Annals of Epidemiology 25 (2015) 120–125

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Neighborhood deprivation and preterm birth: an application of propensity score matching

Xiaoguang Ma MD, PhD [a,b,*], Nancy L. Fleischer PhD [b,c], Jihong Liu ScD [b], James W. Hardin PhD [b], Guang Zhao PhD [b], Angela D. Liese PhD [b,c]

[a] Department of Nutrition and Food Hygiene, School of Public Health, Zhejiang University, Hangzhou, China
[b] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia
[c] Center for Research in Nutrition and Health Disparities, Arnold School of Public Health, University of South Carolina, Columbia

ARTICLE INFO

Article history:
Received 11 July 2014
Accepted 24 November 2014
Available online 28 November 2014

Keywords:
Neighborhood deprivation index
Preterm birth
Propensity score
Matching
Principal component analysis

ABSTRACT

Purpose: On the basis of a neighborhood deprivation index (NDI), this study aims to examine the association between neighborhood deprivation and preterm birth (PTB) by applying propensity score matching (PSM) methods.
Methods: NDI was calculated for all census tracts in South Carolina based on the US Census data. Live births in South Carolina during 2008 to 2009 ($n = 98,456$) were assigned to an NDI quartile group based on residential addresses. PSM was used to create matched pairs by NDI quartiles to avoid any potential inference on imbalanced data. The differences of prevalence of PTB were calculated for exposed and reference deprivation groups.
Results: Neighborhood deprivation was higher among blacks than whites. The overall prevalence of PTB was 8.5% for whites and 12.6% for blacks. Living in neighborhoods with higher deprivation was associated with increased risk of PTB among blacks compared with living in neighborhoods with lower deprivation among blacks. However, random-effect regression models showed that the most deprived whites experienced 1.13 times the odds of having PTB than the least deprived whites.
Conclusions: The racial disparities in adverse birth outcomes might be partially explained by neighborhood deprivation in South Carolina. PSM may be an appropriate approach to avoid imbalanced data inferences.

Introduction

Racial disparities in adverse birth outcomes are well documented but not well explained [1,2]. Racial disparities vary across geographic regions with different political, economic, and social contexts [3,4], which suggests that studies focusing on environmental factors, including of the neighborhood environment, are needed to explain racial disparities in birth outcomes [5]. The neighborhood socioeconomic environment may shape individual biological and behavior risk factors, which may cause adverse birth outcomes through a variety of biological mechanisms [6]. However, relationships between neighborhood socioeconomic factors and adverse birth outcomes are not consistent across studies. Some studies have identified a relationship between deprived neighborhood conditions and adverse birth outcomes [6–10], some have not [11], and some have only demonstrated the associations among certain racial groups [12–15].

A possible explanation for the inconsistency is that various indicators have been used to characterize the neighborhood socioeconomic context, making results difficult to interpret and compare across studies. The neighborhood deprivation index (NDI), which synthesizes multiple dimensions of the neighborhood socioeconomic context, allows comparisons across geographic areas [16]. This index has been linked to adverse birth outcomes, including low birth weight and preterm birth (PTB) [17–19]. However, the distribution of NDI quartiles can be extremely imbalanced across different racial groups; often, more white women live in less deprived areas and more minority women live in more deprived areas. With the addition of covariates in the adjusted models, certain covariate strata may contain imbalanced data or subjects who could never be exposed to a condition, leading to inference based on no actual data [20].

Conflict of interest: None declared.

* Corresponding author. Department of Nutrition and Food Hygiene, School of Public Health, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang 310058, China. Tel.: +86 571-8820-8591; fax: +86 571-8820-8519.
E-mail address: mxg@zju.edu.cn (X. Ma).

http://dx.doi.org/10.1016/j.annepidem.2014.11.021
1047-2797/

Propensity score matching (PSM) is a useful approach for dealing with these issues. A propensity score is defined as the conditional probability of being exposed to a condition [21,22]. The propensity score reduces the dimensionality of a large set of potential confounders to unity, making it conducive to simple pair matching [23]. After exposure groups are matched by propensity scores, they have been balanced on all available relevant covariates. In this way, we reduce the observable bias while maintaining the support of the data.

This study aimed to examine the association between neighborhood deprivation and PTB based on all live births in 2008 to 2009 in South Carolina, stratified by racial groups (whites and blacks). PSM was used to avoid any sparse data among covariate categories caused by an imbalanced distribution of data across quartiles of neighborhood deprivation.

Methods

Study population

Birth certificates of all live births in South Carolina from January 1, 2008, to December 31, 2009, were obtained from the South Carolina Department of Health and Environmental Control. Within the period, there were 123,759 live births. After excluding births without census tract information, multiple births, and births in Hispanic and other racial/ethnic groups (American Indian and Alaska Native, Asian Native Hawaiian, and other Pacific Islander), and extreme outliers of birth weight (±3 SD) and gestational age (<20 weeks), 98,456 white and black births were included in the study.

Measures

PTB was defined as gestational age less than 37 weeks. The PTB data were from the obstetric estimate of gestation from the birth certificate. This measure was added from 2003 in birth certificate and it was based on birth attendant's final estimate of gestation. The algorithm published by Messer et al. [16] was used to create the NDI for each census tract in South Carolina using principal component analysis. The eight census tract-level sociodemographic factors used to compute the NDI included percentage population with less than high school, percentage unemployed population, percentage males in management occupations, percentage crowded housing, percentage households in poverty, percentage female head households with children, percentage households earning less than $30,000 per year, and percentage households on public assistance. We used data from the 2000 census. The NDI was predicted based on the loadings of the eight factors in the first principal component. In this study, only the first principal component had an eigenvalue more than 1, accounting for 61.08% of the total variance. The NDI was standardized to have a mean of 0 and SD of 1 by dividing the index by the square of the eigenvalue. Quartiles of NDI were coded as Q1 (least deprived), Q2, Q3, and Q4 (most deprived). Q1 was considered the reference group. Principal component analysis was conducted using the *pca* program in STATA (Version 10, StataCorp, College Station, TX).

In the PSM analysis, to achieve the best model fit to predict propensity scores, we included all appropriate covariates which were predictive of the exposure of interest and occurred before the outcome of interest. We included all the sociodemographic variables available in the data set, including maternal age, maternal education, the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) participation, and urbanicity. Other covariates included body mass index, maternal smoking, prenatal care, number of previous live births, number of previous PTBs, and maternal risk factors such as infection, chronic and gestational hypertension, and diabetes.

Statistical analysis

Population characteristics were summarized for the pooled sample and for samples in each NDI quartile. Q2, Q3, and Q4 were compared with Q1 based on *t* tests for continuous variables and χ^2 tests for categorical variables.

As shown in Figure 1, the distribution of NDI was imbalanced between white and black women, with approximately 50% of black women living in the most deprived areas. To avoid inference due to the imbalanced distribution of NDI, we used PSM to analyze the relationship between NDI and adverse birth outcomes stratified by race. We used logistic regression to estimate the predicted probability of a mother's exposure to neighborhood deprivation to create matched pairs comparing NDI quartiles. All appropriate covariates discussed previously were included in the models to achieve the best fit. Propensity scores were estimated for each mother and computed separately for whites and blacks. We then matched the mothers living in deprived areas (Q2, Q3, and Q4, separately) with those living in the reference area (Q1) with the same propensity score. The matching procedure was conducted using the *psmatch2* module in STATA. The mothers living in deprived areas were matched 1:1 with replacement (to yield the largest number of matched pairs, the matched observations were returned to the data pool and could be matched again if the propensity score could be matched) to mothers living in reference areas with the same predicted probability of exposure to neighborhood deprivation within a range of ±0.01. We yielded 100% matching between the deprived group and the reference group because of the large sample size. Balance tests were performed to assure the balance of the data after matching. We compared the means and percentage bias before and after matching, and percentage bias reduction, with a goal of a percentage bias reduction of less than 10% indicating sufficient balance. The percentage bias is the percentage difference of the sample means in the deprived and reference group as a percentage of the square root of the average of the sample variances [24]. The graph of propensity score overlap was generated by level of neighborhood deprivation for each racial group. Generally, the graph showed the distribution of propensity score between exposed (deprived groups such as Q2, Q3, and Q4) and unexposed (Q1) groups. Adequate overlap between the distributions showed that we should be able to find a suitable match for an exposed subject. After the subjects were matched, the differences in prevalence of PTB were computed between the matched deprived and

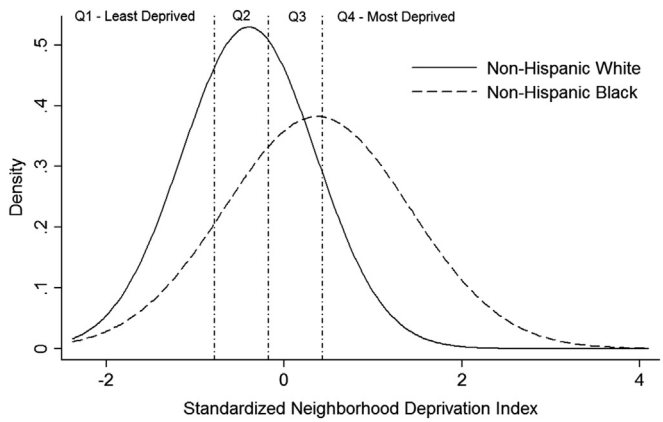

Fig. 1. Distribution of NDI by race in South Carolina.