COMPARING PROPENSITY SCORE METHODS FOR CREATING COMPARABLE COHORTS OF CHIROPRACTIC USERS AND NONUSERS IN OLDER, MULTIPLY COMORBID MEDICARE PATIENTS WITH CHRONIC LOW BACK PAIN

William B. Weeks, MD, PhD, MBA,[a b] Tor D. Tosteson, ScD,[c] James M. Whedon, DC, MS,[d e] Brent Leininger, DC,[f] Jon D. Lurie, MD, MS,[g] Rand Swenson, DC, MD, PhD,[c] Christine M. Goertz, PhD, DC,[h] and Alistair J. O'Malley, PhD[i]

ABSTRACT

Objective: Patients who use complementary and integrative health services like chiropractic manipulative treatment (CMT) often have different characteristics than do patients who do not, and these differences can confound attempts to compare outcomes across treatment groups, particularly in observational studies when selection bias may occur. The purposes of this study were to provide an overview on how propensity scoring methods can be used to address selection bias by balancing treatment groups on key variables and to use Medicare data to compare different methods for doing so.

Methods: We described 2 propensity score methods (matching and weighting). Then we used Medicare data from 2006 to 2012 on older, multiply comorbid patients who had a chronic low back pain episode to demonstrate the impact of applying methods on the balance of demographics of patients between 2 treatment groups (those who received only CMT and those who received no CMT during their episodes).

Results: Before application of propensity score methods, patients who used only CMT had different characteristics from those who did not. Propensity score matching diminished observed differences across the treatment groups at the expense of reduced sample size. However, propensity score weighting achieved balance in patient characteristics between the groups and allowed us to keep the entire sample.

Conclusions: Although propensity score matching and weighting have similar effects in terms of balancing covariates, weighting has the advantage of maintaining sample size, preserving external validity, and generalizing more naturally to comparisons of 3 or more treatment groups. Researchers should carefully consider which propensity score method to use, as using different methods can generate different results. (J Manipulative Physiol Ther 2015;38:620-628)

Key Indexing Terms: *Chiropractic Manipulation; Medicare; Propensity Score*

[a] Professor, The Geisel School of Medicine at Dartmouth, The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH.

[b] Chair, Health Services and Clinical Research Program, Palmer Center for Chiropractic Research, Palmer College of Chiropractic, Davenport, IA.

[c] Professor, The Geisel School of Medicine at Dartmouth, The Dartmouth Institute for Health Policy and Clinical Practice, Davenport, IA.

[d] Instructor, The Geisel School of Medicine at Dartmouth, The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH.

[e] Director, Health Services Research, Southern California University of Health Sciences, Whittier, CA.

[f] Research Fellow, Integrative Health & Wellbeing Research Program, Center for Spirituality & Healing, University of Minnesota, Minneapolis, MN.

[g] Associate Professor, The Geisel School of Medicine at Dartmouth, The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH.

[h] Vice Chancellor, Palmer College of Chiropractic, and Director, Palmer Center for Chiropractic Research, Davenport, IA.

[i] Professor, The Geisel School of Medicine at Dartmouth, Department of Biomedical Data Science, The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH.

Submit requests for reprints to: William B. Weeks, MD, PhD, MBA, Professor, TDI, One Medical Center Drive, Lebanon, NH 03766. (e-mail: *wbw@dartmouth.edu*).

Paper submitted May 13, 2015; in revised form June 16, 2015; accepted June 16, 2015.

0161-4754

http://dx.doi.org/10.1016/j.jmpt.2015.10.005

Health services research methods, such as analysis of large claims databases, are an important tool for research on effectiveness and efficiency of complementary and integrative health services (CIHS) such as spinal manipulative therapy (SMT).[1] These methods may be particularly useful when determining whether patients who use such care for certain conditions might have lower overall care costs when compared with patients who do not,[2] a so-called "medical care cost offset" that has been demonstrated, for instance, with pharmaceutical[3,4] and mental health[5,6] treatment.

The very high variability of health care expenditures for individuals[7] might distort cost-offset analysis when sample sizes are small, as they frequently are in randomized controlled trials that examine the effectiveness of SMT. As a recent report on randomized controlled trials that studied the clinical effectiveness of SMT for neck and low back pain found a median sample size of 95 and an interquartile range of 47 to 199,[8] research efforts searching for medical care cost offsets might need to use large databases or nationally weighted surveys[9] to conduct observational comparative effectiveness studies with adequate sample sizes.

However, the use of observational data from large data sets or surveys to compare groups that do or do not use a particular treatment modality may be confounded by selection bias.[10,11] For instance, studies that have compared users to nonusers of a common type of SMT that is provided by doctors of chiropractic (DCs; known as chiropractic manipulative treatment [CMT]) show that CMT users are younger, wealthier, and healthier than nonusers and more likely to be insured and female.[12–14] These demographic characteristics have been associated with different health and cost outcomes when comparing CMT to medical care for treatment of low back pain.[15] Therefore, it is critical to recognize potential selection bias when attempting to compare treatment groups in observational studies of patients who use CMT.

Although traditional risk adjustment for demographic differences in patient populations through risk stratification and regression adjustment have narrowed cost differences between patients who obtained CMT and those who sought conventional medical care for back pain,[16] newer statistical methods that calculate the propensity of patients to self-select into 1 treatment group or another, based on a variety of demographic and use variables, and use those calculations for further analysis, are increasingly used in light of their improved performance in estimating causal effects.[17,18]

To date, several studies have applied propensity score methods to analysis of survey data in an effort to compare CIHS users to nonusers. In a study that found that most CIHS treatment of back and neck problems was provided by DCs, Martin et al[19] used the Medicare Expenditure Panel Survey to compare costs of treatment of back and neck problems for patients who did and did not use CIHS: propensity scores for the 2 groups were used to match patients who had propensity scores within a "region of common support" and then identify the "nearest neighbor" as a method to develop 2 matched groups for comparison. Weigel et al[20] have applied inverse propensity score weighting to data from "Medicare Current Beneficiary Survey" respondents and to "Assets and Health Dynamics among the Oldest Old" interviews[21] when comparing use and outcomes of chiropractic and conventional medical treatment of back pain in the Medicare population. However, to date, propensity score methods have not been applied directly to Medicare claims data to evaluate costs of care for patients seeking back pain treatment.

This study has 2 purposes. First, the article provides an overview and brief tutorial on how propensity score methods work and the different types of propensity score methods that are commonly used in order to help readers of the CIHS literature understand these fairly complex methods. Second, the study uses Medicare data from 2006 to 2012 on older, multiply comorbid patients who had a chronic low back pain (cLBP) episode to demonstrate the impact of applying different propensity score methods on the balance of demographics of patients who were in 1 of 2 treatment groups: those who received only CMT and those who received no CMT during their episodes.

Methods

Overview of Propensity Score Methods

Propensity scoring methods can be used to address possible confounding in observational studies where investigators have no control over treatment assignment, such as when they retroactively analyze health care claims databases. Conditional on a set of observed covariates, the propensity score for a patient is defined as the probability of a patient with the same observed characteristics being in the treatment group.[22] The application of propensity score methods is designed to construct a new analytic data set in which treatment groups are balanced on observed confounders so that the outcomes for the different treatment groups can be directly compared.

For instance, imagine that there were 100 patients with low back pain who used CMT and 200 who did not, and the 2 groups differed in their age and sex distributions (Fig 1). When examining the actual data, it is evident that the numbers, proportions of males and females, and ages of the 2 different treatment groups differed. Propensity scores based on age and sex could be used to identify and match patients who have a similar likelihood to be in a particular treatment group even if those cases differed individually on age and sex (shown in red in the figure).[23] However, using such a matching process limits the analysis to those with similar scores, and the number of cases that can be used for analysis will decline in both treatment groups. This is generally true whether one uses a 1-to-many (where 1 case