ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Propensity score matching and randomization

George Mnatzaganian[a,*], David C. Davidson[b], Janet E. Hiller[c,d], Philip Ryan[d]

[a]School of Allied Health, Faculty of Health Sciences, Australian Catholic University, Level 2, Daniel Mannix Building, 17 Young Street, Fitzroy, Victoria 3065, Australia

[b]Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia

[c]School of Health Sciences, Faculty of Health, Arts and Design, Swinburne University of Technology, John Street, Hawthorn, Victoria 3122, Australia

[d]Discipline of Public Health, School of Population Health, The University of Adelaide, North Terrace, Adelaide, South Australia 5000, Australia

Accepted 5 January 2015; Published online xxxx

Abstract

Objectives: We used elective total joint replacement (TJR) as a case study to demonstrate selection bias toward offering this procedure to younger and healthier patients.

Study Design and Setting: Longitudinal data from 2,202 men were integrated with hospital data and mortality records. Study participants were followed from recruitment (1996−1999) until TJR, death, or 2007 (end of follow-up). A propensity score (PS) was constructed to quantify each subject's likelihood of undergoing TJR. TJR recipients were later matched to their non-TJR counterparts by PS and year of hospitalization. Ten-year mortality from index admission was compared between cases and controls.

Results: Overall, 819 (37.2%) had TJR. Those were younger, healthier, and belonged to higher socioeconomic classes compared with those who were not proposed for surgery. Of the TJR recipients, 718 were matched to 1,109 controls. Cases and controls had similar characteristics and similar years of follow-up from recruitment till index admission. Nonetheless, controls were more likely to die (39.5%) compared with 14.5% in TJR cases ($P < 0.001$).

Conclusion: Selection for elective procedures may introduce bias in prognostic features not accounted for by PS matching. Caution must be exercised when long-term outcomes are compared between surgical and nonsurgical groups in a population at risk for that surgical procedure.

Keywords: Elective total joint arthroplasty; Selection bias; Channeling bias; Propensity score matching; Confounding; Internal and external validity; Randomization

1. Introduction

In nonrandomized studies, except by chance, the characteristics of patients constituting comparison groups of interest are different. The differences in characteristics between the compared nonrandomized groups (eg, surgical vs. nonsurgical) may be large, systematic, and statistically significant. Often such differences arise from clinically motivated patient selection that is not documented [1,2]. In 1983, Rosenbaum and Rubin [3] proposed a method—propensity score (PS) analysis—as an alternative tool to adjust for confounding and reduce selection bias in such nonrandomized studies. This calculated numerical score describes the expected likelihood for a study participant to receive a treatment (or an exposure of interest) conditional on the participant's observed pretreatment covariates. After its construction, the score can be used to control for confounding in the comparison of outcomes between treatment groups. This can be done through regression adjustment in form of adding the PS as a covariate or a weight into an outcome model, or through stratification or matching [3−5]. In stratification, those with and without the treatment are stratified by the score's categories (eg, quintiles), and then comparisons of outcome are run for each stratum [4]. Rosenbaum and Rubin [4] reported that stratifying on the quintiles of a PS will remove 90% of the bias because of measured confounders and risk factors when estimating a linear treatment effect. Within each stratum, if the PS has been correctly specified, those with and without a treatment of interest will have similar distribution of measured baseline covariates [6]. In the third

Funding: This study was supported by the University of Adelaide.

* Corresponding author. Tel.: +61-3-9953-3637; fax: +61-3-9953-3385.

E-mail address: george.mnatz@gmail.com (G. Mnatzaganian).

http://dx.doi.org/10.1016/j.jclinepi.2015.01.002
0895-4356/

What is new?

Key findings

- Propensity score (PS) matching can balance baseline characteristics of the PS-matched groups that are being compared for a study outcome.
- PSs can only account for known covariates; unknown covariates cannot be accounted for.
- PS matching is not equivalent to randomization.
- With PS matching, we used elective total joint arthroplasty as a case study to demonstrate selection bias toward offering this procedure to younger and healthier patients.
- This is the first study to demonstrate selection bias toward offering TJR to younger, healthier patients, and those belonging to higher socioeconomic classes.
- Selection for elective procedures may introduce bias in prognostic features not accounted for by PS matching.
- Caution must be exercised when long-term outcomes are compared between surgical and nonsurgical groups in a population at risk for that surgical procedure.

method that adjusts for measured confounding, subjects with and without the treatment of interest are matched by their respective PS, thus avoiding the stratified analyses with too many strata. However, often not everyone can be matched, which in turn reduces the sample size and power. Matching by a PS can provide balance in baseline characteristics between the PS-matched groups that are being compared [3−6]. Nonetheless, such scores only account for pretreatment observed covariates. Unobserved factors that influence selection of patients for the treatment or intervention of interest cannot be accounted for [5,7]. Thus, it has been argued that PS matching on observed variables can potentially increase the bias caused by unobserved confounders [7].

A recent PS-matched observational study compared cardiovascular outcomes in a small sample of patients with osteoarthritis (OA) who had and did not have an elective total hip or knee arthroplasty, showing that such elective surgery was associated with a significant reduction in subsequent cardiovascular events [8]. The authors argued that their findings could be explained by the improved physical activity in their 153 patients undergoing arthroplasty and also by the anticipated reduced use of potentially cardiotoxic nonsteroidal anti-inflammatory drugs [8]. However, these findings could have resulted from the initial selection of healthier patients to this elective procedure. Using elective primary total joint replacement (TJR) as a case study, we conducted the current analysis to explore this issue.

The objectives of this present study were not only to show the positive salient features of a constructed PS by demonstrating its prominent balancing properties but also to demonstrate the main limitation in such scores, namely their inability to account for unmeasured confounders, and show that PS matching is not equivalent to randomization. Using TJR as a case study, we show how younger, healthier, and those belonging to higher socioeconomic classes are more likely to be proposed for surgery compared with older, sicker, and more socioeconomically disadvantaged patients.

2. Methods

2.1. Ethics statement

Ethical approval for the study was obtained from the Human Research Ethics Committees of the University of Adelaide (H-106-2009) and WA Department of Health (AHEC EC004220). All analyses used deidentified data. The need for informed consent was waived by the ethical committees due to deidentified data being used.

2.2. Data sources and study population

The study population was described previously [9−11]. Briefly, it was drawn from the Health In Men Study (HIMS) [11], which arose from a randomized population-based trial of ultrasound screening for abdominal aortic aneurysm in men aged 65−83 years living in Perth, Western Australia (WA). In 1996−1999, a total of 12,203 men attended the baseline screening and provided detailed health and other information including information on diet, alcohol consumption, a comprehensive smoking history, medications used, presence of chronic diseases, and information on physical activity and exercise. In addition, study nurses recorded blood pressure, weight, height, and waist and hip circumferences. These clinical data were then integrated with WA hospital morbidity data (HMD) [12], Cancer Registry, Mental Health Services System, and mortality records. Linkage with HMD was used to identify total hip or knee replacement procedures, presence of morbidity, and readmission [10]. All-cause mortality was ascertained through linkage with WA mortality records. The hospital morbidity database includes demographic, diagnostic, and procedural information on all patients discharged from all public and private hospitals in WA. This database allows the inclusion of up to 21 diagnoses and 11 procedure codes for each hospitalization. For each participant, any morbidity or health-related outcome was retrieved from the linked data in the period 1970 through 2007 and this enabled