trial, palliative care

Introduction

Randomized controlled trials (RCTs) are often considered the gold standard of research designs for estimating treatment effects.[1] However, when conducting research on palliative care or other pain and symptom management interventions involving patients with serious illnesses, RCTs are not always feasible or ethical to conduct.[2] Concerns about participant burden are especially salient in this population. In addition, the results of RCTs may not generalize to real-life circumstances as they are focused on efficacy rather than effectiveness. Interventions that require buy-in from patients, physicians, and family members might not be modeled realistically by RCTs.[2]

One solution to the ethical, feasibility, and generalizability concerns of RCTs is for researchers to perform secondary analyses of observational data sets that include information on treatments experienced and outcomes for individuals who did and did not receive a given treatment.[3] In observational data sets, however, there is a high likelihood of encountering confounding because of selection bias. That is, there may be patient characteristics that are associated

Address correspondence to: Melissa M. Garrido, PhD, Geriatric Research, Education and Clinical Center, James J. Peters VA Medical Center, 4A-17, 130 W. Kingsbridge Road, Bronx, NY 10468, USA. E-mail: melissa.garrido@mssm.edu

Accepted for publication: May 28, 2014.

Published by Elsevier Inc. on behalf of American Academy of Hospice and Palliative Medicine.

0885-3924/$ - see front matter
http://dx.doi.org/10.1016/j.jpainsymman.2014.05.014

both with likelihood of receiving treatment and with the experienced outcome. In the later sections, one method for adjusting for observable selection bias, propensity scores, is outlined. Other methods, including instrumental variables and difference-in-difference models, are not discussed here. After briefly describing what propensity scores can and cannot do, an empirical example is used to outline issues with variable selection for propensity scores and comparison of resultant treatment effects with those derived from RCTs, two areas of propensity score analyses that often lead to confusion in practice.

Understanding What Propensity Scores Can Do

To know the true effect of a treatment on an individual (or other unit of observation), one would need to simultaneously observe the individual with and without the treatment. Because this is impossible, researchers create estimates of the counterfactual (the outcome in the unobserved alternate treatment state). In an RCT, this is done by randomly assigning individuals to receive or not receive a treatment. By randomly distributing observed and unobserved characteristics of individuals across the treated and control groups, outcomes in the control group provide the estimate of the counterfactual, and any difference in outcomes across groups (in a well-designed RCT[4]) can be attributed to the treatment. In an observational data set, when individuals are not randomly assigned to treatment, there is a greater chance that individuals will not have the same distribution of observed and unobserved characteristics in the treatment and comparison groups, making it more difficult to find comparable observations from which to estimate the counterfactual. For example, individuals with greater symptom severity may be more likely to receive palliative care than individuals with less severe physical symptoms. These sicker individuals may have systematically different survival times and hospital use than the healthier individuals, regardless of participation in palliative care. Propensity scores provide a way to adjust for observable characteristics that differ between treatment and comparison groups and that might obscure the treatment effect.[5]

Propensity score analysis is based on the assumption of strongly ignorable treatment assignment. This means that given a set of observed covariates, treatment assignment and outcome are independent.[5–7] The theory behind propensity scores states that if individuals with the same distribution of observed characteristics have an equal nonzero chance of receiving treatment, this distribution of characteristics can be compressed into a single score (the propensity score) that represents the likelihood of receiving treatment. Outcomes are then compared among individuals with similar propensity scores in the treatment and comparison groups. This allows the treatment effect to be estimated while adjusting for confounding because of observable selection bias.

Understanding What Propensity Scores Cannot Do

Unobservable characteristics are not balanced by propensity scores. If these unobserved characteristics are also confounders, their absence from the propensity score may exacerbate bias of treatment effect estimates.[8,9] That is, the distance between the estimated and true treatment effect will increase.

In addition, propensity scores can only be used to compare groups of individuals who all have some nonzero chance of receiving treatment. If there are patients who would always or who would never receive treatment, they will not be included in a propensity score-adjusted sample. These patients have no counterpart in the other group from which to estimate a counterfactual, preventing estimation of a treatment effect.

Empirical Example

To illustrate propensity score analysis, consider the following empirical example. Psychological distress and psychological disorders are associated with higher likelihood of multiple hospital admissions,[10] but the degree to which in-hospital mental health care reduces the likelihood of readmission is unclear. Designing an RCT of in-hospital mental health care's effect on all-cause 30-day readmissions would be expensive and difficult to implement. Instead, we can look at readmission rates in a pre-existing data set of seriously ill veterans at risk of hospital readmission.

The data for this example come from medical record review of a sample of 209