Annals of Epidemiology 22 (2012) 691–697

Contents lists available at SciVerse ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Toward a better understanding of when to apply propensity scoring: a comparison with conventional regression in ethnic disparities research

Yu Ye MA [a,*], Jason C. Bond PhD [a], Laura A. Schmidt PhD, MSW, MPH [b], Nina Mulia DrPH [a], Tammy W. Tam PhD [a]

[a] Alcohol Research Group, Public Health Institute, Emeryville, CA
[b] Philip R. Lee Institute for Health Policy Studies and Department of Anthropology, History and Social Medicine, University of California, San Francisco, San Francisco, CA

ARTICLE INFO

Article history:
Received 20 January 2012
Accepted 20 July 2012
Available online 14 August 2012

Keywords:
Propensity scores
Multivariable regression
Marginal effect
Confounding
Effect modification
Reference sample
Disparity

ABSTRACT

Purpose: Despite growing popularity of propensity score (PS) methods used in ethnic disparities studies, many researchers lack clear understanding of when to use PS in place of conventional regression models. One such scenario is presented here: When the relationship between ethnicity and primary care utilization is confounded with and modified by socioeconomic status. Here, standard regression fails to produce an overall disparity estimate, whereas PS methods can through the choice of a reference sample (RS) to which the effect estimate is generalized.

Methods: Using data from the National Alcohol Surveys, ethnic disparities between White and Hispanics in access to primary care were estimated using PS methods (PS stratification and weighting), standard logistic regression, and the marginal effects from logistic regression models incorporating effect modification.

Results: Whites, Hispanics, and combined White/Hispanic samples were used separately as the RS. Two strategies utilizing PS generated disparities estimates different from those from standard logistic regression, but similar to marginal odd ratios from logistic regression with ethnicity by covariate interactions included in the model.

Conclusions: When effect modification is present, PS estimates are comparable with marginal estimates from regression models incorporating effect modification. The estimation process requires a priori hypotheses to guide selection of the RS.

Introduction

Epidemiologic studies investigating racial/ethnic disparities in health have grown exponentially over the past few decades [1], in conjunction with the Institute of Medicine's groundbreaking report on disparities in health care [2] and U.S. national objectives put forth in *Healthy People 2000–2020* [3–5]. Until recently, disparities research relied heavily on conventional regression modeling to document inequalities in health and health care across racial/ethnic groups [1]. One common problem with regression, however, is that the relationship between ethnicity and the health outcomes of interest are often confounded with, and modified by, socioeconomic status [6]. This study addresses why and how newer methods based on propensity scores (PS) are particularly well-suited for disparities research, although they are relevant to any area of epidemiologic research where effect modification is similarly of concern.

There is lively debate over when PS methods offer benefits over standard multivariable linear or logistic regression typically used in disparities research [7–10]. Although there is empirical support for PS's advantages [11,12], in practice, PS methods often seem to generate results quite similar to those from multivariable regression [13,14]. This paper is preoccupied with one common scenario in which PS methods should theoretically produce more interpretable effect estimates than those from multivariable regression. This occurs when the distribution of one or more confounding covariates (e.g., socioeconomic status [SES]), as well as the relationship between these confounding covariates and the health outcome, varies across the ethnic groups being compared. Entering relevant interaction terms addresses the effect modification, but the regression model can only produce disparity estimates at specific SES levels and fails to generate a single overall disparity estimate, which, in most practical cases, is desired. Without including interactions, disparity estimates will in general be biased.

Supported by grants from the U.S. National Institute on Alcohol Abuse and Alcoholism R01 AA017197 and P50 AA005595.

* Corresponding author. 6475 Christie Avenue, Suite 400, Emeryville, CA 94608.
E-mail address: yye@arg.org (Y. Ye).

1047-2797/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.annepidem.2012.07.008

With control variables in the model, standard regression generates *conditional* effect estimates. In contrast, PS methods produce marginal effect estimates that can be interpreted counterfactually. In the language of causal inference, which often focuses on the effect of a treatment or of exposure to a risk factor, PS methods estimate the difference in an outcome for a given population when it is treated/exposed, and the outcome for the same population when it is not treated/exposed. PS methods thus estimate the marginal effect by design, largely through specifying a reference sample (RS) to which the effect estimate is generalized. The RS often used is either the "treated" ("risk-exposed") sample or the sample combining those who were and were not treated/risk-exposed [15–17]. With regression generating conditional effect estimates and PS methods producing marginal estimates, questions have been raised regarding the comparability of the two approaches in nonlinear models [10,18]. Specifically, for a dichotomous outcome (as in the present study), it was suggested that marginal odds ratios (ORs) should be generated from both PS and logistic regression for the two methods to be comparable [19,20]. Little is known about the comparability of the two approaches, however, if effect modification is present.

In the current study of ethnic disparities, we compare estimates from PS approaches generated using various RSs to both conditional and marginal ORs produced from ordinary logistic regression. We show how marginal ORs from logistic regression incorporating effect modification can be produced and comparable with PS estimates. Although some recent work has examined the performance of PS methods under effect modification [21–23], this study compares substantive results obtained from PS and logistic regression models; it also shows how varying the specification of the RS can influence results.

Materials and methods

Dataset and measures

Our empirical analysis focuses on the timely issue of racial/ethnic disparities between Whites and Hispanic Americans in access to primary care interventions for alcohol problems [24]. Data come from the combined 2000 and 2005 U.S. National Alcohol Surveys (NAS), two comparable probability samples of U.S. adults collected using computer-assisted telephone interviews via random digit dialing [25]. Included in this analysis are "at-risk" drinkers who meet the National Institute on Alcohol Abuse and Alcoholism drinking guidelines defined as men/women drinking more than 4/3 drinks on any day or more than 14/7 drinks per week [26]. Self-identified ethnicity, based on the current U.S. Census definition, is the key risk-exposure variable in this analysis. Only those self-identifying as White (N = 2798) or Hispanic (N = 684) were retained for analysis, because the greatest disparities have been found between these two groups [24,27,28]. The outcome of interest is the subject's report of whether she or he had one or more primary care visits in the prior year. Any visit with a private doctor, clinic, or non-emergency medical setting during the prior year counted as a primary care visit. Demographic and SES covariates were used as potential confounders and effect modifiers of interest, including gender, age, education, annual household income, and health insurance coverage.

Statistical analysis

PS stratification and weighting

A PS is defined as the estimated probability of being Hispanic versus White and is modeled as a function of gender, age, education, annual household income, and health insurance coverage using logistic regression. Two approaches utilizing the estimated PS to estimate ethnic disparities are considered in the present study, PS stratification and PS weighting.

PS-stratification [15,16] classifies subjects into five strata (quintiles) using their estimated PS. This process is repeated several times, using "Hispanics," "Whites," and the "combined White and Hispanic sample" separately as the RS. When Hispanics are treated as the RS, the Hispanic sample is divided into five equal-sized strata based on the sorted PS distribution within the Hispanic sample. Using the PS thresholds for quintile generation, the comparison White sample is then divided into five groups, presumably of unequal sizes given that the distribution of PS is different between Whites and Hispanics. This is done analogously for Whites as the RS. When the combined Whites/Hispanic sample is treated as the RS, the pooled sample is divided into five equal-sized strata based on the sorted PS distribution for the total sample. The overall response probability, for each choice of RS, is simply a weighted average of stratum-specific response probabilities. Because the RS is divided into five equal size strata in the current design, equal weights are assigned across the five strata and the overall response probability is the simple average of the stratum-specific probabilities. The marginal OR estimate is then derived using the overall response probabilities (see the Appendix for details as well as [19]).

For PS weighting [29], subjects in the two ethnic comparison groups are weighted based on their estimated PS to construct a 'pseudo-population' in which confounders are no longer associated with ethnicity. When using either the White or the Hispanic group as the RS, "standardized mortality ratio" weights are used that assign a weight of 1 to the group chosen as the RS and the propensity odds $\widehat{e}(X)/(1-\widehat{e}(X)$ (where $\widehat{e}(X)$ is the estimated PS) to the non-RS. When using the combined sample as the RS, the inverse-probability-of-treatment-weighted estimator must be used and derives the effect estimate by using the inverse of the PS $(1/\widehat{e}(X))$ as weights for one group and the inverse of 1 minus the PS $(1/(1-\widehat{e}(X))$ for the other. The marginal ORs are estimated by fitting logistic regression using the ethnic indicator as predictor, with the PS weights used as sampling weights in the model estimation [29].

Standardization and RS

Standardization is the traditional approach in epidemiology to obtain an overall effect estimate when a confounder modifies the relationship between the outcome and a risk factor. Standardization takes a weighted average of the stratum-specific rates or risks, with the strata defined by the effect modifier and with weights corresponding to the number of persons in the RS falling into each category of the effect modifier [30]. As noted, disparity estimates from the PS stratification method are created from a weighted average of stratum-specific estimates, which is essentially a standardization process where the PS serves as the effect modifier (see an illustration in [29] Appendix 1). Sato and Matsuyama [31] also show that PS weighting allows for nonparametric standardization using either the exposure group or the total combined groups as the RS.

Conditional and marginal ORs from logistic regression

We first fit a standard logistic regression model, which generates the conditional OR for the ethnic disparity estimate. Using the fitted model, two predicted response probabilities are generated for each individual by plugging in the actual values of their observed covariates (other than ethnicity) into the regression model. Regardless of an individual's actual ethnicity, the first predicted response probability is generated by assuming the individual is White, the second by assuming he/she is Hispanic. Using these two sets of estimated response probabilities for each individual, the separate "White"/"Hispanic" marginal response probabilities are estimated as the simple average across all respondents (thus using the total combined sample as the RS) where everyone was assumed to be