Published in final edited form as:

Stat Med. 2015 February 20; 34(4): 595–612. doi:10.1002/sim.6357.

Extension of the Peters–Belson method to estimate health disparities among multiple groups using logistic regression with survey data

Y. Li^{a,*†}, B. I. Graubard^b, P. Huang^a, and J. L. Gastwirth^c

^aJoint Program in Survey Methodology, University of Maryland, College Park, MD 20742, U.S.A.

^bBiostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, U.S.A.

^cDepartment of Statistics, George Washington University, 2140 Pennsylvania Ave. NW, Washington, DC 20052 U.S.A.

Abstract

Determining the extent of a disparity, if any, between groups of people, for example, race or gender, is of interest in many fields, including public health for medical treatment and prevention of disease. An observed difference in the mean outcome between an advantaged group (AG) and disadvantaged group (DG) can be due to differences in the distribution of relevant covariates. The Peters–Belson (PB) method fits a regression model with covariates to the AG to predict, for each DG member, their outcome measure as if they had been from the AG. The difference between the mean predicted and the mean observed outcomes of DG members is the (unexplained) disparity of interest. We focus on applying the PB method to estimate the disparity based on binary/multinomial/proportional odds logistic regression models using data collected from complex surveys with more than one DG. Estimators of the unexplained disparity, an analytic variance–covariance estimator that is based on the Taylor linearization variance–covariance estimation method, as well as a Wald test for testing a joint null hypothesis of zero for unexplained disparities between two or more minority groups and a majority group, are provided. Simulation studies with data selected from simple random sampling and cluster sampling, as well as the analyses of disparity in body mass index in the National Health and Nutrition Examination Survey 1999–2004, are conducted. Empirical results indicate that the Taylor linearization variance–covariance estimation is accurate and that the proposed Wald test maintains the nominal level.

Keywords

Taylor linearization; complex survey data; multinomial logistic regression; proportional odds logistic regression; unexplained disparity

*Correspondence to: Y. Li, Joint Program in Survey Methodology, University of Maryland, College Park, MD 20742, U.S.A..

†yli6@umd.edu.

1. Introduction

Determining the extent of disparity between groups of people, for example, based on race, gender, or geography, is of great interest in many fields, including economics for disparity in wages and public health for disparity in the access to medical treatment and the prevalence of medical conditions [1, 2]. After observing a disparity in an outcome of interest, such as the prevalence of obesity measured on individuals from two or more groups (e.g., an advantaged/majority group such as White people and disadvantaged/minority groups such as Black and Hispanic people), a natural next step is to partition the observed disparity into a part that is due to a difference between the distributions of relevant covariates in the groups (e.g., age and education), called the explained disparity, and the remaining part of the observed disparity, which is called the unexplained disparity [3]. The unexplained disparity is often considered a measure of possible discrimination.

A method attributed to Peters [4] and Belson [5], referred in this paper as the Peters–Belson (PB) method, first fits a linear regression model with the relevant covariates as independent variables to the majority (White) group. The estimated model is then used to predict the outcome (such as wages) for each minority (Black) group member as if the Black had been White with the same covariates. The difference between the mean predicted outcomes and the mean observed outcomes of the Black group is the estimate of the unexplained disparity. The partition follows by using the remaining part of the observed disparity after subtracting out the unexplained disparity as the explained disparity. The PB method (also called the Blinder–Oaxaca method) has been used in race and gender wage disparity studies [6–8]. The PB method was extended from continuous to binary outcomes using logistic regression to estimate gender disparity or discrimination in job promotion in legal cases along with obtaining confidence intervals and hypothesis tests of discrimination [8, 9]. Nayak and Gastwirth [10] further showed how the PB method can be used in the context of generalized linear interactive models. Similar extensions to binary outcomes were later proposed to partition the observed disparity for a binary outcome using logistic and probit regression models [11,12]. Bauer and Sinning [13], built on these extensions to apply the PB approach to several other nonlinear models: proportional odds model (an ordinal logistic regression), Poisson regression model, tobit model, and a truncated multiple linear regression model. Fortin et al. [14] provided a comprehensive review of the decomposition of the overall (total) disparity between the group averages related to the PB and Blinder–Oaxaca methods.

National household health surveys are useful data sources for studying disparities for a variety of health outcome measures because these surveys collect a rich set of social, economic, demographic, and health variables. Many national health survey samples are designed to oversample individuals from lower social economic groups or certain race/ethnicity minority groups that provide improved precision when estimating disparities. In the last 15 years, collection of detailed self-reported race/ethnicity including multiple race reporting by US government agencies [15] has allowed US national health surveys such as the National Health Interview Survey (NHIS) 2000–2012 and the National Health and Nutrition Examination Survey (NHANES) to provide more detailed race/ethnicity group identification that is used for understanding health disparities among more specific groups.

A major purpose of this paper is to demonstrate the usefulness of these national health surveys to study health disparities for multiple race/ethnicity groups.

An important consideration when analyzing data from national household surveys such as NHIS or NHANES is to properly account for the complex sample designs of the surveys. These types of sample designs usually involve stratified multistage geographically based cluster sampling that are cost efficient for conducting household surveys but typically inflate variances of estimators compared to simple random samples with the same sample sizes. For example, one needs to properly weight the observations to reflect oversampling, nonresponse, and poststratification adjustments. Fairlie [12] used national economic survey data and conducted sample-weighted binary logistic regression analyses to estimate disparities between Black and White people in rates of self-employment based on the PB approach. Graubard et al. [3] and Fairlie [12] provide analytic variance estimators of the PB disparity under logistic regression that account for the sample weighting and complex sample design of survey data. (In addition, Graubard et al. [3] consider the multiple linear regression case and Fairlie [12] examine the probit regression case for complex survey data.) Statistical software written in STATA (StataCorp LP, College Station, Texas, USA) [16] provides several procedures for estimating the PB (Blinder–Oaxaca) decomposition using the various linear and nonlinear regression models discussed earlier that also incorporate the sample weighting in the estimation and use leaving-one-out jackknife and bootstrap methods to estimate variances of disparities [17–19].

This paper focuses on applying the PB method to binary/multinomial and proportional odds logistic regression models, which are commonly used in medical and public health analyses [20]. PB methods are extended for these models to the estimation and statistical hypothesis testing of disparity among more than two minority groups for data from complex samples of surveys. For example, this methodology allows researchers to take advantage of the detailed multiple race/ethnicity groups available in national health surveys. Analytical variance–covariance estimators are derived using Taylor linearization (i.e., a delta method) for PB estimators of disparity that will be used for inference. Such analytical variances are computationally more efficient than jackknife or bootstrap methods of variance estimation available in existing software [16]. In section 2, we developed the PB-based disparity estimators, variance estimation and statistical hypothesis testing procedures that are applicable to data collected in complex sample surveys. In section 3, simulation results are provided to study the finite sample properties of the methods and demonstrate that the tests maintain nominal Type I error rate. In section 4, we illustrate our methods with an analysis of disparity in the measured body mass index (BMI) among race/ethnicity groups interviewed and examined in the 1999–2004 NHANES, involving stratified multistage cluster sampling. We end with a discussion of our methods and results.

2. Methods

2.1. Notation and models

Suppose we have a survey of n sample individuals. Let R_0 denote the reference group of advantaged/majority individuals (e.g., White individuals) and R_k , for $k = 1, \dots, m$, be m disadvantaged/minority groups. Each sampled individual i in the survey is observed on (T

–1) indicator outcome variables $y_i^{(t)}$ (=1 if individual i has outcome value of t ; 0 otherwise) for $t = 1, \dots, (T-1)$ with T defined as the total number of categories of the outcome variable (e.g., $T = 3$ body weight categories that are defined in Section 4: underweight or normal weight, overweight and obese), a $p \times 1$ covariate vector x_i (e.g., age, smoking status, and income), and m indicator variables $\delta_{R_k i}$ (=1 if individual i is from group R_k ; 0 otherwise) for $k = 0, 1, \dots, m$. The (sample) weighted estimate of the observed proportion of individuals in group R_k with outcome in category t is

$$p_{R_k}^{(t)} = \frac{\sum_{j=1}^n w_j \delta_{R_k j} y_j^{(t)}}{\sum_{j=1}^n w_j \delta_{R_k j}}, \quad (1)$$

where w_j is the sample weight associated with the individual j , that is, the reciprocal of probability of including individual j in the sample [21].

We fit an appropriate binary, multinomial, or proportional odds logistic regression model to the observations from R_0 , depending on the type of categorical outcome. The model coefficients are estimated by maximizing the (sample) weighted pseudo-likelihood. The estimated model coefficients are then used to predict the probability for individual j with covariates x_j in a minority group to have an outcome value of t , denoted as $\hat{p}_{R_0 j}^{(t)}$. In other words, for individual j from a disadvantaged group (R_k , $k = 1, \dots, m$), $\hat{p}_{R_0 j}^{(t)}$ is the predicted probability of its outcome being in category t if this individual had been from the advantaged group R_0 . Accordingly, the sample-weighted mean of the predicted outcome probabilities in category t for individuals in R_k if these individuals had been from R_0 is

$$\hat{p}_{R_0 R_k}^{(t)} = \frac{\sum_{j=1}^n w_j \delta_{R_k j} \hat{p}_{R_0 j}^{(t)}}{\sum_{j=1}^n w_j \delta_{R_k j}}. \quad (2)$$

The difference in estimated means of y in category t for individuals in R_0 and R_k , $p_{R_0}^{(t)} - p_{R_k}^{(t)}$, which we call the ‘observed disparity’, can be partitioned as

$$p_{R_0}^{(t)} - p_{R_k}^{(t)} = \left(p_{R_0}^{(t)} - \hat{p}_{R_0 R_k}^{(t)} \right) + \left(\hat{p}_{R_0 R_k}^{(t)} - p_{R_k}^{(t)} \right). \quad (3)$$

The difference

$$U_{R_k}^{(t)} = \hat{p}_{R_0 R_k}^{(t)} - p_{R_k}^{(t)} \quad (4)$$

is called the ‘unexplained disparity’ (and the remaining difference in (3) is called the ‘explained disparity’). The quantities in (1), (2), and (4) are combined into the column

vectors $\hat{p}_{R_0 R_k} = \left(\hat{p}_{R_0 R_k}^{(1)}, \dots, \hat{p}_{R_0 R_k}^{(t)}, \dots, \hat{p}_{R_0 R_k}^{(T-1)} \right)'$, $p_{R_k} = \left(p_{R_k}^{(1)}, \dots, p_{R_k}^{(t)}, \dots, p_{R_k}^{(T-1)} \right)'$ and $U_{R_k} = \left(U_{R_k}^{(1)}, \dots, U_{R_k}^{(t)}, \dots, U_{R_k}^{(T-1)} \right)'$. The previous literature has considered estimation and inference for the unexplained disparity for only a single minority group R_1 . Here with m

minority groups and a T -category outcome variable, we can estimate the vector of unexplained disparities of length $m \times (T - 1)$ by

$$U = \begin{pmatrix} U_{R_1} \\ \vdots \\ U_{R_k} \\ \vdots \\ U_{R_m} \end{pmatrix} \quad \text{with} \quad U_{R_k} = \begin{pmatrix} U_{R_k}^{(1)} \\ \vdots \\ U_{R_k}^{(T-1)} \end{pmatrix} \quad (5)$$

and test $H_0: \mathcal{V} = 0$, where $\mathcal{V}$ is the unexplained disparity for the population across minority groups and the first $T-1$ outcomes. A test of H_0 can be based on the Wald statistic

$\frac{d-(T-1)+1}{d(T-1)} U' \text{cov}(U) U$, where d is the number of sampled primary units (PSUs), which are units (such as counties or contiguous counties in the USA used in the NHANES) sampled in the first stage in a multistage sample design, minus the number of sampling strata used to stratify the PSUs, and $\text{cov}(U)$ is a sample design consistent estimator for the covariance of U . An F -distribution with $T-1$ and $d - (T-1)+1$ degrees of freedom can be applied as the reference distribution to obtain a p -value for this test [21]. The Taylor linearization method or a replication method such as a leaving-one-out jackknife can be used to compute $\text{cov}(U)$

for a particular survey data set. Correlation between pairs of $U_{R_j}^{(t)}$ and $U_{R_k}^{(s)}$, $j \neq k$, results from the same vector of estimated regression coefficients used to obtain the mean of the predicted outcomes $\hat{p}_{R_0 R_j}^{(t)}$ and $\hat{p}_{R_0 R_k}^{(s)}$, which are in turn used to obtain the estimated unexplained disparities $U_{R_j}^{(t)}$ and $U_{R_k}^{(s)}$. Also, correlation between pairs of $U_{R_j}^{(t)}$ and $U_{R_k}^{(s)}$ can occur for multistage cluster samples when the first stage sampled clusters include sampled individuals from more than one group.

To estimate the unexplained disparity, $\mathcal{V}$, we first need to obtain the predicted probabilities

$\hat{p}_{R_0 i} = (\hat{p}_{R_0 i}^{(1)}, \dots, \hat{p}_{R_0 i}^{(t)}, \dots, \hat{p}_{R_0 i}^{(T-1)})'$ for each individual i in the minority group R_k . In the following, we study the disparity between m minority groups and a majority group using the PB method under multi-nomial logistic regression, that is, the outcome has $T - 2$ nominal categories, where $T = 2$ is binary logistic regression, and proportional odds logistic regression models, that is, the outcome has $T - 2$ ordinal categories.

Multinomial Logistic Regression Models

A multinomial logistic regression model for the observations from R_0 is given by

$$\log \frac{\Pr(y_i^{(t)} = 1 | x_i, i \in R_0)}{\Pr(y_i^{(T)} = 1 | x_i, i \in R_0)} = x_i' \beta_{R_0}^{(t)}$$

for $t = 1, \dots, T - 1$, where the x_i' is the transpose of x_i , and $\beta_{R_0}^{(t)}$ is a $p \times 1$ vector of regression coefficients; here $x_i = (x_{i1}, \dots, x_{ip})'$ with $x_{i1} \equiv 1$ corresponding to the intercepts. The (sample) weighted pseudo-(multinomial logistic regression) likelihood is maximized to the

observations from R_0 to obtain $\hat{\beta}_{R_0}^{(t)}$, a design consistent estimate of $\beta_{R_0}^{(t)}$. Using the fitted multinomial logistic regression, the predicted outcome for each individual i in group R_k for $k = 1, \dots, m$ is

$$\hat{p}_{R_0 i}^{(t)} = \exp \left(x_i' \hat{\beta}_{R_0}^{(t)} \right) / \left[1 + \sum_{s=1}^{T-1} \exp \left(x_i' \hat{\beta}_{R_0}^{(s)} \right) \right]. \quad (6)$$

Proportional odds logistic regression model

An extension of the multinomial logistic regression model for an ordered categorical outcome variable y_i is the proportional odds logistic regression model, given by

$$\log \frac{\Pr(y_i \leq t | x_i, i \in R_0)}{\Pr(y_i > t | x_i, i \in R_0)} = \alpha_{R_0}^{(t)} + x_i' \beta_{R_0}$$

for $t = 1, \dots, T-1$, where β_{R_0} is a $p \times 1$ vector of regression coefficients, and, for convenience, $\alpha_{R_0}^{(t)}$ denotes the intercepts. The proportional odds model assumes that the log odds for y_i being less than or equal to t versus greater than t is the same for all values of x_i , that is, the proportional odds assumption, which, when accurate, can increase statistical power over a multinomial model for testing the regression coefficients. The model is fitted to the observations from R_0 , and the (sample) weighted pseudo-likelihood is maximized to obtain $\hat{\alpha}_{R_0}^{(t)}$ and $\hat{\beta}_{R_0}$ that are sample design consistent estimates of $\alpha_{R_0}^{(t)}$ and β_{R_0} . Using the results of the proportional odds regression model, the predicted outcome for each individual i in group R_k for $k = 1, \dots, m$ is

$$\begin{aligned} \hat{p}_{R_0 i}^{(1)} &= \frac{\exp \left(\hat{\alpha}_{R_0}^{(1)} + x_i' \hat{\beta}_{R_0} \right)}{1 + \exp \left(\hat{\alpha}_{R_0}^{(1)} + x_i' \hat{\beta}_{R_0} \right)} \\ \hat{p}_{R_0 i}^{(t)} &= \frac{\exp \left(\hat{\alpha}_{R_0}^{(t)} + x_i' \hat{\beta}_{R_0} \right)}{1 + \exp \left(\hat{\alpha}_{R_0}^{(1)} + x_i' \hat{\beta}_{R_0} \right)} - \frac{\exp \left(\hat{\alpha}_{R_0}^{(t-1)} + x_i' \hat{\beta}_{R_0} \right)}{1 + \exp \left(\hat{\alpha}_{R_0}^{(t-1)} + x_i' \hat{\beta}_{R_0} \right)} \quad \text{for } t=1, \dots, T-1 \\ \hat{p}_{R_0 i}^{(T)} &= 1 - \frac{\exp \left(\hat{\alpha}_{R_0}^{(T-1)} + x_i' \hat{\beta}_{R_0} \right)}{1 + \exp \left(\hat{\alpha}_{R_0}^{(T-1)} + x_i' \hat{\beta}_{R_0} \right)} = \frac{1}{1 + \exp \left(\hat{\alpha}_{R_0}^{(T-1)} + x_i' \hat{\beta}_{R_0} \right)}. \end{aligned} \quad (7)$$

Depending on the regression model, the predicted outcome probabilities for sampled individual i , are given by either (6) or (7). The sample-weighted mean of these predicted probabilities for each category t for individuals in each group R_k is used to obtain the unexplained disparity estimate U given by (5). The proportional odds assumption can be tested using the method of Peterson and Harrell [22] that has been adapted for complex survey data in Proc Surveylogist in SAS [23].

2.2. Variance estimation of the PB measure of unexplained disparity U

The Taylor linearization method is used for variance estimation of the PB measure of disparity U . In general, this variance estimator involves first computing the Taylor deviate of

the estimator for each observation (individual), which is a measure of the influence or change in the value of the statistic when that observation is deleted. The details follow.

$$Cov(U) = \left\{ Cov(U_{R_k}, U_{R_l}) \right\}_{(T-1)m \times (T-1)m} \quad \text{for any } k, l = 1, 2, \dots, m,$$

where

$$Cov(U_{R_k}, U_{R_l}) = Cov \left(\frac{\sum_{j=1}^n w_j \delta_{R_{kj}} (\hat{p}_{R_{0j}} - y_j)}{\sum_{j=1}^n w_j \delta_{R_{kj}}}, \frac{\sum_{j=1}^n w_j \delta_{R_{lj}} (\hat{p}_{R_{0j}} - y_j)}{\sum_{j=1}^n w_j \delta_{R_{lj}}} \right)$$

The Taylor deviate for sampled individual i in minority group k for U_{R_k} can be derived by differentiating a sample-weighted estimator with respect to its weights [24], which gives

$$\begin{aligned} z_{R_{ki}} &= \frac{\partial}{\partial w_i} \left(\frac{\sum_{j=1}^n w_j \delta_{R_{kj}} (\hat{p}_{R_{0j}} - y_j)}{\sum_{j=1}^n w_j \delta_{R_{kj}}} \right) \\ &= \frac{(\delta_{R_{ki}} (\hat{p}_{R_{0i}} - y_i) + \sum_{j=1}^n w_j \delta_{R_{kj}} \frac{\partial}{\partial w_i} (\hat{p}_{R_{0j}})) (\sum_{j=1}^n w_j \delta_{R_{kj}}) - (\sum_{j=1}^n w_j \delta_{R_{kj}} (\hat{p}_{R_{0j}} - y_j)) \delta_{R_{ki}}}{(\sum_{j=1}^n w_j \delta_{R_{kj}})^2} \end{aligned}$$

where

$$\frac{\partial}{\partial w_i} (\hat{p}_{R_{0j}}) = \frac{\partial}{\partial \hat{\beta}_{R_0}} (\hat{p}_{R_{0j}}) \frac{\partial}{\partial w_i} (\hat{\beta}_{R_0}) \quad (8)$$

is derived from the weighted pseudo-likelihood estimating equations for β_{R_0} evaluated at $\hat{\beta}_{R_0}$, that is,

$$S(\hat{\beta}_{R_0}) = \sum_{j=1}^n w_j \delta_{R_{0j}} \hat{D}_j' (diag(\hat{p}_{R_{0j}}) - \hat{p}_{R_{0j}} \hat{p}_{R_{0j}}')^{-1} (y_j - \hat{p}_{R_{0j}}) = 0,$$

where $\hat{D}_j = \frac{\partial}{\partial \hat{\beta}_{R_0}} (\hat{p}_{R_{0j}})$, the matrix for partial derivatives of $p_{R_{0j}}$, with respect to β_{R_0} ,

evaluated at $\hat{\beta}_{R_0}$. See Appendix A for details for the derivation for $\frac{\partial}{\partial w_j} (\hat{p}_{R_{0j}})$ under the multinomial and proportional odds logistic models.

For multistage stratified cluster sampling used in household surveys such as the NHANES, the target population of individuals is partitioned into PSUs, which are usually geographically based clusters consisting of single counties, contiguous counties, cities, or parts of cities. The PSUs are grouped into H strata that are formed to be approximately homogeneous with respect to certain characteristics of the populations of the PSUs, for example, the population sizes or demographic characteristics. At the first stage of sampling

t_h , PSUs are randomly sampled from each stratum $h = 1, \dots, H$. At the second and further stages, stratification and cluster sampling can be used to sample individuals within the first-stage sampled PSUs. Let n_{hi} be the number of individuals ultimately sampled from PSU i in stratum h . For the NHANES, Hispanic and Black populations are oversampled to increase the sample sizes of individuals from these groups [25]. Often the proportion of PSUs sampled at the first stage from each of the strata is small so that the sample of the PSUs can be approximated as a stratified with-replacement sample [26, p. 246]. In this case, a design consistent estimator for the covariance of U is given by

$$C\hat{\sigma}_v(U) = \sum_{h=1}^H \frac{t_h}{t_h - 1} \sum_{i=1}^{t_h} \left(z_{hi} - \bar{z}_h \right) \left(z_{hi} - \bar{z}_h \right)', \quad (9)$$

where z_{hi} is the vector of sample-weighted totals of the observation-level Taylor deviates for

$$z_{hi} = \begin{pmatrix} \sum_{j=1}^{n_{hi}} w_{hij} z_{R_1} h_{ij} \\ \sum_{j=1}^{n_{hi}} w_{hij} z_{R_2} h_{ij} \\ \vdots \\ \sum_{j=1}^{n_{hi}} w_{hij} z_{R_m} h_{ij} \end{pmatrix}$$

PSU i sampled from stratum h , that is, sample weights w_{hij} for $j = 1, \dots, n_{hi}$.

For non-cluster sample designs, each sampled individual is its own sampled PSU so that $C\hat{\sigma}_v(U)$ in (9) simplifies with $n_{hi} = 1$ and t_h the number of sampled individuals in stratum h .

2.3. Using scores for the outcome categories in multinomial and proportional odds logistic models

For an ordinal outcome variable y , a score value, $s^{(t)}$, can be assigned to each category, $t = 1, \dots, T$ [13]. Accordingly, for each individual j , the expectation of y_j can be estimated by

$$\hat{E}(y_j) = \left(s^{(1)}, \dots, s^{(t)}, \dots, s^{(T)} \right) \left(\hat{p}_{R_{0j}}^{(1)}, \dots, \hat{p}_{R_{0j}}^{(t)}, \dots, \hat{p}_{R_{0j}}^{(T)} \right)'.$$

The PB estimator can then be conveniently computed as

$$U = \begin{pmatrix} U_{R_1} \\ \vdots \\ U_{R_k} \\ \vdots \\ U_{R_m} \end{pmatrix} \quad \text{with} \quad U_{R_k} = \frac{\sum_{j=1}^n w_j \delta_{R_k j} \left(\hat{E}(y_j) - s_j \right)}{\sum_{j=1}^n w_j \delta_{R_k j}},$$

where s_j is the observed score for individual j , that is, if $y_j^{(t)} = 1$ then $s_j = s^{(t)}$. Note that the length of the vector U is reduced from $m(T-1)$ in equation (5) to m . Similar to the previous section, a Taylor linearization variance–covariance estimator of U is derived by taking the derivative of U with respect to the weight w_i to obtain Taylor deviate for observation i , where

$$\frac{\partial}{\partial w_i} \hat{E}(y_j) = \frac{\partial}{\partial \hat{p}_{R_{0j}}} \hat{E}(y_j) \frac{\partial}{\partial w_i} (\hat{p}_{R_{0j}}),$$

where $\frac{\partial}{\partial \hat{p}_{R_{0j}}} \hat{E}(y_j) = (s^{(1)} \dots s^{(t)} \dots s^{(T)})$ and $\frac{\partial}{\partial w_i} (\hat{p}_{R_{0j}})$ is derived according to (9) in section 2.2.

3. Simulations

Three limited simulation studies were conducted to evaluate the finite-sample performance of the Taylor linearization estimation of the $Cov(U)$ and the Type I error of the proposed hypothesis tests of $H_0: \mathcal{V}=0$, corresponding to the following cases: (1) a majority group (White people), 2 minority groups (Black and Hispanic people) and a binary outcome, (2) a majority group (White people), a minority group (Black people) and a multinomial outcome with 3 categories, and (3) a majority group (White people), a minority group (Black people), and an ordinal outcome with 4 categories, respectively. For each simulation study, a sample size of $n=2,000$ individuals was repeatedly selected 3,000 times from a finite population of size 200,000.

In the first simulation, membership in the three race groups was generated from a multinomial distribution with probabilities of 0.6, 0.2, and 0.2, corresponding to ‘Whites’, ‘Blacks’, and ‘Hispanics’, respectively. We generated a binary outcome variable Y based on the binary logistic regression model

$$\text{logit}Pr(Y=1|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

where X_1 is a binary variable distributed as binomial with a probability of 0.5, the covariates X_2 and X_3 follow standard normal distributions. The values of β_0 , β_1 , β_2 , and β_3 are specified to be -1 , 0.5 , 1 , and 1 , respectively. We are testing the null hypothesis of no unexplained disparity among the White, Black and Hispanic people, that is, $H_0: \mathcal{V}=0$ where $\mathcal{V}$ is by

$U = \begin{pmatrix} \hat{p}_{R_{0,Black}} - p_{Black} \\ \hat{p}_{R_{0,Hispansic}} - p_{Hispansic} \end{pmatrix}$. Because the data is generated under the null hypothesis, the Type I error should be approximately at the nominal level of 5%.

In the second simulation, membership in the two race groups (White people vs. Black people) was generated following a binomial distribution with probability of 0.6 for being White and 0.4 for Black. A 3-category outcome variable Y was generated based on the multinomial logistic regression model

$$\log \frac{Pr(Y^{(t)}=1|X)}{Pr(Y^{(3)}=1|X)} = \beta_{t0} + \beta_{t1} X_1 + \beta_{t2} X_2 + \beta_{t3} X_3$$

for $t=1$ and 2 , corresponding to each of the first 2 outcome categories and the third outcome category $T=3$ is the reference category. We specify $\beta_{10}=1$, $\beta_{20}=0$, and $\beta_{11}=\beta_{12}=\beta_{13}=0.5$

for $t=1$ and 2 for both race groups. For White people, X_1 is distributed as a binomial distribution with probability 0.4, and X_2 and X_3 follow standard normal distributions, denoted by $N(0,1)$. For Black people, X_1 is distributed as a binomial distribution with probability 0.6, $X_2 \sim N(1.5, 1)$ and $X_3 \sim N(2, 1)$. Because we are generating data under a null hypothesis of no unexplained disparity between White people and Black people across the

three outcome groups, that is, $H_0: \mathcal{V}=0$ where $\mathcal{V}$ is estimated by $\begin{pmatrix} \hat{p}_{R_0, Black}^{(t=1)} - p_{Black}^{(t=1)} \\ \hat{p}_{R_0, Black}^{(t=2)} - p_{Black}^{(t=2)} \end{pmatrix}$ the Type error rate should be approximately at the 5% nominal level. We also assigned equally spaced score values of 2, 3, and 4 to each of the three outcome categories $t=1, 2$, and 3,

respectively, and tested $H_0: \mathcal{V}=0$ where $\mathcal{V}$ is estimated by $(2, 3, 4) \begin{pmatrix} \hat{p}_{R_0, Black}^{(t=1)} - p_{Black}^{(t=1)} \\ \hat{p}_{R_0, Black}^{(t=2)} - p_{Black}^{(t=2)} \\ \hat{p}_{R_0, Black}^{(t=3)} - p_{Black}^{(t=3)} \end{pmatrix}$.

We also extended the simulation 2 by varying the intercepts β_{20} for Black people to have values of .4 and .8, respectively, while keeping the other parameter values unchanged with β_{20} for White people equal to 0, to study the power of the proposed tests.

In the third simulation, a 4-category ordinal outcome variable Y is generated based on the proportional odds logistic regression model

$$\log \frac{Pr(Y \leq t|X)}{Pr(Y > t|X)} = \alpha_t + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

for $t=1, 2$, and 3, corresponding to the each of the first three outcome categories and the fourth outcome group $T=4$ is the reference category. X_1 is a binary variable that follows the same distribution as in the second simulation. For both race groups (i.e. White and Black people), X_2 and X_3 follow standard normal distributions. We specify $\alpha_1=1$, $\alpha_2=0.2$, $\alpha_3=-1$, and $\beta_1=\beta_2=\beta_3=0.5$. Because we are generating the data under a null hypothesis of no unexplained disparity between White and Black people across the three outcome categories,

that is, $H_0: \mathcal{V}=0$ where $\mathcal{V}$ is estimated by $\begin{pmatrix} \hat{p}_{R_0, Black}^{(t=1)} - p_{Black}^{(t=1)} \\ \hat{p}_{R_0, Black}^{(t=2)} - p_{Black}^{(t=2)} \\ \hat{p}_{R_0, Black}^{(t=3)} - p_{Black}^{(t=3)} \end{pmatrix}$, the Type I error rate should be approximately at the 5% nominal level. We further assigned score values of 1, 2, 3, and 4 to each of the four ordinal outcome categories, and tested $H_0: \mathcal{V}=0$ where $\mathcal{V}$ is estimated by

$$(1, 2, 3, 4) \begin{pmatrix} \hat{p}_{R_0, Black}^{(t=1)} - p_{Black}^{(t=1)} \\ \hat{p}_{R_0, Black}^{(t=2)} - p_{Black}^{(t=2)} \\ \hat{p}_{R_0, Black}^{(t=3)} - p_{Black}^{(t=3)} \\ \hat{p}_{R_0, Black}^{(t=4)} - p_{Black}^{(t=4)} \end{pmatrix}.$$

Table I shows the results from the three simulation studies with simple random sampling (SRS), where the $n=2,000$ individuals are randomly selected from the population of $N=200,000$. Table II shows the results from the same three simulation studies, but with cluster

sampling. Specifically, the target population is sorted by the weighted mean of y scores, where the weights are the probabilities of the scores of $Y=y$. In total, 4,000 PSU's are formed in the population by sequentially grouping the adjacent 50 individuals. We randomly selected 40 PSU's from the 4,000 PSU's, resulting in a cluster sample where the y scores are correlated for individuals within PSU's. From Tables I and II, we make the following observations: First, estimates of unexplained disparities (U) are consistently close to zero (< 0.01). Second, the Taylor linearization estimates of the variance–covariance matrices are close to the empirical estimates computed across the 3,000 simulations, indicating that the Taylor linearization variance estimation performs well in estimating the true variance of U . Finally, the Type I error rates under the null hypothesis were near the 5% nominal level. Additional sets of score values were considered with similar results to those for simulations 2 and 3 shown in Tables I and II. We found that the power is 37% or 93.6% when β_{20} for Black people versus White people are (0.4 vs. 0) or (0.8 vs. 0) for the extended simulation 2 (results not shown).

4. Data Examples

The data for this application is taken from the 1999–2004 NHANES, which is a continuing series of 2-year survey cycles designed to assess the health and nutritional status in a representative sample of non-institutionalized adults and children in the USA. Stratified multistage cluster sampling is applied in 1999–2004 NHANES, where at the first stage of sampling 2 PSUs are sampled from each of 42 strata and three PSUs sampled from one stratum.

We estimated disparities in measured BMI, which is calculated as measured body weight in kilograms divided by the square of measured height in meters, among White, Black and Hispanic groups of women (above 20 years old and non-pregnant). We used the PB methods for binary, multinomial, and proportional odds logistic regression, where the majority group was the White people. The categorical outcomes are based on the World Health Organization (WHO) categories of BMI for BMI < 18.5 , $18.5 \leq \text{BMI} < 25.0$, $25.0 \leq \text{BMI} < 30.0$ and $\text{BMI} \geq 30.0$ that define underweight, normal weight, and obesity, respectively [27]. Because of the small sample size of the underweight individuals, we combine the WHO normal weight category with the underweight category and call the resulting category (BMI < 25) the normal weight category. The covariates considered in these analyses were age, age², poverty index ratio (PIR), number of drinks alcohol per week, smoking status (never, former, or current smoker), total amount of leisure-time physical activity, and health insurance coverage (yes or no) (see supplemental tables in Appendix B for the estimated regression coefficients for the covariates used in each of our data examples).

All estimates were weighted using the NHANES full sample Mobile Examination Center (MEC) examination (sample) weights that were rescaled according to the US National Center for Health Statistics guidelines by multiplying the 4-year MEC exam weights for persons sampled during 1999–2002 by two-thirds, and the 2-year MEC exam weights for persons sampled during 2003–2004 by one-third to approximately reflect the population size over the entire 1999–2004 period [25]. We computed variances for the estimates using the Taylor linearization approach. The MASS and nnet packages of R [28] were used to

compute the estimates, and the R programming language [29] was also used to compute the Taylor deviates and resulting variance estimates.

Table III shows the observed and predicted prevalence of obesity for each of the three race/ethnicity groups under the binary logistic regression model. The Hosmer–Lemeshow goodness-of-fit test for survey data, implemented using SUDAAN (RTI International, Cary, North Carolina, USA) with the procedure Proc Rlogist [30], with p -value of 0.13 does not reject the validity of the use of binary logistic regression model. The unexplained disparity between the White and the Black people, and the White and the Hispanic people was -15.75% and -0.99% , respectively, and jointly, these two disparities are significantly different from zero ($p < 0.0001$). This result is driven by the large disparity between the White and the Black people ($p < 0.0001$), whereas there is an insignificant ($p = 0.59$) disparity between the White and the Hispanic people. About 26% of the observed disparity between the White and the Black people could be explained by the covariates in the binary logistic regression model, which suggests that a sizable amount of the unexplained disparity could be due to societal or cultural factors as well as unmeasured covariates, such as, past diet.

Table IV shows the observed and predicted prevalence of overweight and obesity for the White and the Black women. The predicted values are obtained using the multinomial logistic regression model that is fitted to the data for the White people, where the normal weight category is the reference outcome category for the model. We apply the Hosmer–Lemeshow test for each categorical outcome (i.e., overweight and obesity), one at a time, compared to the normal weight category outcome, producing the p -values of 0.07 and 0.03, respectively, which indicate that the multinomial logistic models may not fit the data well. The unexplained disparity in the Black people was -2.48% and -15.73% for overweight and obesity, respectively. There was a highly significant ($p < 0.0001$) unexplained disparity across the overweight and obesity outcome categories between White and Black. This result was primarily due to the large unexplained disparity in obesity ($p < 0.0001$) whereas there is no significant ($p = 0.12$) unexplained disparity in overweight. About 26% of the difference in the observed disparity for obesity for the Black people could be explained by the covariates in the multinomial logistic regression model. These results for disparity in obesity for the Black people from the multinomial model are almost identical to the results for disparity in obesity for the Black people that were estimated from the binary logistic regression in Table III, that is, an unexplained disparity of -15.75% and 26% of the observed disparity explained by the covariates. This would be expected because the same covariates are used in both models and in such cases numerical results from multinomial models will agree very close with comparable binary logistic models. Two sets of score values are considered in the score analysis: (1) one set of score values of 1, 0.93, and 1.30 are the approximate adjusted relative hazard ratios of all-cause mortality for the normal, overweight, and obesity categories, respectively, [31] and (2) the other set of score values of 21.87, 26.79, and 33.27 are the sample-weighted medians of BMI for the normal, overweight and obesity categories, respectively. The results from the two score analyses showed a similar pattern, in terms of the percent of disparity explained by the covariates, as

the result for the disparity for obesity (29% and 21% vs. 26%). This was expected because the explained disparity for overweight was not significant.

Table V shows the observed and predicted prevalence of overweight and obesity for White and Black women. The predicted values are obtained using the proportional odds logistic regression model fitted to the White women, where women with normal BMI are the reference outcome category. Although the proportionality assumption was tested and found violated with $p < 0.0001$ (implemented by the SAS procedure Proc Surveylogistic), for illustrative purposes we report the results under this fitted proportional odds logistic regression model. The unexplained disparity in the Black women was 0.41% and -17.53% for overweight and obesity, respectively. The hypothesis of no unexplained disparity jointly across the overweight and obesity outcome categories was rejected ($p < 0.0001$). This result was primarily due to the large unexplained disparity in obesity ($p < 0.0001$), whereas there is no significant ($p = 0.80$) unexplained disparity in overweight for Black women. About 18% of the difference in the observed disparity for obesity for the Black women could be explained by the covariates. Unlike table IV, the difference between the predicted value and the observed value for the Black women who are overweight becomes slightly positive, although the difference is insignificant. Results from the score analyses showed a similar pattern as those in table II.

The results of PB analysis for continuous BMI using multiple linear regression provided in Appendix C are similar to the results earlier for the logistic regression analyses.

5. Discussion

In this paper, the PB method is extended to estimate the unexplained disparity between multiple minority/disadvantaged groups and a majority/advantaged group under binary, multinomial, and proportional odds logistic regression models. We provide estimators of the unexplained disparity, an analytic variance–covariance estimator that is based on the Taylor linearization variance–covariance estimation method and a Wald test for testing a joint null hypothesis of zero unexplained disparities between two or more minority groups and a majority group. Our method applies to either simple random samples or complex samples that involve stratification and clustering. Limited simulation results indicate that the Taylor linearization variance–covariance estimation is accurate, and that the proposed Wald test maintains the nominal level. The analytical variance and covariance estimators that are presented in this paper require computation of derivatives from the weighted estimating equations. More computer intensive replication methods that do not require differentiation, such as leaving-one-out jackknife and bootstrap methods, can also be used to obtain variance–covariance estimates for complex estimators such as the unexplained disparity estimates [32]. Another possible replication approach is a parametric bootstrap as described by Mandel [33], but this approach should be investigated for complex sample designs.

In cases where the proportional odds is violated for a subvector of the covariates z_i and not violated for the remaining covariates v_i , then a partial proportional odds model can be applied, given by

$$\log \frac{Pr(y_i \leq t | v_i, z_i, i \in R_0)}{Pr(y_i > t | v_i, z_i, i \in R_0)} = \alpha_{R_0}^{(t)} + v_i' \eta_{R_0} + z_i' \gamma_{R_0}^{(t)}$$

for $t = 1, \dots, T - 1$, where η_{R_0} and $\gamma_{R_0}^{(t)}$ are $s \times 1$ and $r \times 1$ vectors, $p = s + r$, of regression coefficients and $\alpha_{R_0}^{(t)}$ denote the intercepts [20]. As indicated by the notation, the partial proportional odds model is fitted to the observations from R_0 , and the (sample) weighted pseudo-likelihood corresponding to the partial proportional odds model is maximized to obtain the sample design consistent estimators $\hat{\alpha}_{R_0}^{(t)}$, $\hat{\eta}_{R_0}$ and $\hat{\gamma}_{R_0}^{(t)}$. Using the results of the partial proportional odds regression model, the predicted probability of an outcome for each individual i in each group R_k for $k = 1, \dots, m$ is

$$\begin{aligned} \hat{p}_{R_0 i}^{(1)} &= \frac{\exp(\hat{\alpha}_{R_0}^{(1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(1)})}{1 + \exp(\hat{\alpha}_{R_0}^{(1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(1)})} \\ \hat{p}_{R_0 i}^{(t)} &= \frac{\exp(\hat{\alpha}_{R_0}^{(t)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(t)})}{1 + \exp(\hat{\alpha}_{R_0}^{(1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(t)})} \\ &\quad - \frac{\exp(\hat{\alpha}_{R_0}^{(t-1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(t-1)})}{1 + \exp(\hat{\alpha}_{R_0}^{(t-1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(t-1)})} \quad \text{for } t=2, \dots, T-1 \\ \hat{p}_{R_0 i}^{(T)} &= 1 - \frac{\exp(\hat{\alpha}_{R_0}^{(T-1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(T-1)})}{1 + \exp(\hat{\alpha}_{R_0}^{(T-1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(T-1)})} \\ &= \frac{1}{1 + \exp(\hat{\alpha}_{R_0}^{(T-1)} + v_i' \hat{\eta}_{R_0} + z_i' \hat{\gamma}_{R_0}^{(T-1)})}. \end{aligned}$$

The sample-weighted mean of these predicted probabilities for category t for individuals in each group R_k is used to obtain the PB unexplained disparity U given by (5). The variance and the tests can be obtained by following the same derivation as described in the methods section.

The validity of the PB method depends on how well the regression model fits the majority group. Various approaches with weaker parametric assumptions have been used when it is suspected that a regression model does not fit the data well, such as, using local linear regression [34] or semi-parametric propensity methods [35]. However, these methods have limited ability to adjust for more than a few covariates unless there are very large samples. As in all applications, one needs to assess how well the data fit the model. The validity of the PB methods, however, only needs to fit the majority group as the objective is to determine whether the minority group is treated similarly. Another important consideration is that the PB method, like all regression techniques, including propensity methods, can lead to biased results if there is insufficient overlap in the covariate distributions between the majority group and the minority groups. In our applications, the covariates, insurance, and smoking status, were categorical with at least 167 individuals in each category within each race/ethnicity group, which should be of sufficient sample size. Moreover, there was a substantial overlap in the distribution of continuous covariates of age, PIR, number of drinks

alcohol per week, and total amount of leisure-time physical activity between minority and majority groups. Finally, another important consideration when using regression methods is whether there are important omitted variables. Not including such important variables can result in either underestimation or overestimation of the unexplained disparity, depending on the differences between their distributions in the majority and minority groups and the sign of the regression coefficients of the missing variables. Careful consideration of the substantive relationships needs to be used when choosing covariates to be included in any regression model.

Appendix A: Derivatives of p^{R0j} with respect to sample weight of each individual under multinomial and proportional odds logistic models, used to obtain the Taylor deviates for the Taylor linearization variance estimators

The weighted pseudo-likelihood estimating equations for β_{R0} evaluated at $\hat{\beta}_{R0}$ is

$$S(\hat{\beta}_{R0}) = \sum_{j=1}^n w_j \delta_{R0j} \hat{D}_j' \left(\text{diag}(\hat{p}_{R0j}) - \hat{p}_{R0j} \hat{p}_{R0j}' \right)^{-1} (y_j - \hat{p}_{R0j}) = 0,$$

where $\hat{D}_j = \frac{\partial}{\partial \hat{\beta}_{R0}} (\hat{p}_{R0j})$. Define

$$Q(\hat{\beta}_{R0}) = -\frac{\partial}{\partial \hat{\beta}_{R0}} S(\hat{\beta}_{R0}) = \sum_{j=1}^n w_j \delta_{R0j} \hat{D}_j' \left(\text{diag}(\hat{p}_{R0j}) - \hat{p}_{R0j} \hat{p}_{R0j}' \right)^{-1} \hat{D}_j.$$

Taking the derivative of $S(\hat{\beta}_{R0})$ with respect to $w_{j'}$, and solving $\frac{\partial}{\partial w_{j'}} S(\hat{\beta}_{R0}) = 0$ for

$\frac{\partial}{\partial w_{j'}} (\partial \hat{\beta}_{R0})$ we have

$$\frac{\partial}{\partial w_{j'}} (\partial \hat{\beta}_{R0}) = Q^{-1}(\hat{\beta}_{R0}) \left\{ \delta_{R0j'} \hat{D}_{j'}' \left(\text{diag}(\hat{p}_{R0j'}) - \hat{p}_{R0j'} \hat{p}_{R0j'}' \right)^{-1} (y_{j'} - \hat{p}_{R0j'}) \right\}$$

Therefore,

$$\begin{aligned} \frac{\partial}{\partial w_{j'}} (\hat{p}_{R0j}) &= \frac{\partial}{\partial \hat{\beta}_{R0}} (\hat{p}_{R0j}) \frac{\partial}{\partial w_{j'}} (\partial \hat{\beta}_{R0}) \\ &= \hat{D}_j Q^{-1}(\hat{\beta}_{R0}) \left\{ \delta_{R0j'} \hat{D}_{j'}' \left(\text{diag}(\hat{p}_{R0j'}) - \hat{p}_{R0j'} \hat{p}_{R0j'}' \right)^{-1} (y_{j'} - \hat{p}_{R0j'}) \right\} \end{aligned}$$

Under the multinomial logistic regression model, we derive

$\hat{D}_j = \left(\text{diag}(\hat{p}_{R0j}) - \hat{p}_{R0j} \hat{p}_{R0j}' \right) \otimes x_j'$; under proportional odds logistic regression model,

$$\hat{D}_j = \begin{pmatrix} \text{diag}(q_{R_{0j}}) V \\ \hat{q}'_{R_{0j}} V \otimes x_j' \end{pmatrix} \text{ where the } (T-1) \times (T-1) \text{ matrix } V = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & 0 & 1 & -1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

and $(T-1)$ -dimensional column vector

$$q_{R_{0j}} = (F_{R_{0j}}(1)(1 - F_{R_{0j}}(1)), \dots, F_{R_{0j}}(t)(1 - F_{R_{0j}}(t)), \dots, F_{R_{0j}}(T-1)(1 - F_{R_{0j}}(T-1)))',$$

and $F_{R_{0j}}(t) = \Pr(y_j \leq t | x_j, j \in R_0)$ for $t = 1, 2, \dots, (T-1)$; and under partial proportional odds

$$\hat{D}_j = \begin{pmatrix} \text{diag}(q_{R_{0j}}) V \\ \hat{q}'_{R_{0j}} V \otimes x_j' \\ \text{diag}(q_{R_{0j}}) \text{diag}(z_j') V \end{pmatrix} \text{ with } (T-1)\text{-dimensional column vector}$$

logistic regression model [22],

$$q_{R_{0j}} = (F_{R_{0j}}(1)(1 - F_{R_{0j}}(1)), \dots, F_{R_{0j}}(t)(1 - F_{R_{0j}}(t)), \dots, F_{R_{0j}}(T-1)(1 - F_{R_{0j}}(T-1)))',$$

where $F_{R_{0j}}(t) = \Pr(y_j \leq t | x_j, j \in R_0)$ for $t = 1, 2, \dots, (T-1)$

For the special case of binary logistic regression, $S(\hat{\beta}_{R_0}) = \sum_{j=1}^n w_j \delta_{R_{0j}} x_j (y_j - \hat{p}_{R_{0j}}) = 0$ and

$$\begin{aligned} \frac{\partial}{\partial w_{j'}} (\hat{p}_{R_{0j}}) &= \frac{\partial}{\partial \hat{\beta}_{R_0}} (\hat{p}_{R_{0j}}) \frac{\partial}{\partial w_{j'}} (\partial \hat{\beta}_{R_0}) \\ &= (\hat{p}_{R_{0j}}) (1 - \hat{p}_{R_{0j}}) x_j' \left[\sum_{j=1}^n w_j \delta_{R_{0j}} x_j x_j' (\hat{p}_{R_{0j}}) (1 - \hat{p}_{R_{0j}}) \right]^{-1} \delta_{R_{0j}'} x_{j'}' (y_{j'} - \hat{p}_{R_{0j}'}) \end{aligned}$$

Appendix B: Results from binary and multinomial logistic regression and proportional odds regression analysis of National Health and Nutrition Examination Survey data

Table B.1

Binary logistic regression analysis of obesity (=0 if body mass index (BMI)< 30; 1 if BMI ≥ 30) among White non-pregnant women (>20 years).

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr >ChiSq
Intercept	-2.643	0.428	38.187	<.0001
Age	0.119	0.017	50.103	<.0001
Age ²	-0.001	0.000	52.895	<.0001
PIR	-0.166	0.032	27.557	<.0001

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr >ChiSq
Alcohol	-0.075	0.019	16.356	<.0001
Former Smoker	-0.107	0.104	1.056	0.304
Current Smoker	-0.265	0.105	6.405	0.011
Physical Activity ($\times 10^{-3}$)	-0.066	0.025	7.077	0.008
Insurance	-0.037	0.145	0.063	0.801

PIR, poverty index ratio.

Table B.II

Multinomial logistic regression analysis of categorical body mass index (BMI) (=1 if BMI< 25; 2 if 25 BMI< 30; and 3 if BMI 30) among White non-pregnant women (> 20 years).

Parameter	Catg.BMI	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	3	-2.586	0.470	30.308	<.0001
Intercept	2	-1.821	0.499	13.303	0.000
Age	3	0.138	0.019	53.118	<.0001
Age	2	0.058	0.021	7.849	0.005
Age ² ($\times 10^{-3}$)	3	-1.270	0.178	52.479	<.0001
Age ² ($\times 10^{-3}$)	2	0.400	0.195	4.150	0.042
PIR	3	-0.198	0.034	34.125	<.0001
PIR	2	-0.075	0.035	4.699	0.030
Alcohol	3	-0.083	0.020	17.502	<.0001
Alcohol	2	-0.018	0.014	1.616	0.204
Former Smoker	3	-0.077	0.127	0.368	0.544
Former Smoker	2	0.064	0.147	0.191	0.662
Current Smoker	3	-0.284	0.103	7.525	0.006
Current Smoker	2	-0.047	0.172	0.076	0.783
Physical Activity ($\times 10^{-3}$)	3	0.069	0.026	6.977	0.008
Physical Activity ($\times 10^{-3}$)	2	-0.009	0.007	1.799	0.180
Insurance	3	-0.185	0.155	1.430	0.232
Insurance	2	-0.403	0.198	4.126	0.042

PIR, poverty index ratio.

Table B.III

Proportional odds regression analysis of categorical body mass index (BMI) (=1 if BMI< 25; 2 if 25 BMI< 30; and 3 if BMI 30) among White non-pregnant women (> 20 years)¹.

Parameter	Estimate	Standard error	Wald chi-square	Pr >ChiSq
Intercept 3	-2.860	0.382	55.983	<.0001
Intercept 2	-1.636	0.377	18.824	<.0001
Age	0.114	0.015	56.188	<.0001
Age ²	-0.001	0.000	55.332	<.0001
PIR	-0.157	0.027	34.853	<.0001
Alcohol	-0.052	0.013	15.154	<.0001

Parameter	Estimate	Standard error	Wald chi-square	Pr >ChiSq
Former smoker	-0.042	0.098	0.185	0.667
Current smoker	-0.193	0.085	5.175	0.023
Physical activity ($\times 10^{-3}$)	-0.042	0.012	13.244	0.000
Insurance	-0.188	0.122	2.360	0.125

¹ Quasi-score test for the proportional odds assumption: p-value < 0.0001

Appendix C: Peters–Belson analyses of body mass index using multiple linear regression of National Health and Nutrition Examination Survey data

We use multiple linear regression to estimate the unexplained disparity by treating body mass index (BMI) as the continuous dependent variable. The estimated coefficients for the advantaged group $\hat{\beta}_{R_0}$ are obtained using weighted linear regression, weighting by the sample weights [21, pp. 92–93]. Using the same formulation as in this paper, the unexplained disparity is obtained as in (4) but replacing the $y_j^{(t)}$ for each individual j by the BMI value for individual j in (1) and $\hat{p}_{R_0j}^{(t)} = x_j' \hat{\beta}_{R_0}$ in (2), where the index is no longer needed and is dropped. The variance and covariance estimators for stratified multistage cluster samples are obtained by using the expression for equation (9) with the Taylor deviates modified according to [21]. Below the $\hat{\beta}_{R_0}$ for the multiple linear regression among the advantage group members is given in Table C.I and the results of the PB analysis are given Table C.II.

Table C.I

Multiple linear regression analysis of body mass index among White non-pregnant women (> 20 years).

Parameter	Estimate	Standard error	t value	Pr > t
Intercept	21.150	1.240	17.063	<.0001
Age	0.401	0.050	7.536	<.0001
Age ²	-0.004	0.000	-7.642	<.0001
PIR	-0.579	0.096	-6.050	<.0001
Alcohol	-0.194	0.032	-6.050	<.0001
Former Smoker	0.295	0.320	0.932	0.358
Current Smoker	-0.955	0.324	-2.946	0.006
Physical Activity ($\times 10^{-3}$)	-0.119	0.029	-4.099	<.0001
Insurance	-0.631	0.420	-1.502	0.142

PIR, poverty index ratio.

Table C.II

Multiple linear regression analysis of disparity in body mass index (BMI) among non-pregnant women (> 20 years) by race/ethnicity.

Group	Sample size	Observed Mean BMI	Predicted mean	Unexplained Disparity (SE) ¹	Percent of disparity explained ²	F value	P value
White	2647	27.74	27.74				
Black	1013	31.56	28.4	-3.16	0.17	151.02	<.0001
Hispanic	1379	28.87	28.3	-0.58	0.49	3.37	0.07
Overall						89.08	<.0001

¹Unexplained disparity is (observed mean BMI (White) – predicted mean BMI (Black or Hispanic)).

²Percent disparity explained is (1 – unexplained disparity)/(observed mean BMI(White) – observed BMI (Black or Hispanic))*100.

References

- Altonji, JG.; Blank, RM. Race and Gender in the Labor Market.. In: Ashenfelter, O.; Card, D., editors. Handbook of Labor Economics. Vol. 3. Elsevier Science B. V.; New York: 1999. p. 3143-3259.
- Cutler, DM.; Lleras-Muney, A.; Vogl, T. Socioeconomic Status and Health: Dimensions and Mechanisms.. In: Glied, S.; Smith, P., editors. Oxford Handbook on Health Economics. Oxford University Press; Oxford, UK: 2011. p. 124-163.
- Graubard BI, Rao RS, Gastwirth JL. Using the Peters-Belson method to measure health care disparities from complex survey data. Statistics in Medicine. 2005; 24:2659–2668. [PubMed: 16118808]
- Peters CC. A method of matching groups for experiment with no loss of population. The Journal of Educational Research. 1941; 34:606–612.
- Belson WA. A technique for studying the effects of a television broadcast. The Royal Statistical Society Series C-Applied Statistics. 1956; 5:195–202.
- Blinder AA. Wage discrimination: reduced form structural estimates. Journal of Human Resources. 1973; 8:433–455.
- Oaxaca R. Male-female differentials in urban labor markets. International Economic Review. 1973; 14:693–709.
- Gastwirth JL, Greenhouse SW. Biostatistical concepts and methods in the legal setting. Statistics in Medicine. 1995; 14:1641–1653. [PubMed: 7481200]
- Sinclair MD, Pan Q. Using the Peters-Belson method in equal employment opportunity personnel evaluations. Law, Probability and Risk. 2009; 8:95–117.
- Nayak, TK.; Gastwirth, JL. The Peters-Belson Approach to Measures of Economic and Legal Discrimination.. In: Johnson, NL.; Balakrishnan, N., editors. In Advances in the Theory and Practice of Statistics: A volume in honor of Samuel Kotz. Wiley; New York: 1997. p. 587-601.
- Fairlie RW. The absence of the African-American owned business: an analysis of the dynamics of self-employment. Journal of Labor Economics. 1999; 17:80–108.
- Fairlie RW. An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. Journal of Economic and Social Measurement. 2005; 30:305–316.
- Bauer TK, Sinning M. An extension of the Blinder-Oaxaca decomposition to nonlinear models. Advances in Statistical Analysis. 2008; 92:197–206.
- Fortin, N.; Lemieux, T.; Firpo, S. Decomposition methods in economics, Technical Report Working paper 16045. National Bureau of Economic Research; Cambridge, MA.: (Available from: <http://www.nber.org/papers/w16045>), 2010. [11/03/2014]

15. Office of Management and Budget. [11/03/2014] Provisional guidance on the implementation of the 1997 standards for the collection of federal data on race and ethnicity. Dec 15. 2000 (Available from: http://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/re_guidance2000update.pdf).
16. StataCorp.. Stata: Release 11. Statistical Software. StataCorp LP.; College Station. TX: 2009.
17. Sinning M, Hahn M, Bauer TK. The Blinder-Oaxaca decomposition for nonlinear regression models. *The Stata Journal*. 2008; 8:480–492.
18. Jann, B. A Stata implementation of the Blinder-Oaxaca decomposition, Technical Report Working Paper No. 5. Swiss Federal Institute of Technology Zurich Sociology; 2008.
19. Buis ML. Direct and indirect effects in a logit model. *The Stata Journal*. 2010; 10:11–29. [PubMed: 22468140]
20. Hosmer, DW.; Lemeshow, S. *Applied Logistic Regression* 2nd ed. Wiley; New York, NY: 2000.
21. Korn, EL.; Graubard, BI. *Analysis of Health Surveys*. Wiley; New York, NY: 1999.
22. Peterson B, Harrell FE. Partial proportional odds models for ordinal response variables. *Applied Statistics*. 1990; 39:205–217.
23. SAS Institute Inc. SAS/STAT® User's Guide, Version 9. SAS Institute Inc.; Cary, NC: 2002-2012.
24. Shah BV. Comment on “linearization variance estimators for survey data” by A Demanti and JNK Rao. *Survey Methodology*. 2004; 30:29.
25. Johnson CL, Paulose-Ram R, Ogden CL, Carroll MD, Kruszan-Moran D, Dohrmann SM, Curtin LR. National health and nutrition examination survey, analytic guidelines, 1999-2010. National Center for Health Statistics. *Vital Health Statistics*. 2013; 2(161)
26. Lohr, SL. *Sampling Design and Analysis*. 2nd ed.. Brooks/Cole; Boston, MA: 2010.
27. World Health Organization. Report of a WHO expert committee. Vol. 854. World Health Organ Tech Rep Ser; 1995. Physical status: the use and interpretation of anthropometry.; p. 1-452.
28. Venables, WN.; Ripley, BD. *Modern Applied Statistics with S* 4th ed. Springer; NY: 2014. (Available from: <http://cran.r-project.org/web/packages/MASS/index.html>, <http://cran.r-project.org/web/packages/nnet/index.html>) [June 24]
29. Development Core Team, R. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria: 2010. (Available from: <http://www.R-project.org>) [June 24, 2014]
30. [Oct 06, 2014] (Available from: http://www.rti.org/sudaan/pdf_files/110Example/Logistic%20Example%201.pdf)
31. Flegal KM, Graubard BI, Williamson DF, Gail MH. Cause-specific excess deaths associated with underweight, overweight, and obesity. *JAMA*. 2007; 298(17):2028–2037. [PubMed: 17986696]
32. Rust KF, Rao JNK. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*. 1996; 5:283–310. [PubMed: 8931197]
33. Mandel M. Simulation-based confidence intervals for functions with complicated derivatives. *The American Statistician*. 2013; 67(2):76–81.
34. Hikawa H, Bura E, Gastwirth JL. Local linear logistic Peters-Belson regression and its application in employment discrimination cases. *Statistics and its Interface*. 2010; 3:125–144.
35. Barsky R, Bound J, Charles KK, Lupton JP. Accounting for the Black–White wealth gap: a nonparametric approach. *Journal of the American Statistical Association*. 2002; 97:663–673.

Table 1

Simulation results under simple random sampling based on 3000 replications for Wald test of the null hypothesis of no unexplained disparity, $H_0 : \mathcal{V} = 0$.

Numerator of Wald test	Logistic Regression model	Average $U(\times 10^{-4})$	Empirical est. VarCov of $U(\times 10^{-4})$	Average Taylor linearization estimate of VarCov of $U(\times 10^{-4})$	Type I Error
$\begin{pmatrix} \hat{p}_{R_0,Black} - p_{Black} \\ \hat{p}_{R_0,Hispanic} - p_{Hispanic} \end{pmatrix}$	Binary	$\begin{pmatrix} 16.96 \\ -28.63 \end{pmatrix}$	$\begin{pmatrix} 5.46 & 1.27 \\ 1.27 & 5.59 \end{pmatrix}$	$\begin{pmatrix} 5.75 & 1.42 \\ 1.42 & 5.64 \end{pmatrix}$	0.050
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \end{pmatrix}$	Multinomial	$\begin{pmatrix} 8.21 \\ -26.86 \end{pmatrix}$	$\begin{pmatrix} 1.29 & -1.11 \\ -1.11 & 9.76 \end{pmatrix}$	$\begin{pmatrix} 1.37 & -1.20 \\ -1.20 & 10.03 \end{pmatrix}$	0.048
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \end{pmatrix}$	Multinomial	10.44	10.48	10.70	0.057
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \end{pmatrix}$	Proportional odds	$\begin{pmatrix} 21.31 \\ 3.58 \\ 14.02 \end{pmatrix}$	$\begin{pmatrix} 4.55 & -2.08 & -1.11 \\ -2.08 & 3.98 & -0.84 \\ -1.11 & -0.84 & 2.62 \end{pmatrix}$	$\begin{pmatrix} 4.39 & -1.91 & -1.11 \\ -1.91 & 3.89 & -0.76 \\ -1.11 & -0.76 & 2.65 \end{pmatrix}$	0.048
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \\ \hat{p}_{R_0,Black}^{(i=4)} - p_{Black}^{(i=4)} \end{pmatrix}$	Proportional odds	42.75	24.46	25.13	0.049

Table II

Simulation results under cluster sampling based on 3000 replications of Wald test of the null hypothesis of no unexplained disparity, $H_0 : \psi=0$.

Numerator of Wald test	Logistic Regression model	Average $U(\times 10^{-4})$	Empirical est. VarCov of $U(\times 10^{-4})$	Average Taylor linearization estimate of VarCov of $U(\times 10^{-4})$	Type I Error
$\begin{pmatrix} \hat{p}_{R_0,Black} - p_{Black} \\ \hat{p}_{R_0,Hispanic} - p_{Hispanic} \end{pmatrix}$	Binary	$\begin{pmatrix} 25.82 \\ -21.40 \end{pmatrix}$	$\begin{pmatrix} 5.37 & 1.18 \\ 1.18 & 5.56 \end{pmatrix}$	$\begin{pmatrix} 5.61 & 1.44 \\ 1.44 & 5.68 \end{pmatrix}$	0.049
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \end{pmatrix}$ (1, 2, 3)	Multinomial	$\begin{pmatrix} 14.38 \\ -38.46 \end{pmatrix}$	$\begin{pmatrix} 1.47 & -1.25 \\ -1.25 & 9.91 \end{pmatrix}$	$\begin{pmatrix} 1.50 & -1.26 \\ -1.26 & 10.24 \end{pmatrix}$	0.060
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \end{pmatrix}$ (1, 2, 3)	Multinomial	9.69	10.82	11.20	0.058
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \end{pmatrix}$ (1, 2, 3, 4)	Proportional odds	$\begin{pmatrix} 29.08 \\ 31.85 \\ 8.15 \end{pmatrix}$	$\begin{pmatrix} 4.32 & -1.87 & -0.90 \\ -1.87 & 3.94 & -0.88 \\ -0.90 & -0.88 & 2.61 \end{pmatrix}$	$\begin{pmatrix} 4.25 & -1.80 & -1.04 \\ -1.80 & 3.93 & -0.86 \\ -1.04 & -0.86 & 2.71 \end{pmatrix}$	0.051
$\begin{pmatrix} \hat{p}_{R_0,Black}^{(i=1)} - p_{Black}^{(i=1)} \\ \hat{p}_{R_0,Black}^{(i=2)} - p_{Black}^{(i=2)} \\ \hat{p}_{R_0,Black}^{(i=3)} - p_{Black}^{(i=3)} \\ \hat{p}_{R_0,Black}^{(i=4)} - p_{Black}^{(i=4)} \end{pmatrix}$ (1, 2, 3, 4)	Proportional odds	-15.37	25.85	25.44	0.048

Table III

Binary logistic regression analysis of disparity in obesity among non-pregnant women (>20 years) by race groups.

Group	Sample size	Observed prevalence of obesity	Predicted prevalence of obesity	Unexplained disparity (SE) ¹	Percent of disparity explained ²	F value	P value
White	2647	30.78%	30.78%	0			
Black	1013	52.08%	36.33%	-15.75% (0.020)	26%	86.64	< 0.0001
Hispanic	1379	37.41%	36.41%	-0.99% (0.022)	85%	0.30	0.59
Overall						39.82	< 0.0001

¹Unexplained disparity is (Observed prevalence of obesity (White) – predicted prevalence of obesity (Black or Hispanic))

²Percent disparity explained is (1– unexplained disparity)/(observed prevalence of obesity(White) – observed prevalence of obesity (Black or Hispanic))*100

Table IV

Multinomial logistic regression analysis of disparity in overweight/obesity among non-pregnant women (>20 years) by race groups.

Group	Outcome	Sample size	Observed prevalence or expectation of outcome	Predicted prevalence or expectation of outcome ¹	Unexplained disparity (SE) ²	Percent of disparity explained ³	F value	P value
White	Overweight	761	27.67%	27.67%	0			
	Obesity	819	30.78%	30.78%	0			
Black	Overweight	289	27.55%	25.07%	-2.47% (0.019)	2043%	2.48	0.12
	Obesity	529	52.08%	36.35%	-15.73% (0.020)	26%	85.31	<.0001
Overall							50.45	<.0001
<u>Score of (normal, overweight, obesity) = (1, 0.93, 1.3)</u>								
White		2647	1.07	1.07	0			
Black		1013	1.14	1.09	-0.045 (0.0068)	29%	44.56	<.0001
<u>Score of (normal, overweight, obesity) = (21.87, 26.79, 33.27)</u>								
White		2647	26.74	26.74	0			
Black		1013	29.16	27.25	-1.92 (0.22)	21%	94.89	<.0001

¹ Expectation of the outcome using the predicted prevalence.

² Unexplained disparity is (observed prevalence or expectation of outcome (White) – predicted prevalence or expectation of outcome using predicted prevalence (Black)).

³ Percent disparity explained is (1 – unexplained disparity) / (observed prevalence or expectation of outcome (White) – observed prevalence or expectation of outcome using predicted prevalence (Black)) * 100.

Table V

Proportional odds logistic regression analysis of disparity in overweight/obesity among non-pregnant women (< 20 years) by race groups.

Group	Outcome	Sample size	Observed prevalence (%) or expectation of outcome	Predicted prevalence (%) or expectation of outcome ¹	Unexplained disparity (SE) ²	Percent of disparity explained ³	F value	P value
White	Overweight	761	27.67%	27.73%	0			
	Obesity	819	30.78%	30.85%	0			
Black	Overweight	289	27.55%	27.96%	0.41% (0.019)	-222%	0.065	0.80
	Obesity	529	52.08%	34.55%	-17.53% (0.020)	18%	105.34	<.0001
Overall							50.23	<.0001
<u>Score of (normal, overweight, obesity) = (1, 0.93, 1.3)</u>								
White		2647	1.07	1.07	0			
Black		1013	1.14	1.08	-0.053 (0.0067)	17%	63.25	<.0001
<u>Score of (normal, overweight, obesity) = (21.87, 26.79, 33.27)</u>								
White		2647	26.74	26.75	0			
Black		1013	29.16	27.18	-1.98 (0.20)	18%	102.09	<.0001

¹ Expectation of the outcome using the predicted prevalence.

² Unexplained disparity is (observed prevalence or expectation of outcome (White) – predicted prevalence or expectation of outcome using predicted prevalence (Black)).

³ Percent disparity explained is (1 – unexplained disparity) / (observed prevalence or expectation of outcome (White) – predicted prevalence or expectation of outcome using predicted prevalence (Black)) * 100.