4-17-2012

Adjusting for Confounding by Neighborhood Using a Proportional Odds Model and Complex Survey Data

Babette A. Brumback

Amy B. Dailey
Gettysburg College

Hao W. Zheng

Brumback, B. A., Dailey, A. B., & Zheng, H. W. (2012). Adjusting for Confounding by Neighborhood Using a Proportional Odds Model and Complex Survey Data. American Journal of Epidemiology, 175(11), 1133-1141. http://dx.doi.org/10.1093/aje/kwr452

Adjusting for Confounding by Neighborhood Using a Proportional Odds Model and Complex Survey Data

Abstract

In social epidemiology, an individual's **neighborhood** is considered to be an important determinant of health behaviors, mediators, and outcomes. Consequently, when investigating health disparities, researchers may wish to adjust for **confounding** by unmeasured **neighborhood** factors, such as local availability of health facilities or cultural predispositions. With a simple random sample and a binary outcome, a conditional logistic regression analysis that treats individuals within a **neighborhood** as a matched set is a natural method to use. The authors present a generalization of this method for ordinal outcomes and **complex** sampling designs. The method is based on a **proportional odds model** and is very simple to program **using** standard software such as SAS PROC SURVEYLOGISTIC (SAS Institute Inc., Cary, North Carolina). The authors applied the method to analyze racial/ethnic differences in dental preventative care, **using 2008** Florida Behavioral Risk Factor Surveillance System **survey data**. The ordinal outcome represented time since last dental cleaning, and the authors adjusted for individual-level **confounding** by gender, age, education, and health insurance coverage. The authors compared results with and without additional adjustment for **confounding** by **neighborhood**, operationalized as zip code. The authors found that adjustment for **confounding** by **neighborhood** greatly affected the results in this example.

Keywords

Neighborhoods, Dental Medicine, Models and Simulations, Socioeconomic Factors, health, epidemiology

Disciplines

Behavior and Behavior Mechanisms | Other Medicine and Health Sciences

American Journal of Epidemiology

Vol. 175, No. 11
DOI: 10.1093/aje/kwr452
Advance Access publication:
April 17, 2012

Practice of Epidemiology

Adjusting for Confounding by Neighborhood Using a Proportional Odds Model and Complex Survey Data

Babette A. Brumback*, Amy B. Dailey, and Hao W. Zheng

* Correspondence to Babette A. Brumback, Department of Biostatistics, College of Public Health and Health Professions and College of Medicine, University of Florida, P.O. Box 117450, Gainesville, FL 32610 (e-mail: brumback@ufl.edu).

Initially submitted May 2, 2011; accepted for publication November 9, 2011.

In social epidemiology, an individual's neighborhood is considered to be an important determinant of health behaviors, mediators, and outcomes. Consequently, when investigating health disparities, researchers may wish to adjust for confounding by unmeasured neighborhood factors, such as local availability of health facilities or cultural predispositions. With a simple random sample and a binary outcome, a conditional logistic regression analysis that treats individuals within a neighborhood as a matched set is a natural method to use. The authors present a generalization of this method for ordinal outcomes and complex sampling designs. The method is based on a proportional odds model and is very simple to program using standard software such as SAS PROC SURVEYLOGISTIC (SAS Institute Inc., Cary, North Carolina). The authors applied the method to analyze racial/ethnic differences in dental preventative care, using 2008 Florida Behavioral Risk Factor Surveillance System survey data. The ordinal outcome represented time since last dental cleaning, and the authors adjusted for individual-level confounding by gender, age, education, and health insurance coverage. The authors compared results with and without additional adjustment for confounding by neighborhood, operationalized as zip code. The authors found that adjustment for confounding by neighborhood greatly affected the results in this example.

confounding factors (epidemiology); healthcare disparities; health status disparities; health surveys; logistic models; minority health; proportional hazards models; residence characteristics

Abbreviations: BRFSS, Behavioral Risk Factor Surveillance System; CI, confidence interval.

The determinants of racial/ethnic and socioeconomic disparities in health have been a concern in the United States for decades. National initiatives, such as Healthy People (1), have explicitly included goals to reduce or eliminate disparities. Understanding the role of neighborhood in causing and perpetuating racial/ethnic and socioeconomic disparities in health has been a prominent area of research in recent years. Neighborhood has been conceptualized to influence disparities in health and health-related behaviors through a variety of mechanisms, including access to material and social resources (2), residential segregation (3), ethnic group density (4), characteristics of the built environment (5), and environmental exposures (6). Investigators may be interested in 1) quantifying overall neighborhood effects; 2) isolating a particular neighborhood attribute; 3) controlling for unmeasured neighborhood factors to identify residual disparities not accounted for by neighborhood factors; or 4) use of neighborhood as a proxy for socioeconomic conditions that have not been estimated appropriately at the individual level.

Standard multilevel models which include a random intercept for neighborhood can be used to jointly quantify neighborhood and individual effects when the sample sizes within neighborhoods are large and the within-neighborhood variation of individual-level covariates is sufficient. When these conditions are not satisfied, as is often the case, standard use of multilevel models fails to adjust for confounding of the individual effect by unmeasured neighborhood factors (7–10). Likewise, the unmeasured neighborhood effects are not estimated accurately. These failures are due to violation of the assumption that the random effects are uncorrelated with the measured covariates; this violation is implied when there is confounding of individual-level effects by the unmeasured neighborhood effects. One might attempt to solve this problem by modifying the multilevel model to include the neighborhood

averages of individual-level covariates as additional terms in the model (7–9, 11, 12). However, Brumback et al. (9) showed that for multilevel models using a logit or other nonidentity link function, this method requires additional strong assumptions in order to consistently adjust for confounding due to unmeasured neighborhood factors. Specifically, one must assume that the neighborhood effect is a linear function of the neighborhood averages plus a random term that is independent of the individual-level variables. When this assumption is violated, the method is prone to bias (9); whether this bias can be substantial is a topic for further research. Moreover, recent generalizations of generalized linear mixed models for complex survey data (13) have been demonstrated to produce inconsistent estimators of fixed-effects parameters with or without confounding by unmeasured neighborhood factors, when the cluster sizes are small and the sampling is informative (10, 13). To attempt to adjust for confounding by neighborhood with the generalized software (13), one could include the weighted neighborhood averages of the individual-level covariates in the model (10). The resulting estimators are inconsistent, however, because the estimating equation based on the derivative of the log pseudolikelihood is a nonlinear function of the individual-level sampling weights, and as such can be strongly biased for small clusters (10, 13).

With noncomplex survey data, no additional strong assumptions are necessary if one uses a conditional maximum likelihood approach—for example, conditional logistic regression with a binary outcome (14, 15). Very recently, conditional logistic regression has been generalized for use with complex survey data (16, 17). In the present article, we extend the method to accommodate ordinal outcomes via a proportional odds model, which reduces to a logistic regression model in the case of a binary outcome. It is noteworthy that even for simple random samples, regression methods based on conditional likelihood (14, 15) have not previously been developed for the proportional odds model (18, 19). Our development has the added advantage that it consistently estimates the proportional odds with complex survey data. Furthermore, our method is extremely easy to program in SAS (SAS Institute Inc., Cary, North Carolina) using PROC SURVEYLOGISTIC, or in similar statistical software packages. Statistical methods for epidemiologic analyses of complex survey data are of recent interest (20, 21), because of the easy availability of freely downloadable public-use data sets that are nationally representative, such as Behavioral Risk Factor Surveillance System (BRFSS) data or National Health Interview Survey data.

To illustrate the new method, we apply it to 2008 Florida BRFSS survey data to assess racial/ethnic disparities in oral health care. Racial/ethnic or socioeconomic disparities have been observed across a spectrum of oral health outcomes, including the presence of untreated dental caries (22, 23), other oral health problems (e.g., toothaches, tooth loss, or periodontal disease) (24–28), and self-rated or parent-rated dental health (26, 27, 29–31). Access to dental care, particularly preventative care such as routine dental cleanings, is an important determinant of oral health (32). According to a review of oral health disparities in the United States by Chattopadhyay (25), non-Hispanic African Americans and Hispanics are less likely to access dental care than whites or members of other racial/ethnic groups. Factors that have been shown to influence disparities in oral health or access to dental care include place of residence (e.g., urban or rural setting), education, income, insurance status, gender, and health behaviors such as nutrition and smoking (23–30, 33). For persons accessing dental services through Medicaid, low reimbursement rates, low provider participation, and shortages of dentists and clinics in poor areas may also play a role in access to preventive oral health care (34, 35). With our method, we can additionally adjust individual-level effects for unmeasured neighborhood factors, which may be serving as proxies for unmeasured individual-level socioeconomic factors or as determinants separate from socioeconomic status.

For the sake of comparison, we also apply the multilevel modeling approach, which includes the weighted neighborhood averages of individual-level covariates in the model (10). We let the individual-level weights be the BRFSS sampling weights, and the neighborhood-level weights equal 1. Because the generalization (13) of the multilevel modeling approach for complex survey data can lead to biased estimators (10, 13), we apply it both with and without the complex survey weights. When we apply it without the complex survey weights, we include the unweighted neighborhood averages of individual-level covariates in the model, rather than the weighted averages. We use the GLLAMM procedure in Stata, version 11.0 (StataCorp LP, College Station, Texas), for the computation.

BRFSS EXAMPLE

Our ordinal outcome represents how long it has been since an individual has had his or her teeth cleaned by a dentist or dental hygienist (within the past year, within the past 2 years, within the past 5 years, 5 or more years ago, or never). Our covariate of primary interest is race/ethnicity, which we categorized into non-Hispanic white, non-Hispanic African American, Hispanic, and other. We included the individual-level confounders gender, age (18–34, 35–54, 55–64, or $\geq$65 years), education (less than high school vs. high school or more), and health insurance status (covered vs. uncovered). We also included a neighborhood variable, operationalized as zip code. For ease of implementation, as we will discuss, we permitted the neighborhood effect to differ for portions of the zip code nested within different BRFSS survey strata. The Florida BRFSS uses a stratified sampling design with 134 strata formed by 2 telephone density strata crossed with 67 Florida counties. This led to 1,968 neighborhood effects in the model, with each neighborhood on average containing 4.6 sampled individuals; 39% contained 1 sampled individual, 15% contained 2 individuals, 11% contained 3 individuals, 8% contained 4 individuals, 5.5% contained 5 individuals, 12% contained 6–10 individuals, 4% contained 11–15 individuals, and 5% contained 16–83 individuals. Thus, the sample sizes in the neighborhoods were relatively small. In a second set of analyses, we also included annual household income ($<$\$15,000, \$15,000–$<$\$25,000, \$25,000–$<$\$50,000, or $\geq$\$50,000).

The Florida BRFSS uses disproportionate stratified sampling, in which only 1 person per household can be selected. In 2008, the BRFSS sampled 10,874 Floridians, including 9,745 Floridians with teeth. Each individual in the BRFSS is assigned a sampling weight, representing the inverse probability of being selected into the sample multiplied by a poststratification adjustment, constructed so the joint distribution of

race/ethnicity, gender, and age matches that of the most recent state census. For our analysis, we will need to estimate the inverse probability of selecting each possible pair of individuals within a given neighborhood. We are approximating this as the product of the 2 individual sampling weights. This approximation would be nearly exact if no poststratification adjustment had been made. However, if we assume that the inverse of the poststratification factor represents the conditional probability of responding to the survey given the survey design variables and race/ethnicity, gender, and age, and that the probability of one individual in a pair responding is independent of whether the other responded, then our approximation is valid. We also point out that the probability of pairs within the same household within a neighborhood being selected into the sample is zero. Strictly speaking, this violates the "positivity" (36) assumption, that is, that all pairs within a given neighborhood have a positive chance of selection into the sample. However, even if the BRFSS were to allow multiple individuals per household to be selected into the sample, such persons would represent a negligible fraction of the sample. Thus, for all practical purposes, the positivity assumption is satisfied, in that its violation in our context results in negligible bias.

We excluded persons with missing data or a "don't know" response to any of the questions we used in our analysis, except for the question on whether or not the individual had teeth. This resulted in a final sample of 8,989 individuals; that is, we excluded 7.8% of Floridians in the original sample who had teeth. In the second set of analyses including household income as a confounder, exclusion of missing data resulted in a final sample of 8,079 individuals, excluding 17.1% of the original sample who had teeth.

MODEL AND METHOD

We first define a proportional odds model for the population of M neighborhoods, with N_i individuals in the population residing in neighborhood i. Let Y_{ij} be an ordinal outcome with categories $k = 1,\ldots, K$, for individual $j, j = 1,\ldots, N_i$, in neighborhood $i, i = 1,\ldots, M$, and let $X_i = (X_{i1},\ldots, X_iN_i)$ be the vector of individual covariates X_{ij} from neighborhood i. Let $V_{kij} = 1$ if $Y_{ij} \leq k$, and 0 otherwise. The proportional odds model (15) can then be written as

$$\operatorname{logit}\big(E\big(V_{kij} \mid X_i, b_i\big)\big) = X_{ij}\beta + \alpha_\kappa + b_i,\ k = 1,\ldots,k-1. \quad (1)$$

The goal is to estimate β, the coefficients of the individual-level covariates. In our example, we are primarily interested in the coefficients corresponding to the race/ethnicity categorical variable; the other covariates are included as confounders. Unlike the simpler logistic model for binary outcomes, model 1 does not admit a proper conditional likelihood (15) when $K > 2$ (18); therefore, an analog of conditional logistic regression for the proportional odds model has not been previously developed (18, 19).

In the present paper, we develop conditional logistic regression for the proportional odds model using a conditional pseudolikelihood rather than a proper conditional likelihood, which extends previous methods for binary outcomes (16, 17, 37, 38). The conditional pseudolikelihood is constructed using all pairs (V_{kij}, V_{kil}) matched on neighborhood i and category k, and pretending that the matched pairs are independent even though they are not. Only pairs in which one observation is a case and the other is a control contribute any information to the pseudolikelihood, and the rest can be ignored. Specifically, we consider the conditional likelihood for the pair (V_{kij}, V_{kil}), matched on neighborhood i and category k, supposing that we had selected (ij, il) completely at random from all possible pairs matched on neighborhood i. The conditional likelihood, or probability, that $V_{kij} = 1$ and $V_{kil} = 0$ given that their sum equals 1 (i.e., that exactly 1 member of the pair is a case) is the basis for analysis of matched case-control studies, and it equals

$$\frac{\exp\big(\big(X_{ij} - X_{il}\big)\beta\big)}{1 + \exp\big(\big(X_{ij} - X_{il}\big)\beta\big)}. \quad (2)$$

Note that when $V_{kij} = V_{kil}$, the conditional likelihood is equal to 1 (and hence does not involve β), and also note that we can arbitrarily order the pair such that $V_{kij} = 1$ and $V_{kil} = 0$. The first derivative of the logarithm of equation 2, known as the score equation S_{kijl}, has an expected value equal to zero if we select the pair completely at random from all pairs matched on neighborhood and category. With complex survey data, let W_{ijl} be the inverse probability of selecting pair (V_{kij}, V_{kil}) into the sample; then the weighted score equation $W_{ijl}S_{kijl}$ has an expected value of zero, provided that each of the pairs matched on neighborhood and category has a nonzero probability of being selected into the sample. This latter condition is typically referred to as a positivity condition (36).

Therefore, the sum of the weighted score equations over all k and all pairs (ij, il) in the sample also has an expected value equal to zero, and we can estimate β with the $\hat{\beta}$ that solves $\hat{S} = \sum_{K_{(ij,il)}} W_{ijl}S_{kijl} = 0$, provided that the positivity condition holds for all such pairs in the sample. When the positivity condition is not exactly satisfied, as turned out to be the case with our BRFSS example (because only 1 member per household can be sampled), one can sometimes argue, as we did, that the bias incurred is negligible, because the expected proportion in the sample of excluded population pairs under an analogous sampling design that permits all pairs to be sampled is negligible.

The properties of the estimator $\hat{\beta}$ derive from the theory of unbiased estimating equations. The estimating equation $\hat{S}$ that we have constructed from conditional likelihood components treats all pairs as independent even though they are not, and it also weights the components by inverse probabilities of selecting the pairs; in both of these senses, $\hat{S}$ is a pseudolikelihood, which we call a conditional pseudolikelihood, because of its construction from conditional likelihood components. Although we have included all pairs in our construction of $\hat{S}$ because our intuition is that it would be more efficient, one might consider instead including only a randomly selected subset of those pairs, or even just 1 randomly selected pair per neighborhood. Asymptotic relative efficiency of alternative methods is an area for future research.

The variance of $\hat{\beta}$ can be estimated with a sandwich estimator based on a Taylor series linearization (39, 40). For complex survey sampling involving primary strata and primary sampling

Table 1. Odds Ratios for More Recent Dental Cleaning Derived Using a Conditional Pseudolikelihood Approach, Adjusted for Gender, Age, Education, and Health Insurance and Unadjusted and Adjusted for Confounding by Neighborhood and Income, Florida BRFSS Survey, 2008

	Model 1[a]		Model 2[b]		Model 3[c]		Model 4[d]	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Race/ethnicity								
Non-Hispanic white	1.00		1.00		1.00		1.00	
Non-Hispanic African-American	0.65	0.49, 0.86	0.92	0.55, 1.54	0.82	0.60, 1.13	1.45	0.83, 2.53
Hispanic	1.15	0.87, 1.54	2.34	1.29, 4.26	1.41	1.04, 1.90	3.67	1.79, 7.52
Other	1.26	0.82, 1.92	1.38	0.73, 2.59	1.25	0.79, 1.96	1.10	0.51, 2.37
Gender								
Male	1.00		1.00		1.00		1.00	
Female	1.16	0.98, 1.38	1.31	1.00, 1.72	1.28	1.06, 1.54	1.46	1.06, 2.03
Age, years								
18–34	1.00		1.00		1.00		1.00	
35–54	1.43	1.13, 1.81	1.72	1.22, 2.41	1.35	1.04, 1.75	1.85	1.23, 2.79
55–64	1.66	1.27, 2.16	2.18	1.41, 3.36	1.74	1.31, 2.32	2.57	1.58, 4.17
≥65	1.79	1.39, 2.30	1.84	1.27, 2.69	2.37	1.80, 3.11	3.02	1.90, 4.81
Education								
Less than high school	1.00		1.00		1.00		1.00	
High school or more	2.50	1.81, 3.43	1.90	1.16, 3.13	1.79	1.25, 2.55	1.25	0.70, 2.22
Health insurance status								
No insurance	1.00		1.00		1.00		1.00	
Insurance	3.07	2.42, 3.88	2.92	1.80, 4.73	1.99	1.54, 2.57	1.88	1.12, 3.17
Annual income, dollars								
<15,000					1.00		1.00	
15,000–<25,000					1.39	1.00, 1.93	1.11	0.49, 2.52
25,000–<50,000					2.45	1.77, 3.40	1.91	0.92, 3.95
≥50,000					4.82	3.41, 6.82	4.31	2.12, 8.77

Abbreviations: BRFSS, Behavioral Risk Factor Surveillance System; CI, confidence interval; OR, odds ratio.
[a] Multivariate proportional odds model, not adjusting for neighborhood.
[b] Multivariate proportional odds model, adjusting for neighborhood.
[c] Multivariate proportional odds model, not adjusting for neighborhood but adjusting for household income.
[d] Multivariate proportional odds model, adjusting for neighborhood and household income.

units (clusters) within the primary strata, we need the neighborhoods to be either nested within the primary sampling units or nested within the strata and independent of one another. For the Florida BRFSS data, we have effectively defined neighborhood as the intersection of zip code and primary stratum; furthermore, individuals sampled within one neighborhood are independent of those sampled within another neighborhood, because they are in different households. Although the estimation of and inference regarding β may seem complex, we have discovered a very simple method based on standard software for logistic regression with complex survey data, such as SAS PROC SURVEYLOGISTIC. Our method capitalizes on the relation between conditional logistic regression for matched pairs and ordinary logistic regression (15, 17, 38). Specifically, we can interpret equation 2 as the likelihood for an ordinary logistic regression model with covariates $X^*_{ijl} = X_{ij} - X_{il}$, outcome $Y^*_{ijl} = 1$, and no intercept. Given pairs (V_{kij}, V_{kil}) such that $V_{kij} = V_{kil}$ do not contribute to the likelihood and given that we can arbitrarily order the other pairs such that $V_{kij} = 1$ and $V_{kil} = 0$, the estimator $\widehat{\beta}$ can be computed using standard weighted logistic regression with 1 observation per discordant $(V_{kij} \neq V_{kil})$ pair, weight equal to W_{ijl}, constant outcome Y^*_{ijl}, no intercept, and covariates X^*_{ijl}. Inference can be based on the sandwich estimator explained above, which is implemented in SAS PROC SURVEYLOGISTIC. However, unlike SAS PROC LOGISTIC, SAS PROC SURVEYLOGISTIC will not accept a constant outcome. Therefore, we select an arbitrary number of pairs (e.g., the first 1,000) for reordering, with $Y^*_{ijl} = 0$ and $X^*_{ijl} = X_{il} - X_{ij}$. SAS code for implementing the method is provided in the Appendix.

RESULTS OF ANALYSIS OF BRFSS DATA

The crude odds ratios representing the association between race/ethnicity and more recent dental cleaning,

Table 2. Odds Ratios for More Recent Dental Cleaning Derived Using a Multilevel Model Approach, Adjusted for Gender, Age, Education, Health Insurance, and Neighborhood Averages; Unadjusted and Adjusted for Income; and Unadjusted and Adjusted for Survey Weights, Florida BRFSS Survey, 2008

	Model 5[a]		Model 6[b]		Model 7[c]		Model 8[d]	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Race/ethnicity								
Non-Hispanic white	1.00		1.00		1.00		1.00	
Non-Hispanic African-American	0.77	0.61, 0.95	0.86	0.56, 1.31	1.06	0.83, 1.35	1.02	0.70, 1.48
Hispanic	1.48	1.16, 1.90	1.02	0.64, 1.62	1.97	1.51, 2.57	1.45	0.95, 2.23
Other	0.75	0.60, 0.95	1.18	0.75, 1.88	0.84	0.66, 1.08	1.12	0.73, 1.73
Gender								
Male	1.00		1.00		1.00		1.00	
Female	1.28	1.15, 1.42	1.30	1.06, 1.59	1.43	1.27, 1.60	1.42	1.13, 1.79
Age, years								
18–34	1.00		1.00		1.00		1.00	
35–54	1.26	1.08, 1.48	1.17	0.83, 1.66	1.15	0.97, 1.37	1.24	0.88, 1.76
55–64	1.31	1.12, 1.54	1.28	0.88, 1.85	1.29	1.08, 1.55	1.50	1.06, 2.11
≥65	1.26	1.07, 1.49	1.41	0.96, 2.07	1.56	1.29, 1.88	2.17	1.50, 3.14
Education								
Less than high school	1.00		1.00		1.00		1.00	
High school or more	2.63	2.20, 3.15	2.04	1.33, 3.10	2.00	1.64, 2.43	1.26	0.77, 2.04
Health insurance status								
No insurance	1.00		1.00		1.00		1.00	
Insurance	3.32	2.84, 3.87	3.55	2.48, 5.08	2.19	1.86, 2.58	2.28	1.63, 3.19
Annual income, dollars								
<15,000					1.00		1.00	
15,000–<25,000					1.76	1.46, 2.13	1.62	1.02, 2.57
25,000–<50,000					3.04	2.50, 3.70	2.88	1.77, 4.68
≥50,000					6.06	4.96, 7.40	6.12	3.94, 9.51

Abbreviations: BRFSS, Behavioral Risk Factor Surveillance System; CI, confidence interval; OR, odds ratio.
[a] Multivariate proportional odds model, not adjusting for income or survey weights.
[b] Multivariate proportional odds model, not adjusting for income but adjusting for survey weights.
[c] Multivariate proportional odds model, adjusting for income but not adjusting for survey weights.
[d] Multivariate proportional odds model, adjusting for income and survey weights.

with non-Hispanic whites used as the reference group, were 0.55 (95% confidence interval (CI): 0.42, 0.71) for non-Hispanic African Americans, 0.79 (95% CI: 0.6, 1.03) for Hispanics, and 0.88 (95% CI: 0.59, 1.31) for persons of other race/ethnicity. Results of adjusted analyses are presented in Table 1. Model 1 adjusts for gender, age, education, and health insurance, whereas model 2 additionally adjusts for neighborhood. Model 3 adds income to model 1 (but incurs much more missing data), and model 4 adds income to model 2. We observe in model 1 that adjusting for simple demographic and socioeconomic factors accounts for the near-statistically significant overall disparity observed for Hispanics but not for that of non-Hispanic African Americans. We see from the other 3 models that when we adjust for income or neighborhood, the disparity disappears for non-Hispanic African Americans and reverses for Hispanics. The effect of adjusting for confounding by neighborhood on both odds ratios is dramatic, with or without income in the model.

To compare our approach with existing methods, we also applied the multilevel modeling approach detailed in the Introduction; results are presented in Table 2. Odds ratios for the neighborhood averages of individual-level covariates are not presented because of space limitations. Models 5 and 7 ignore the BRFSS survey weights, whereas models 6 and 8 incorporate them. Models 7 and 8 additionally adjust for confounding by income. We observe no statistically significant association for non-Hispanic African Americans for all models but model 5. We observe a statistically significant positive association for Hispanics in models 5 and 7 but not for models 6 and 8. It is reassuring that the positive association for Hispanics that we observe with our new method is also detected with the multilevel modeling approach that ignores the survey weights. We are not concerned that the multilevel modeling analysis that

incorporates survey weights does not detect that association, because, as we mentioned in the Introduction, that approach enlists a biased estimating equation.

DISCUSSION

We have presented a new method of adjusting for confounding by neighborhood using the proportional odds model and complex survey data. We have applied the method to analyze disparities in dental preventative care assessed as an ordinal outcome, but we note that the method we developed has a much broader reach. First, the method is also applicable to binary outcomes. Second, the method is applicable to simple random samples or data obtained from ordinary cluster sampling, for which no methods for the proportional odds model based on a conditional likelihood currently exist (18, 19). Third, whereas we have conceptualized neighborhood as a geographic region, another possibility is to conceptualize it more generally as any set of cluster-defining characteristics (i.e., as a matched set).

Our application to the 2008 Florida BRFSS survey data served mainly as an illustration, but the results are intriguing. In a model adjusting for gender, age, education, and health insurance, non-Hispanic African Americans were less likely to receive more recent dental cleanings than non-Hispanic whites, and Hispanics were not statistically different from non-Hispanic whites. However, once estimates were adjusted for neighborhood, we saw a striking change in the estimate for Hispanics, in the direction of Hispanics' receiving more dental cleanings than non-Hispanic whites. For non-Hispanic African Americans, we observed an attenuation of the originally observed disparity once confounding by neighborhood was taken into account. While these shifts in the estimates by race/ethnicity are difficult to explain given the limited variables available in the BRFSS, one might conclude that there is better outreach for dental-care access to Hispanic/Latino populations than to persons of other races/ethnicities. There has been increased attention given to oral health care and research in the Latino population since Hispanics/Latinos became the nation's largest and fastest-growing minority group (41, 42). However, it may be shortsighted to concentrate only on the role of ''access'' to dental care. It is likely that the broader social and cultural determinants of oral health practices need to be taken into account. Our example highlights the fact that once neighborhood is taken into account, factors influencing utilization of preventive cleanings may be different for different cultural groups. Patrick et al. (32) provide a conceptual framework for the major influences on oral health disparities, which include factors such as cultural environment (e.g., beliefs about oral health), stressors (e.g., dentist-patient interaction), health behaviors (e.g., alcohol use and smoking), and individual psychology (e.g., fear of dentists). Thus, these relations are likely to be complex, and additional research is needed to understand racial/ethnic differences in utilization of dental cleanings.

For comparison with existing methods, we also applied multilevel modeling approaches that adjusted for confounding due to neighborhood by including either unweighted or weighted neighborhood averages of individual-level covariates as additional terms in the model. This approach to adjusting for neighborhood-level confounding is known as fixed-effects regression (12, 43) in the social sciences, and it is known as the poor man's approach to conditional likelihood methods (8) in the statistical literature. For our purpose of estimating odds ratios corresponding to the individual-level covariates, this method is equivalent to replacing each individual-level covariate with its deviation from its respective neighborhood average (i.e., centering about the neighborhood average) and additionally including the neighborhood averages in the model. It is also possible to center the individual-level covariates but not include the neighborhood averages (44); however, this approach is less robust (8).

Our example analysis was subject to some limitations. The BRFSS survey weights attempt to adjust for unit (individual-level) nonresponse, but cannot do so perfectly. Furthermore, we have handled item (question-level) nonresponse using a complete-case analysis—for our analyses without income, this is not much of a limitation, because the percentage of missing data is small. Finally, it would be optimal to include information about dental insurance in the model, but information on this variable is not included in the BRFSS. According to a 2010 brief by the National Center for Health Statistics (45), non-Hispanic African Americans were more likely to have dental insurance than Hispanics, who were in turn less likely to have dental insurance than non-Hispanic whites. Therefore, we may have underestimated the odds ratios for Hispanics and overestimated them for non-Hispanic African Americans. It would be of further interest to investigate whether the disparities in access to dental insurance reverse when neighborhood effects are taken into account.

ACKNOWLEDGMENTS

Author affiliations: Department of Biostatistics, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, Florida (Babette A. Brumback, Hao W. Zheng); and Department of Health Sciences, Gettysburg College, Gettysburg, Pennsylvania (Amy B. Dailey).

This work was supported by the National Science Foundation; the US Department of Agriculture (National Agricultural Statistics Service and Economic Research Service; the US Department of Education (National Center for Educational Statistics); and the Social Security Administration (grant NSF SES-1115618).

The authors thank Dr. Youjie Huang, Jamie Forrest, and Melissa Murray from the Florida BRFFS office for their helpful support.

Conflict of interest: none declared.

REFERENCES

1. US Department of Health and Human Services. *About Healthy People*. Washington, DC: US Department of Health and Human Services; 2011. (http://www.healthypeople.gov/2020/about/default.aspx). (Accessed November 8, 2011).

2. Carpiano RM. Neighborhood social capital and adult health: an empirical test of a Bourdieu-based model. *Health Place*. 2007; 13(3):639–655.
3. Kramer MR, Hogue CR. Is segregation bad for your health? *Epidemiol Rev*. 2009;31(1):178–194.
4. Pickett KE, Wilkinson RG. People like us: ethnic group density effects on health. *Ethn Health*. 2008;13(4):321–334.
5. Feng J, Glass TA, Curriero FC, et al. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place*. 2010;16(2):175–190.
6. Hoffmann B, Moebus S, Dragano N, et al. Residential traffic exposure and coronary heart disease: results from the Heinz Nixdorf Recall Study. *Biomarkers*. 2009;14(suppl 1):74–78.
7. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*. 1998;54(2):638–645.
8. Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *J R Stat Soc Series B*. 2006;68(5):858–872.
9. Brumback BA, Dailey AB, Brumback LC, et al. Adjusting for confounding by cluster using generalized linear mixed models. *Stat Probab Lett*. 2010;80(21-22):1650–1654.
10. Brumback BA, Dailey AB, He Z, et al. Efforts to adjust for confounding by neighborhood using complex survey data. *Stat Med*. 2010;29(18):1890–1899.
11. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications; 2002.
12. Allison PD. *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage Publications; 2009.
13. Rabe-Hesketh S, Skrondal A. Multilevel modeling of complex survey data. *J R Stat Soc A*. 2006;169(4):805–827.
14. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
15. Agresti A. *Categorical Data Analysis*. 2nd ed. New York, NY: John Wiley & Sons, Inc; 2002.
16. Graubard BI, Korn EL. Conditional logistic regression with survey data. *Stat Biopharm Res*. 2011;3(2):398–408.
17. Brumback BA, He Z. Adjusting for confounding by neighborhood using complex survey data. *Stat Med*. 2011;30(9): 965–972.
18. Agresti A, Natarajan R. Modeling clustered ordered categorical data: a survey. *Int Stat Rev*. 2001;69(3):345–371.
19. Liu I, Agresti A. The analysis of ordered categorical data: an overview and a survey of recent developments. *Test*. 2005; 14(1):1–73.
20. Bieler GS, Brown GG, Williams RL, et al. Estimating model-adjusted risks, risk differences, and risk ratios from complex survey data. *Am J Epidemiol*. 2010;171(5):618–623.
21. Brumback BA, Bouldin ED, Zheng HW, et al. Testing and estimating model-adjusted effect-measure modification using marginal structural models and complex survey data. *Am J Epidemiol*. 2010;172(9):1085–1091.
22. Cheng NF, Han PZ, Gansky SA. Methods and software for estimating health disparities: the case of children's oral health. *Am J Epidemiol*. 2008;168(8):906–914.
23. Tellez M, Sohn W, Burt BA, et al. Assessment of the relationship between neighborhood characteristics and dental caries severity among low-income African-Americans: a multilevel approach. *J Public Health Dent*. 2006;66(1):30–36.
24. Ahn S, Burdine JN, Smith ML, et al. Residential rurality and oral health disparities: influences of contextual and individual factors. *J Prim Prev*. 2011;32(1):29–41.
25. Chattopadhyay A. Oral health disparities in the United States. *Dent Clin North Am*. 2008;52(2):297–318, vi.
26. Sabbah W, Tsakos G, Chandola T, et al. Social gradients in oral and general health. *J Dent Res*. 2007;86(10):992–996.
27. Sabbah W, Tsakos G, Sheiham A, et al. The effects of income and education on ethnic differences in oral health: a study in US adults. *J Epidemiol Community Health*. 2009;63(7): 516–520.
28. Sanders AE, Turrell G, Slade GD. Affluent neighborhoods reduce excess risk of tooth loss among the poor. *J Dent Res*. 2008;87(10):969–973.
29. Bramlett MD, Soobader MJ, Fisher-Owens SA, et al. Assessing a multilevel model of young children's oral health with national survey data. *Community Dent Oral Epidemiol*. 2010;38(4): 287–298.
30. Turrell G, Sanders AE, Slade GD, et al. The independent contribution of neighborhood disadvantage and individual-level socioeconomic position to self-reported oral health: a multilevel analysis. *Community Dent Oral Epidemiol*. 2007;35(3):195–206.
31. Wu B, Plassman BL, Liang J, et al. Differences in self-reported oral health among community-dwelling black, Hispanic, and white elders. *J Aging Health*. 2011;23(2):267–288.
32. Patrick DL, Lee RS, Nucci M, et al. Reducing oral health disparities: a focus on social and cultural determinants. *BMC Oral Health*. 2006;6(suppl 1):S4. (doi:10.1186/1472-6831-6-S1-S4).
33. Doty HE, Weech-Maldonado R. Racial/ethnic disparities in adult preventive dental care use. *J Health Care Poor Underserved*. 2003;14(4):516–534.
34. Castañeda H, Carrion IV, Kline N, et al. False hope: effects of social class and health policy on oral health inequalities for migrant farmworker families. *Soc Sci Med*. 2010;71(11): 2028–2037.
35. Grembowski D, Spiekerman C, Milgrom P. Disparities in regular source of dental care among mothers of Medicaid-enrolled preschool children. *J Health Care Poor Underserved*. 2007;18(4): 789–813.
36. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6): 656–664.
37. Liang KY. Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models. *Biometrics*. 1987;43(2): 289–299.
38. Breslow NE, Day NE, Halvorsen KT, et al. Estimation of multiple relative risk functions in matched case-control studies. *Am J Epidemiol*. 1978;108(4):299–307.
39. Binder DA. On the variances of asymptotically normal estimators from complex surveys. *Int Stat Rev*. 1983;51(3): 279–292.
40. Korn EL, Graubard BI. *Analysis of Health Surveys*. New York, NY: John Wiley & Sons, Inc; 1999.
41. Ramos-Gomez F, Cruz GD, Watson MR, et al. Latino oral health: a research agenda toward eliminating oral health disparities. *J Am Dent Assoc*. 2005;136(9):1231–1240.
42. Mejia GC, Kaufman JS, Corbie-Smith G, et al. A conceptual framework for Hispanic oral health care. *J Public Health Dent*. 2008;68(1):1–6.
43. Allison PD. *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Institute Inc; 2005.
44. Morenoff JD, House JS, Hansen BB, et al. Understanding social disparities in hypertension prevalence, awareness, treatment, and control: the role of neighborhood context. *Soc Sci Med*. 2007;65(9):1853–1866.
45. Bloom B, Cohen RA. *Dental Insurance for Persons Under Age 65 Years With Private Health Insurance: United States, 2008*. (NCHS data brief no. 40). Hyattsville, MD: National Center for Health Statistics; 2010.

APPENDIX

The programming involves 3 main steps: 1) forming all pairs, which we do using SAS PROC SQL; 2) coding the new outcomes and covariates, which we do with SAS PROC SQL and using a SAS DATA step; and then 3) implementing the weighted logistic regression, which we do using SAS PROC SURVEYLOGISTIC. For the BRFSS example, the first 2 steps were completed using the following SAS code. Note that the ordinal outcome denclean has 5 levels: 1, 2, 3, 4, and 5, with 1 indicating the most recent dental cleaning and 5 indicating never. Other variables used in the program are the BRFSS stratum (_strstr), the neighborhood variable (zipstr), the BRFSS sampling weight (_finalwt), education (edu), age (age2, age3, age4), gender (female), race/ethnicity (black, hisp, other), and insurance status (ins). Note that we use the variable "zipstr" as our primary sampling unit in SAS SURVEYLOGISTIC ("cluster zipstr").

```
data denclean;
   set cleanbrfss08;
   if denclean<=1 then den1=1; else den1=0;
   if denclean<=2 then den2=1; else den2=0;
   if denclean<=3 then den3=1; else den3=0;
   if denclean<=4 then den4=1; else den4=0;
run;

%macro ordinal;
%do i=1 %to 4;

proc sql;
   create table den&i.pairs as
   select a._ststr, a.zipstr, a._finalwt*b._finalwt as wt,
   a.edu-b.edu as edu_diff, a.age2-b.age2 as age2_diff,
   a.age3-b.age3 as age3_diff, a.age4-b.age4 as age4_diff,
   a.female-b.female as gender_diff, a.black-b.black as black_diff,
   a.other-b.other as other_diff, a.hisp-b.hisp as hisp_diff,
   a.ins-b.ins as ins_diff,
   a.den&i.-b.den&i. as den_diff
   from denclean as a, denclean as b where (a.zipstr=b.zipstr and a.den&i.>b.den&i.);
quit;

*set an arbitrary number of observations to 0;
data den&i.pairs;
   set den&i.pairs;
   if_n_<=1000 then do;
   den_diff=0;
   edu_diff=-edu_diff;
   age2_diff=-age2_diff;
   age3_diff=-age3_diff;
   age4_diff=-age4_diff;
   ins_diff=-ins_diff;
   gender_diff=-gender_diff;
   black_diff=-black_diff;
   hisp_diff=-hisp_diff;
   other_diff=-other_diff;
   end;
run;

%end;
%mend;

%ordinal;

data allden;
   set den1pairs den2pairs den3pairs den4pairs;
run;

proc surveylogistic data=allden;
   strata_ststr;
   cluster zipstr;
```

```
   model den_diff (event='1')=edu_diff ins_diff gender_diff age2_diff age3_diff age4_diff black_diff hisp_diff other_diff/
noint;
   weight wt;
run;
```

Although we have used SAS macro programming for simplicity, one could have instead cut and pasted the code within the macro 4 times in a row, taking care to make sure each block of code matched the corresponding level of the ordinal variable.