



Practice of Epidemiology

Methods of Estimating or Accounting for Neighborhood Associations With Health Using Complex Survey Data

Babette A. Brumback*, Zhuangyu Cai, and Amy B. Dailey

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

Initially submitted October 14, 2013; accepted for publication February 12, 2014.

Reasons for health disparities may include neighborhood-level factors, such as availability of health services, social norms, and environmental determinants, as well as individual-level factors. Investigating health inequalities using nationally or locally representative data often requires an approach that can accommodate a complex sampling design, in which individuals have unequal probabilities of selection into the study. The goal of the present article is to review and compare methods of estimating or accounting for neighborhood influences with complex survey data. We considered 3 types of methods, each generalized for use with complex survey data: ordinary regression, conditional likelihood regression, and generalized linear mixed-model regression. The relative strengths and weaknesses of each method differ from one study to another; we provide an overview of the advantages and disadvantages of each method theoretically, in terms of the nature of the estimable associations and the plausibility of the assumptions required for validity, and also practically, via a simulation study and 2 epidemiologic data analyses. The first analysis addresses determinants of repeat mammography screening use using data from the 2005 National Health Interview Survey. The second analysis addresses disparities in preventive oral health care using data from the 2008 Florida Behavioral Risk Factor Surveillance System Survey.

conditional logistic regression; generalized linear mixed models; health disparities; health surveys; neighborhood associations; survey analysis

Abbreviations: GLMM, generalized linear mixed model; MSE, mean-squared error; NHIS, National Health Interview Survey; SEP, socioeconomic position.

There is growing interest in how and to what extent neighborhood or residential context influences health (1, 2). Our interest in this problem originated with an investigation of the determinants of mammography use (3). In that study, neighborhood was incorporated into the statistical model through the inclusion of a measured area-level socioeconomic position (SEP) variable. Subsequently, we developed an interest in statistical methods capable of addressing whether unmeasured neighborhood-level factors, in addition to measured individual-level factors, could account for racial/ethnic health disparities (4–10). Additionally, we became interested in methods to quantify the independent contribution of unmeasured neighborhood-level factors beyond that of measured individual- and neighborhood-level factors (10).

Statistical methods for accomplishing the above objectives are reasonably well established in cases in which the data

analysis assumes a simple random sample or ordinary cluster sample, although the matter is not without some confusion. A complicating feature of our investigations was our use of nationally or locally representative complex survey data for the analyses. With complex survey data, individuals typically have unequal probabilities of selection into the sample, and stratification and/or clustering also need to be taken into account. The goal of the present article was to review and compare methods of estimating or accounting for neighborhood associations with complex survey data. We considered 3 types of methods, each generalized for use with complex survey data: ordinary regression, conditional likelihood regression, and generalized linear mixed-model (GLMM) regression. The methods can be programmed using statistical software for complex survey data, available in recent versions of SAS (SAS Institute, Inc., Cary, North Carolina) or Stata (StataCorp

Table 1. Odds Ratios for Repeat Mammography Use, National Health Interview Survey, 2005

Variable	Model 1 ^a		Model 2 ^b		Model 3 ^c	
	OR	95% CI	OR	95% CI	OR	95% CI
Area-level SEP						
Quartile 1	1.0	Referent			1.0	Referent
Quartile 2	0.74	0.61, 0.90			0.62	0.46, 0.83
Quartile 3	0.69	0.55, 0.85			0.78	0.50, 1.22
Quartile 4	0.61	0.48, 0.78			0.61	0.39, 0.85
Age, years						
<50	0.65	0.53, 0.81	0.71	0.45, 1.12	0.61	0.43, 0.86
50–64	1.0	Referent	1.0	Referent	1.0	Referent
65–79	0.84	0.65, 1.07	0.68	0.40, 1.17	1.13	0.76, 1.68
Race/ethnicity						
Hispanic	1.02	0.74, 1.40	1.00	0.19, 5.25	1.19	0.58, 2.42
Non-Hispanic white	1.0	Referent	1.0	Referent	1.0	Referent
Non-Hispanic black	1.30	1.02, 1.64	1.07	0.60, 1.90	1.41	0.98, 2.04
Other	0.60	0.37, 0.97	0.12	0.03, 0.55	0.28	0.14, 0.58
Marital status						
Married	1.0	Referent	1.0	Referent	1.0	Referent
Widowed, separated, or divorced	0.87	0.74, 1.01	0.83	0.60, 1.15	0.74	0.58, 0.94
Never married	0.85	0.62, 1.16	1.31	0.71, 2.43	1.05	0.65, 1.71
Educational level						
<High school	0.54	0.41, 0.70	0.42	0.23, 0.76	0.42	0.24, 0.72
High school	0.69	0.56, 0.86	0.70	0.41, 1.18	0.61	0.42, 0.87
Some college	0.81	0.65, 1.00	0.68	0.39, 1.18	0.78	0.53, 1.17
Completed college	1.0	Referent	1.0	Referent	1.0	Referent
Household income						
<\$20,000	0.67	0.49, 0.92	0.70	0.37, 1.32	0.65	0.39, 1.10
\$20,000–\$44,999	0.96	0.74, 1.25	0.87	0.50, 1.50	0.81	0.55, 1.21
\$45,000–\$74,999	0.93	0.72, 1.19	0.76	0.43, 1.37	0.76	0.52, 1.12
≥\$75,000	1.0	Referent	1.0	Referent	1.0	Referent
>\$20,000 but not otherwise specified	0.82	0.62, 1.07	0.73	0.41, 1.30	0.83	0.45, 1.53
Do not know/refused to answer	0.85	0.60, 1.21	0.76	0.26, 2.22	0.73	0.42, 1.26
Years in the United States						
Native born	1.0	Referent	1.0	Referent	1.0	Referent
Foreign born; <10 years in the United States	0.73	0.39, 1.38	1.28	0.26, 6.34	0.94	0.38, 2.31
Foreign born; ≥10 years in the United States	1.15	0.86, 1.55	0.99	0.44, 2.23	1.11	0.76, 1.64

Table continues

LP, College Station, Texas). As we shall see, the methods rely on different assumptions; therefore, the best choice of method will vary from one research study to another. For that reason, we consider not just 1 but 2 example research studies for illustration in this article. We also conduct a simulation study.

TWO EXAMPLE RESEARCH STUDIES

The first example is an investigation of the determinants of mammography use, as in the study by Dailey et al. (3), using data from the 2005 National Health Interview Survey (NHIS) (11) merged with census tract-level data from the 2000 US

Census. We used the same variables that were used in the investigation by Dailey et al. (3), except with more categories for income and insurance coverage. We re-analyzed the data using 3 competing methods. The first column of Table 1 shows the variables included in the analyses. The response is a dichotomous measure of repeat mammography screening use, for which respondents needed to have reported having 3 or more mammograms in the past 6 years and having the most recent mammogram within the past 2 years. The sole measured neighborhood-level determinant is area-level SEP, created from the census tract-level data and based on the work of Krieger et al. (12), which uses a standardized *z* score combining

Table 1. Continued

Variable	Model 1 ^a		Model 2 ^b		Model 3 ^c	
	OR	95% CI	OR	95% CI	OR	95% CI
Insurance coverage						
None	0.48	0.37, 0.63	0.65	0.38, 1.11	0.55	0.36, 0.86
Public only	0.83	0.62, 1.11	1.10	0.61, 1.96	1.01	0.66, 1.56
Private only	1.0	Referent	1.0	Referent	1.0	Referent
Public and private	1.20	0.88, 1.62	1.85	0.93, 3.69	1.04	0.67, 1.63
Usual place of care						
No	1.0	Referent	1.0	Referent	1.0	Referent
Yes	1.77	1.27, 2.48	6.72	2.84, 15.9	2.56	1.30, 5.04
Saw a physician in last year						
No	1.0	Referent	1.0	Referent	1.0	Referent
Yes	3.48	2.84, 4.26	3.92	2.41, 6.40	3.97	2.61, 6.04
Mammogram recommended in past year						
No	0.23	0.20, 0.27	0.21	0.14, 0.32	0.21	0.16, 0.27
Yes	1.0	Referent	1.0	Referent	1.0	Referent
Do not know/refused to answer	0.03	0.01, 0.05	0.02	0.01, 0.08	0.03	0.01, 0.06
Perceived breast cancer risk						
Low	0.83	0.71, 0.98	0.59	0.41, 0.83	0.77	0.57, 1.03
Medium	1.0	Referent	1.0	Referent	1.0	Referent
High	1.31	1.01, 1.70	1.12	0.59, 2.15	1.10	0.66, 1.82
Do not know/refused to answer	0.81	0.61, 1.08	0.53	0.28, 1.01	0.80	0.49, 1.31
Family history of cancer						
No	1.0	Referent	1.0	Referent	1.0	Referent
Yes	1.20	1.03, 1.40	1.08	0.79, 1.49	1.21	0.97, 1.52
Smoking status						
Current smoker	0.61	0.50, 0.75	0.47	0.32, 0.71	0.55	0.38, 0.77
Former smoker; quit <6 years ago	0.90	0.65, 1.24	0.61	0.27, 1.35	0.73	0.44, 1.19
Former smoker; quit ≥6 years	1.19	1.00, 1.42	1.01	0.69, 1.46	1.23	0.94, 1.60
Never	1.0	Referent	1.0	Referent	1.0	Referent
Activity limitations						
No	1.0	Referent	1.0	Referent	1.0	Referent
Yes	0.71	0.60, 0.84	0.46	0.31, 0.69	0.63	0.49, 0.82

Abbreviations: CI, confidence interval; OR, odds ratio; SEP, socioeconomic position.

^a Standard logistic regression for complex survey data.

^b Logistic regression using a conditional pseudolikelihood.

^c Logistic regression using a generalized linear mixed model and a marginal pseudolikelihood.

data on percentage that are working class, percentage that are unemployed, percentage that live below the US poverty line, percentage without a high school education, median housing values, and median household income. A higher score indicates a higher degree of deprivation. As in the study by Dailley et al. (3), we translated the SEP index into within-state quartiles so that the most advantaged census tracts in one state equated with the most advantaged ones in another.

The second example is obtained from the recent literature (8, 10) and addresses whether measured and/or unmeasured neighborhood-level factors could account for racial/ethnic health disparities in preventive oral health care, using data from the 2008 Florida Behavioral Risk Factor Surveillance

System Survey (13) merged with 2000 US Census Zip Code Tabulation Area data. The response is an ordinal measure of recency of preventive dental care. Neighborhood is defined as zip code crossed with Behavioral Risk Factor Surveillance System survey stratum.

METHODS TO ESTIMATE OR ACCOUNT FOR A NEIGHBORHOOD ASSOCIATION

For ease of exposition, we present each of the methods assuming a binary outcome and a logit link. Generalization to other types of outcomes and links is discussed with references to the literature. Let Y_{ij} be the outcome for individual

j in neighborhood i , X_{ij} be the measured individual-level covariates, and M_i be the measured neighborhood-level covariates. SAS and Stata code for implementation are provided in the Appendix.

Ordinary regression with complex survey data

We consider 3 ordinary logistic regression models for use with complex survey data:

$$\text{Logit}(P(Y_{ij} = 1|X_{ij}, i, M_i)) = X_{ij}\beta, \quad (1)$$

$$\text{Logit}(P(Y_{ij} = 1|X_{ij}, M_i)) = X_{ij}\beta + b_i, \quad (2)$$

and

$$\text{Logit}(P(Y_{ij} = 1|X_{ij}, i, M_i)) = X_{ij}\beta + M_i\alpha, \quad (3)$$

where β and α represent the associations with the individual and neighborhood covariates and b_i is estimated as a fixed effect that results from including i as a categorical covariate in the model. Estimations of β , α , and the b_i are based on a sum of weighted score functions (14) for each individual, using the complex survey weights W_{ij} . The sampling distributions used to obtain confidence intervals for these parameters are estimated under the assumption that the weighted score functions cluster into independent groups within the primary strata of the complex survey design. For example, if the primary sampling unit for the complex survey design is either the neighborhood or the individual, then the neighborhoods represent the independent groups; alternatively, if the neighborhoods are each nested into exactly one primary sampling unit, then the primary sampling units represent the independent groups. Either a sandwich estimator of variance for complex survey data (15, 16) coupled with an asymptotic normal assumption or the bootstrap for complex survey data (16) can be used to estimate the sampling distribution. Standard software such as SAS PROC SURVEYLOGISTIC implements the former, with the variable denoting the independent groups as the clustering variable. If there are primary strata, the independent groups must each be nested into exactly 1 stratum. Ignoring primary stratification is also possible; this will result in conservative (larger) confidence intervals, although often not much larger.

In all 3 models, β represents the effects of unit increases in selected components of X_{ij} on average in the population, provided that the other components and the included functions of i or M_i form a sufficient set of confounders. With model 2, the b_i represents the effects of unmeasured neighborhood-level variables. For approximately consistent estimation, the sample sizes within neighborhoods need to be large. Therefore, for many research studies, this approach might fail. However, for the other research studies, this approach enables adjustment for confounding by unmeasured neighborhood-level variables. It can also be used to quantify the unmeasured neighborhood-level effects. Model 3 can be regarded as a compromise between models 1 and 2. It adjusts for confounding by measured neighborhood covariates M_i but not confounding by unmeasured neighborhood covariates. The parameter α represents the effect of the neighborhood covariates assuming the individual-level components of X_{ij} form a sufficient set of confounders.

For other outcome types, generalizations of models 1–3 can be used analogously to the ones for binary outcomes. Standard software such as SAS PROC GENMOD can be used if there is no need to adjust for primary stratification in estimating the sampling distribution.

Consistent estimation may not be achieved unless the complex survey design satisfies a positivity assumption (17); that is, each individual in the population must have a positive probability of selection into the sample. When the positivity assumption is not satisfied, other assumptions must hold to guard against bias. For example, in some cases, it may be that the positivity assumption is approximately satisfied, whereas in others it may be that the excluded individuals are a random subset of the eligible individuals.

Conditional pseudolikelihood regression with complex survey data

There is just one method for binary outcomes in this category, and it is based on model 2. The estimation uses a sum of weighted conditional score functions for each within-neighborhood pair of observations such that $Y_{ij} \neq Y_{il}$. The weights W_{ijl} represent the probability that pair (ij, il) is selected into the survey sample. By using conditional score functions, we can consistently estimate β and adjust for unmeasured neighborhood-level variables even with small within-neighborhood sample sizes. The price for this is that the b_i become inestimable; it is also not possible to include measured neighborhood covariates in the model. Each conditional score function stems from the conditional likelihood that $Y_{ij} = a$ and $Y_{il} = b$ given that $Y_{ij} + Y_{il} = 1$, where a and b equal 0 or 1. Let Y_{ijl} equal 1 if $Y_{ij} = 1$ and $Y_{il} = 0$ and let it equal 0 if $Y_{ij} = 0$ and $Y_{il} = 1$. As it turns out, the conditional likelihood equals the ordinary logistic regression likelihood for a model with Y_{ijl} as the outcome and with no intercept, differenced covariates $X_{ij} - X_{il}$, and parameter β (7–9, 18, 19). Therefore, standard software such as SAS PROC SURVEY LOGISTIC can be used to estimate β with weights W_{ijl} , and estimation of the sampling distribution is much the same as previously described, based on independent groups of the paired observations. Analogous methodology has been developed for other outcome types, such as ordinal, multinomial, and non-negative outcomes; see Brumback et al. (8, 9) for details.

As with ordinary regression with complex survey data, a positivity assumption is generally necessary for consistent estimation. Because of the pairwise weights, the survey design must place a positive probability on the selection of each within-neighborhood pair in the population. As before, violation of this positivity assumption requires alternative assumptions entailing negligible bias. Methods of this type are only practical when the pairwise weights W_{ijl} can be adequately approximated. For some types of surveys and specifications of neighborhood, knowledge of W_{ijl} may be out of reach.

GLMM regression with complex survey data

These methods are based on the following GLMM, which augments model 2 with distributional assumptions about the b_i :

$$\text{Logit}(P(Y_{ij} = 1|X_{ij}, i, M_i)) = X_{ij}\beta + b_i \quad (4)$$

$b_i = M_i\alpha + \delta_i$, where the δ_i are independent mean zero normal random variables with standard deviation τ and δ_i is jointly independent of M_i and X_{ij} , where M_i again represents a set of measured neighborhood covariates. Two popular options are to let M_i equal the intercept and \bar{X}_i , which denotes the within-neighborhood averages of the X_{ij} (20–24), or to let M_i contain the intercept only.

Estimands of possible interest include components of β , α , and τ . The magnitude of independent unmeasured neighborhood effects is quantified through τ . The parameters β , α , and τ can be estimated in 2 ways. The first way was introduced by Rabe-Hesketh and Skrondal (25) and involves the weights W_{ji} and W_i , which are the inverse probability of sampling individual j given neighborhood i and the inverse probability of sampling neighborhood i . W_{ji} is used to weight the likelihood corresponding to the distribution of Y_{ij} given X_{ij} and δ_i , and W_i is used to weight the resulting pseudolikelihood after integrating over the distribution of δ_i . This method can produce inconsistent estimators when the sample sizes within a neighborhood are small (10), because the weights W_{ji} enter into the score function nonlinearly. Therefore, an alternative method of estimation was introduced by Brumback et al. (10) and involves the pairwise weights W_{ijl} described above. W_{ijl} is used to weight the marginal likelihood (integrated over δ_i) of the within-neighborhood pair (ij, il) . Because the resulting score function is a linear function of the W_{ijl} , the method produces consistent estimators even when the sample sizes within a neighborhood are small. Both methods are easily programmed using the GLLAMM macro within Stata (10, 25), and the same software can handle other outcome types with analogous versions of model 4.

For validity, the first method requires the positivity assumption that each individual within a neighborhood has a positive probability of selection and that each neighborhood has a positive probability of selection. The second method requires that each pair of individuals within a neighborhood has a positive probability of selection. If these assumptions are violated, other assumptions entailing negligible bias must be made.

There is another subtle issue affecting the validity of both methods when the sample sizes within a neighborhood are small and M_i is set equal to \bar{X}_i . The issue is that model 4 is a model for the population, in which \bar{X}_i is a census average within neighborhood i . When sampled data are used to estimate the parameters of model 4, typically one would use a weighted sample average \bar{X}_i^w of the X_{ij} within neighborhood i (with weights W_{ij}) in place of \bar{X}_i , as census data on the individual-level covariates are often not available. However, this can lead to inconsistent estimators of β , α , and τ (10).

SIMULATION STUDIES

We conducted 2 simulation studies to compare these methods in terms of the mean-squared error (MSE) for estimating β based on 500 iterations. In the first study, we generated 1,000 clusters and 1,000 individuals per cluster. We sampled U_i as $N(0,1)$, X_{ij} as Bernoulli with mean $\text{expit}(U_i) = \exp(U_i)/(1 + \exp(U_i))$, δ_i as $N(0,1)$, $b_i = -5\bar{X}_i + \delta_i$, and Y_{ij} as Bernoulli with mean $\text{expit}(X_{ij} \beta + b_i)$. We then sampled

Table 2. Results of the 2 Simulation Studies Reporting on 500 Estimates of β^a

Model	Study 1		Study 2	
	Mean (SE)	MSE	Mean (SE)	MSE
O1	-2.47 (0.005)	0.570	0.932 (0.002)	0.189
O2	0.155 (0.011)	0.176	0.912 (0.011)	0.227
O3	0.435 (0.005)	0.019	0.182 (0.003)	0.105
O4	0.091 (0.006)	0.186	0.172 (0.002)	0.109
C	0.496 (0.008)	0.033	0.514 (0.006)	0.019
G1	0.492 (0.008)	0.028	0.367 (0.005)	0.028
G2	0.172 (0.009)	0.144	0.385 (0.004)	0.022

Abbreviations: MSE, mean-squared error; SE, standard error.
^a Truth = 0.5.

observations with probability 0.002 if $X_i = Y_i$ and 0.004 otherwise to induce a complex sampling design.

In the second study, we generated 1,000 clusters and 4 individuals per cluster. We sampled U_i as $N(0,0.25^2)$, X_{ij} as $N(U_i, 1)$, $b_i = 2 \max(X_{ij}$ in cluster i) if $U_i > 0$, and $b_i = 2 \min(X_{ij}$ in cluster i) if $U_i < 0$. We then sampled observations with probability of 0.4 if $U_i > 0$ and $Y_{ij} = 1$ or $U_i < 0$ and $Y_{ij} = 0$; otherwise, we used a probability of 0.8.

We applied 7 methods of estimation for each study. Let O1 and O2 denote the first 2 ordinary logistic regression methods and let O3 and O4 denote the third method with $M_i = \bar{X}_i$ and $M_i = \bar{X}_i^w$, respectively. Let C denote the conditional pseudolikelihood method with pairwise weights, and let G1 and G2 denote the GLMM methods with pairwise weights and $M_i = \bar{X}_i$ and $M_i = \bar{X}_i^w$, respectively. The results are presented in Table 2. For study 1, methods O2, C, and G1 are based on correctly specified models. However, the O2 estimator is severely biased, with a relatively large MSE, because of the small within-cluster sample sizes. Surprisingly, although the estimator for method O3 is biased, with a mean more than twice its standard error away from the truth (0.5), its MSE is the smallest among all of the methods, followed by G1 and C, which have relatively similar MSEs. For study 2, only methods O2 and C are based on correctly specified methods, and again because of the small within-cluster sample sizes, method C outperforms all of the methods in terms of its MSE. Methods G1 and G2 are biased, with somewhat larger MSEs.

COMPARISON IN TERMS OF THEORY AND SIMULATIONS

If a large within-neighborhood sample size is available, then ordinary regression with model 2 or the GLMM method enable estimation of β and of the distribution of the b_i . With either approach, the estimate of β will be adjusted for confounding due to unmeasured neighborhood covariates. With model 2, the distribution of b_i can be estimated as the sample distribution of the estimated b_i , whereas with the GLMM approach, it is approximated as Gaussian with a mean of $M_i\alpha$ and a standard deviation of τ . The ordinary regression methods are the easiest to implement.

If the within-neighborhood sample sizes are small, then ordinary regression with model 3 is an easy method to implement; however, we see from the second simulation study that it can lead to a biased estimator of β . Furthermore, it requires correct knowledge of M_i . The conditional pseudolikelihood method is perhaps the best overall choice because it produces consistent estimates and competitive MSEs. The GLMM methods are a reasonable alternative if either census data are available for \bar{X}_i or a sufficient set of measured confounders M_i are available.

A COMPARISON OF THE METHODS IN TERMS OF THE 2 EXAMPLE STUDIES

For the first example research study, we applied 3 different methods; see Table 1 for results. We applied methods that did not require large sample sizes within a given neighborhood because the sample sizes within the neighborhoods ranged from 1 to 8. The first method used ordinary logistic regression for complex survey data with model 3, where we let M_i be the SEP variable. This approach enabled us to estimate the association of SEP with repeat mammography use, adjusted for confounding by individual-level factors. We can observe in the first column of Table 1 that higher levels of deprivation are associated with lower odds of repeat mammography use. We also were able to estimate the associations of certain individual-level covariates after adjustment for confounding by the other ones and by the neighborhood covariate SEP. For example, we found that non-Hispanic blacks had increased odds of repeat mammography use, after adjustment, relative to non-Hispanic whites.

The second method we applied was the conditional pseudolikelihood method based on model 2. We specified neighborhood as the secondary sampling unit of the NHIS. We formed the pairwise weights W_{ijl} as $W_i W_{jl} W_{li}$ according to the in-house survey design probabilities available through the National Center for Health Statistics Research Data Center. This method enabled us to estimate the associations of certain individual-level covariates with the outcome while fully adjusting for any unmeasured confounding due to neighborhood covariates. It is not possible to estimate the associations with measured neighborhood covariates using this approach. We can observe in the second column of Table 1 that according to the results of applying this method, non-Hispanic blacks do not have a statistically significant increased odds of repeat mammography use, after adjustment, relative to non-Hispanic whites.

The third method we applied was the GLMM method based on model 4 with M_i specified as the SEP variable. We used the second GLMM method of estimation with the secondary sampling unit as the neighborhood and the same pairwise weights W_{ijl} specified above. The results are similar to those based on the first method, because the estimate of τ^2 was 1.51×10^{-13} (standard error, 7.14×10^{-11}), which is essentially zero. Because of the small within-neighborhood sample sizes and the inclusion of several individual-level covariates, this approach is underpowered for estimating the effect of unmeasured neighborhood covariates; this explains the near-zero estimate of τ^2 . Differences between results of the first and third methods are due to the use of a pairwise

pseudolikelihood for the third method but an ordinary pseudolikelihood for the first.

For this research study, the positivity assumption required for validity of the second and third methods is violated. Only 1 random adult per household is surveyed for the NHIS mammography questions, which means that the probability of sampling some pairs within each secondary sampling unit neighborhood is zero. Therefore, the results of the second and third method are prone to bias unless we assume that the pairs with zero probability of selection are essentially a random subset of all eligible pairs. As the pairs with zero probability are from the same household and thus tend to be genetically related and share similar lifestyle habits, we are reluctant to rely on this strong assumption. We conclude that for the first example research study, the first method we applied is the best method available.

The second example research study comes from recent the literature (8, 10). Tables 1 and 2 in the first study (8) and Table 1 in the second (10) show the results of applying 3 different types of methods to estimate adjusted racial/ethnic disparities in health behaviors. We include this example mainly to demonstrate that there are some research studies for which the use of methods relying on the pairwise weights W_{ijl} is viable, unlike for our first example research study. In this study, neighborhood is specified as the intersection of zip code and Behavioral Risk Factor Surveillance System stratum, and the weights W_{ijl} can be well approximated as $W_{ij} W_{il}$ because of the approximate independent sampling of individuals. The pairwise positivity assumption is reasonable, even though it is not strictly correct. The study design samples only 1 adult per household, as with the first research study, but if we were to allow independent sampling of multiple adults per household, we would sample very few pairs within a zip code who were also within the same household. Therefore, the positivity assumption holds up to a negligible violation.

As with the first research study, in the second research study the GLMM method leads to an estimated τ^2 of essentially zero. We suspect that this will occur in many research studies with several individual-level covariates and small within-neighborhood sample sizes. Thus, the GLMM approaches may be severely underpowered for capturing the associations with unmeasured neighborhood covariates in these settings. In the second research study, we prefer the second method based on the conditional pseudolikelihood because the main focus is on adjusting the estimated racial/ethnic disparities for all neighborhood covariates, both measured and unmeasured. In that study, it furthermore turned out that neighborhood served as a good proxy for income, which is notoriously difficult to measure.

DISCUSSION

We have presented and compared several methods of estimating or accounting for neighborhood associations with health using complex survey data. We have focused on design-based methods rather than model-based methods for accounting for the complex survey design (26–28), and as such we have considered only frequentist methods. Model-based methods utilize a regression model that reflects features of the complex sampling design, for example conditioning on covariates that

control the sample selection (26–28). Bayesian methods further incorporate prior distributions for the model parameters (26–28). We are unaware of existing model-based methods that account for the complex survey design while simultaneously adjusting for confounding due to unmeasured neighborhood characteristics.

We emerge from our investigation with a preference and recommendation for using methods based on ordinary regression for complex survey data when it is difficult to determine the probability of sampling within-neighborhood pairs or when the pairwise positivity assumption is violated. Furthermore, sampling weights for complex surveys typically incorporate post-stratification adjustment for nonresponse, and this requires further assumptions when forming pairwise weights (8). When pairwise weights are feasibly obtained and the pairwise positivity assumption is approximately correct, we recommend the conditional pseudolikelihood methods because they allow for adjustment for unmeasured neighborhood covariates. However, the GLMM methodology performed well when M_i was correctly specified, and it also has the advantage of quantifying the unmeasured neighborhood associations. Therefore, it will continue to play a role in research on neighborhood associations with health.

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, Zhuangyu Cai); and Department of Health Sciences, Gettysburg College, Gettysburg, Pennsylvania (Amy B. Dailey).

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

Conflict of interest: none declared.

REFERENCES

1. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci.* 2010;1186:125–145.
2. Schempf AH, Kaufman JS. Accounting for context in studies of health inequalities: a review and comparison of analytic approaches. *Ann Epidemiol.* 2012;22(10):683–690.
3. Dailey AB, Brumback BA, Livingston MD, et al. Area-level socioeconomic position and repeat mammography screening use: results from the 2005 National Health Interview Survey. *Cancer Epidemiol Biomarkers Prev.* 2011;20(11):2331–2344.
4. 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.
5. 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.
6. Graubard BI, Korn EL. Conditional logistic regression with survey data. *Stat Biopharm Res.* 2011;3(2):398–408.
7. Brumback BA, He Z. Adjusting for confounding by neighborhood using complex survey data. *Stat Med.* 2011;30(9):965–972.
8. Brumback BA, Dailey AB, Zheng HW. Adjusting for confounding by neighborhood using a proportional odds model and complex survey data. *Am J Epidemiol.* 2012;175(11):1133–1141.
9. Brumback BA, Cai Z, He Z, et al. Conditional pseudolikelihood methods for clustered ordinal, multinomial, or count outcomes with complex survey data. *Stat Med.* 2013;32(8):1325–1335.
10. Brumback BA, Zheng HW, Dailey AB. Adjusting for confounding by neighborhood using generalized linear mixed models and complex survey data. *Stat Med.* 2013;32(8):1313–1324.
11. Botman SL, Moore TF, Moriarity CL, et al. Design and estimation for the National Health Interview Survey, 1995–2004. National Center for Health Statistics. *Vital Health Stat.* 2000;2(130):1–31.
12. Krieger N, Chen JT, Waterman PD, et al. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. *Am J Epidemiol.* 2002;156(5):471–482.
13. Remington PL, Smith MY, Williamson DF, et al. Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981–87. *Public Health Rep.* 1988;103(4):366–375.
14. Faraway JJ. *Extending the Linear Model With R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models.* Boca Raton, FL: Chapman & Hall; 2006.
15. Binder DA. On the variances of asymptotically normal estimators from complex surveys. *Int Stat Rev.* 1983;51(3):279–292.
16. Korn EL, Graubard BI. *Analysis of Health Surveys.* New York, NY: John Wiley & Sons; 1999.
17. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168(6):656–664.
18. 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.
19. Agresti A. *Categorical Data Analysis.* 2nd ed. Hoboken, NJ: John Wiley & Sons; 2002.
20. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics.* 1998;54(2):638–645.
21. Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *J R Stat Soc Series B Stat Methodol.* 2006;68(5):859–872.
22. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA: Sage Publications; 2002.
23. Allison PD. *Fixed Effects Regression Models.* Los Angeles, CA: Sage; 2009.
24. Allison PD. *Fixed Effects Regression Methods for Longitudinal Data Using SAS.* Cary, NC: SAS Institute Inc.; 2005.

25. Rabe-Hesketh S, Skrondal A. Multilevel modelling of complex survey data. *J R Stat Soc Ser A Stat Soc.* 2006;169(4):805–827.
26. Binder DA, Roberts GR. Design-based and model-based methods for estimating model parameters. In: Chambers RL, Skinner CJ, eds. *Analysis of Survey Data*. West Sussex, England: John Wiley & Sons; 2003:29–48.
27. Little RJ. The Bayesian approach to sample survey inference. In: Chambers RL, Skinner CJ, eds. *Analysis of Survey Data*. West Sussex, England: John Wiley & Sons; 2003: 49–57.
28. Gelman A. Struggles with survey weighting and regression modeling. *Stat Sci.* 2007;22(2):153–164.

APPENDIX

SAS CODE FOR CONSTRUCTING PAIRWISE DATA SETS FOR METHODS C, G1, and G2:

```
*For methods G1 and G2;
proc sql;
create table match as
  select
    one.i as i,
    one.j as one_j, two.j as two_j,
    one.x_ij as x_ij_1, two.x_ij as x_ij_2,
    one.y_ij as y_ij_1, two.y_ij as y_ij_2,
    one.wt*two.wt as weight_prod,
    one.b_i as b_i,
    one.xi_bar as xi_bar,
    one.xi_hat as xi_hat
  from sim.samp one, sim.samp two
  where (one.i=two.i and one.j<two.j);
quit;

*For method C;
data pairs;
  set match;
  if y_ij_1=y_ij_2 then delete;
  if y_ij_1=1 and y_ij_2=0 then do; y_ij1=1; diff=x_ij_1-x_ij_2; end;
  if y_ij_1=0 and y_ij_2=1 then do; y_ij1=0; diff=x_ij_1-x_ij_2; end;
run;
```

SAS CODE FOR METHODS O1, O2, O3, and O4:

```
proc surveylogistic data=sim.samp;
ods output ParameterEstimates=ord1;
model Y_ij (ref='0') = X_ij;
Weight wt;
run;

proc surveylogistic data=sim.samp;
ods output ParameterEstimates=ord2;
class i;
model Y_ij (ref='0') = X_ij i;
Weight wt;
run;

proc surveylogistic data=sim.samp;
ods output ParameterEstimates=ord3;
model Y_ij (ref='0') = X_ij Xi_bar;
Weight wt;
run;
```



```
proc surveylogistic data=sim.samp;
ods output ParameterEstimates=ord4;
model Y_ij (ref='0') = X_ij Xi_hat;
Weight wt;
run;
```

SAS CODE FOR METHOD C:

```
proc surveylogistic data=pairs;
ods output ParameterEstimates=cplout1;
cluster i;
model y_ij1(ref='0')=diff/link=logit noint;
Weight Weight_prod;
run;
```

STATA CODE FOR METHODS G1 and G2:

```
ren x_ij_1 x_ij1
ren y_ij_1 y_ij1
ren x_ij_2 x_ij2
ren y_ij_2 y_ij2

gen final_id=_n

reshape long x_ij y_ij, i(final_id) j(pair_id)
gen wt2=weight_prod/10000
gen wt1=1

gllamm y_ij x_ij xi_bar, i(final_id) pweight(wt) link(logit)
family(binomial) cluster(i) robust adapt trace

gllamm y_ij x_ij xi_hat, i(final_id) pweight(wt) link(logit)
family(binomial) cluster(i) robust adapt trace
```