Annals of Epidemiology 22 (2012) 683–690

Contents lists available at SciVerse ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Accounting for context in studies of health inequalities: a review and comparison of analytic approaches

Ashley H. Schempf PhD [a,*], Jay S. Kaufman PhD [b]

[a] Office of Epidemiology, Policy & Evaluation, Maternal and Child Health Bureau, Health Resources and Services Administration, Rockville, MD 20857
[b] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

A R T I C L E I N F O

Article history:
Received 19 October 2011
Accepted 25 June 2012
Available online 2 August 2012

Keywords:
Health status disparities
Minority health
Residence characteristics
Multilevel analysis
Econometric models
Statistical models
Gestational age
Preterm birth

A B S T R A C T

Background: A common epidemiologic objective is to evaluate the contribution of residential context to individual-level disparities by race or socioeconomic position.
Purpose: We reviewed analytic strategies to account for the total (observed and unobserved factors) contribution of environmental context to health inequalities, including conventional fixed effects (FE) and hybrid FE implemented within a random effects (RE) or a marginal model.
Methods: To illustrate results and limitations of the various analytic approaches of accounting for the total contextual component of health disparities, we used data on births nested within neighborhoods as an applied example of evaluating neighborhood confounding of racial disparities in gestational age at birth, including both a continuous and a binary outcome.
Results: Ordinary and RE models provided disparity estimates that can be substantially biased in the presence of neighborhood confounding. Both FE and hybrid FE models can account for cluster level confounding and provide disparity estimates unconfounded by neighborhood, with the latter having greater flexibility in allowing estimation of neighborhood-level effects and intercept/slope variability when implemented in a RE specification.
Conclusions: Given the range of models that can be implemented in a hybrid approach and the frequent goal of accounting for contextual confounding, this approach should be used more often.
Published by Elsevier Inc.

Introduction

Accounting for the contribution of residential context to health inequalities is an increasingly common goal of epidemiologists. This goal derives from the recognition that place plays an important role in influencing population health, coupled with the extent of residential segregation by race and class in the United States and other countries. Investigators sometimes seek to address the confounding that results from the imbalance of an individual-level factor across contexts in the design of a study, for example by selecting subjects in only balanced neighborhoods [1] or through balanced or stratified selection of each race group across a variety of neighborhoods as in a frequency matched design. However, the selection of only balanced neighborhoods may restrict generalizability and does not provide estimates of the crude disparity that are necessary to quantify the magnitude of confounding. Although the matched approach could provide an estimate of the crude disparity if sampling fractions were known, design-based approaches are not an option for secondary data analyses.

In the analytic phase of population-based studies, specific measured contextual factors can be evaluated in multilevel or generalized estimating equation (GEE) models that appropriately account for the nonindependence of observations within a contextual unit. Propensity score methods also can be used to account for observed contextual factors and may provide greater model parsimony and the opportunity to check for covariate balance [2,3]. Unfortunately, not everything that affects an outcome is known and reliably measured. For example, racially segregated neighborhoods may be different in many ways (e.g., environmental pollution, crime, access/quality of health care, walkability, food environment) irrespective of neighborhood socioeconomic status [4–6], which is often the only measured contextual characteristic, or at least a crude proxy for integral features. Although it is important to assess all theoretically relevant factors, it may not always be feasible, and residual disparities may therefore be falsely attributed to other domains unless unobserved contextual differences also have been controlled. Thus, the total or overall contribution of contextual factors to individual-level

* Corresponding author. Office of Epidemiology, Policy & Evaluation, Maternal and Child Health Bureau, Health Resources and Services Administration, 5600 Fishers Lane, Room 13–45, Rockville, MD 20857.
E-mail address: aschempf@hrsa.gov (A.H. Schempf).

1047-2797/$ – see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.annepidem.2012.06.105

disparities should be evaluated and compared with the contribution of specific contextual factors that are observed.

Statistical methods that account for the total contribution of contextual factors essentially form matched contrasts within contextual units and are generally referred to as "fixed effects" (FE) models [7–9]. For example, contrasting or stratifying individuals according to neighborhoods provides complete control for contextual confounding because all group-level characteristics are the same within a given neighborhood. This is the same logic as matching in case-control studies, in which contrasts are conditioned on the matched set. Fixed-area or within-area contrasts can be achieved through a number of common model specifications, including a special form of the random effects (RE) model [10,11]. Although these techniques have been reviewed elsewhere for nonspatial clusters (e.g., patients within health care sites, siblings within families) [12–16], they are still surprisingly infrequent in epidemiologic studies of neighborhoods or other geographic clusters [17–22]. We therefore aim to provide a review of these techniques that can be used to control analytically for contextual confounding with the aid of an applied example of evaluating neighborhood confounding of racial disparities in gestational age at birth, including both a continuous and binary outcome. We compare and contrast the various options, address some potential misconceptions, and provide recommendations.

A review of analytic methods with an applied example

To illustrate and describe the various analytic approaches of accounting for the total contextual component of health disparities, we use the example objective of determining the neighborhood (census block group) contribution to black-white racial disparities in continuously measured gestational age and a dichotomous measure of preterm birth (PTB, <37 weeks) using 1999–2001 birth certificate data for Durham and Wake Counties, North Carolina. The goal is to quantify the contribution of neighborhood inequalities (i.e., residential segregation) to racial disparities in these birth outcomes by comparing unadjusted disparities with those from analytic approaches that control for neighborhood-level confounding through within-neighborhood inference. The hypothetical experiment to control for neighborhood confounding would involve randomizing women of both races to live across the distribution of neighborhoods. However, neighborhood randomization can often fail because of selective nonparticipation or noncompliance, which cast doubts on both external and internal validity and would then require analyses that treat the data as observational [23]. Of course, the observational approaches reviewed here are also prone to self-selection because choice of neighborhood is dependent on many factors including income. For simplicity in this illustrative example, we ignore selection effects but have recently published an empirical examination of this research question with control for observable characteristics related to neighborhood selection using one of the approaches detailed herein [24]. Table 1 provides the model specifications for all models discussed below and Table 2 provides accompanying Stata and SAS code to fit these models.

Baseline model of crude disparity

A general baseline estimate of unadjusted racial disparities without respect to context can be obtained through conventional ordinary least squares linear regression for continuous gestational age or maximum likelihood unconditional logistic regression for PTB. Model 1 of Table 1 shows the specification for the continuous outcome, where Y_{ij} is the gestational age in weeks for a birth to mother i in neighborhood j, $RACE_{ij}$ is the individual-level exposure of interest, a binary indicator for race (black = 1, white = 0), and the error terms for both cluster (ε_{2j}) and individual within cluster (ε_{1ij}) will be subsumed within the same undifferentiated term with mean zero and homoskedastic variance due to the single level specification. One can use cluster-robust SEs in single-level regression models to account for the nonindependence of observations within clusters [9]. The degree of neighborhood confounding depends on the magnitudes of associations between neighborhood and outcome, and between neighborhood and $RACE_{ij}$.

Accounting for context: obtaining the average within-neighborhood effect

FE models

Models with FE for neighborhoods can be considered a gold standard for complete control of neighborhood-level attributes when one is examining the effect of an individual-level exposure. In this approach, each neighborhood typically receives a fixed intercept or effect by adding k-1 indicator (i.e., "dummy") variables for the k neighborhoods [7–9,17,18,20]. This captures all neighborhood characteristics related to the outcome, whether observed or unobserved, that are distinct from race (Model 2). All inference on model parameters, such as the estimated coefficient for the race term, is within-neighborhood and purged of neighborhood-level confounding because there is no longer any residual between-neighborhood error that could be correlated with covariates (i.e., ε_{2j} in Model 1 is eliminated, having been replaced by the fixed neighborhood intercepts γ_j in Model 2a). Only neighborhoods with variation in the exposure variable (e.g., births to both black and white women), will inform the within-neighborhood FE estimate, and results will be generalizable to the population living in neighborhoods with births to both races (97% in our example). The FE estimate that removes confounding by neighborhood would not be identifiable in the case of complete segregation since there would be no within-neighborhood variation in race and therefore complete correlation between race and neighborhood [25].

Complete control for unmeasured confounding in observational data seems almost too good to be true, and indeed, there are potential costs and limitations to the FE approach. One is that the effects of factors at the neighborhood level cannot be estimated, because all neighborhood variability has been accounted for and comparisons are strictly within-neighborhood. By extension, variability across neighborhoods in the effect of a factor (i.e., slope variability or cross-level interactions) cannot be examined. Another limitation is that some inefficiency may result if there is little within-cluster variation in the exposure or if the number of clusters is large relative to the number of observations (i.e., small cluster size) because each additional term in the model requires a degree of freedom. However, there is no minimum requirement for the number of clusters or average cluster size beyond two [9].

In the continuous outcome case, an alternative approach to account for neighborhood FEs is to subtract the neighborhood mean from the outcome and covariates (Model 2b). This requires adjustment for the degrees of freedom needed to estimate each cluster mean, however, so ultimately the results are identical to the use of dummy variables [7]. In a nonlinear model, the dummy variable approach can also produce upwardly biased estimates when many clusters have few observations, known as the "incidental parameters problem" [7–9,15]. The use of conditional likelihoods as is done in matched case-control studies produces the within-cluster effect estimate without explicitly estimating a separate coefficient for each cluster. Nonetheless, efficiency may still be compromised because clusters with non-varying outcomes—all 0 or 1—will be dropped

Table 1
Model specifications assessing racial disparities in a continuous outcome

Model no.	Type	Equation	Inference and distinguishing features
1	Conventional single-level	$Y_{ij} = \beta_0 + \beta_1 Race_{ij} + \varepsilon_{1ij} + \varepsilon_{2j}$	Individual-level racial disparity, adjusted for any other covariates that may be in the model
2a	FE: dummy variable	$Y_{ij} = \beta_0 + \beta_1 \mathrm{Race}_{ij} + \sum_{j=1}^{k-1} \gamma_j \mathrm{Neighborhood}_j + \varepsilon_{ij}$	Within-neighborhood individual-level racial disparity (i.e., adjusted for neighborhood differences/segregation) Fixes/conditions on neighborhood so neighborhood-level covariates cannot be examined
2b	FE: conditional (“de-meaned”)	$Y_{ij} - \overline{Y}_{.j} = \beta_0 + \beta_1 (\mathrm{Race}_{ij} - \overline{\mathrm{Race}}_{.j}) + \varepsilon_{ij}$	Within-neighborhood individual-level racial disparity (i.e., adjusted for neighborhood differences/segregation) Fixes/conditions on neighborhood so neighborhood-level covariates cannot be examined
3a	RE: intercept	$Y_{ij} = \beta_{0j} + \beta_1 \mathrm{Race}_{ij} + \varepsilon_{ij}$ $\beta_{0j} = \gamma_{00} + \mu_{0j} \quad \mu_{0j} \sim \mathrm{N}(0, \tau_0^2)$	Individual-level racial disparity, generally confounded by neighborhood when segregation is present Possible to estimate neighborhood-level covariate effects Models variability in the outcome across neighborhoods
3b	RE: intercept + slope	$Y_{ij} = \beta_{0j} + \beta_{1j} \mathrm{Race}_{ij} + \varepsilon_{ij}$ $\beta_{0j} = \gamma_{00} + \mu_{0j} \; \begin{bmatrix} \mu_{0j} \\ \mu_{1j} \end{bmatrix} \sim \mathrm{N}(0, \Omega)$ $\beta_{1j} = \gamma_{10} + \mu_{1j} \; \Omega = \begin{bmatrix} \tau_0^2 & \\ \tau_{01} & \tau_1^2 \end{bmatrix}$	Individual-level racial disparity, generally confounded by neighborhood when segregation is present Possible to estimate neighborhood-level covariate effects Models variability in the outcome and racial disparity across neighborhoods (random intercept + slope)
4a	Hybrid FE*: cluster-mean centering	$Y_{ij} = \beta_{0j} + \beta_w (\mathrm{Race}_{ij} - \overline{\mathrm{Race}}_{.j}) + \varepsilon_{ij}$ $\beta_{0j} = \gamma_{00} + \mu_{0j} \quad \mu_{0j} \sim \mathrm{N}(0, \tau_0^2)$	Within-neighborhood individual-level racial disparity (i.e., adjusted for neighborhood differences/segregation), but other variables in the model will not be adjusted for neighborhood racial segregation Possible to estimate neighborhood-level covariate effects Models variability in the outcome across neighborhoods
4b	Hybrid FE*: cluster-mean centering + cluster-mean adjustment	$Y_{ij} = \beta_{0j} + \beta_w (\mathrm{Race}_{ij} - \overline{\mathrm{Race}}_{.j}) + \varepsilon_{ij}$ $\beta_{0j} = \gamma_{00} + \beta_b \overline{\mathrm{Race}}_{.j} + \mu_{0j} \quad \mu_{0j} \sim \mathrm{N}(0, \tau_0^2)$	Within-neighborhood individual-level racial disparity (i.e., adjusted for neighborhood differences/segregation) and the between-neighborhood ecological effect of race (context + composition) Possible to estimate neighborhood-level covariate effects Models variability in the outcome across neighborhoods
4c	Hybrid FE*: cluster-mean adjustment	$Y_{ij} = \beta_{0j} + \beta_w \mathrm{Race}_{ij} + \varepsilon_{ij}$ $\beta_{0j} = \gamma_{00} + (\beta_b - \beta_w) \overline{\mathrm{Race}}_{.j} + \mu_{0j} \quad \mu_{0j} \sim \mathrm{N}(0, \tau_0^2)$	Within-neighborhood individual-level racial disparity (i.e., adjusted for neighborhood differences/segregation) and the contextual effect of neighborhood racial composition Possible to estimate neighborhood-level covariate effects Models variability in the outcome across neighborhoods

b = between neighborhood; FE = fixed effects; i = individual; j = neighborhood; OLS = ordinary least squares; RE, random effects; w = within neighborhood.
* Shown in a random effects (intercept) specification but marginal models are also applicable.

from the analysis and only those clusters with discordant exposures *and* outcomes will inform the effect estimates [16].

RE models

The RE model can be considered a reduced form of the FE model that requires more assumptions and may trade bias for efficiency [7]. RE models provide neighborhood-specific or within-neighborhood inference by incorporating a random normal distribution of neighborhood intercepts $\mu_{0j} \sim \mathrm{N}(0, \tau_0^2)$ around a cluster mean (γ_{00}) so the black–white contrast (β_1) is conditioned on a random neighborhood intercept (Model 3a) [9,26,27]. However, neighborhood RE models only account for the association between cluster and outcome and not the association between exposure and cluster as in the neighborhood FE specification. Thus, a valid estimation of the exposure effect is dependent on the assumption that the exposure is not conditionally correlated with the neighborhood random intercepts, i.e., that *cov* (RACE_{ij}, μ_{0j}) = 0. This assumption is necessarily violated when there is contextual confounding of race and neighborhood as the result of residential segregation. Yet, because the neighborhood variability is not explicitly removed, RE models carry several important benefits, including the ability to examine the effects of neighborhood factors, partition the variance at individual and neighborhood levels, and also evaluate random slopes or neighborhood variability in the coefficients of individual-level variables (e.g., whether racial disparities vary across neighborhoods, Model 3b).

Hybrid fixed effects models

An alternative method of removing the neighborhood-level confounding of the race effect within a RE or marginal model is to incorporate the between-cluster variation associated with race, which can be captured by the neighborhood mean of race ($\overline{\mathrm{Race}}_{.j}$ or % black at neighborhood level) [9,12–16,28,29]. A convenient property of the cluster-mean exposure is that it accounts for the between-cluster variation in exposure that produces cluster-level confounding. When this is controlled, the level 1 exposure parameter reflects the within-cluster effect that is purged of all cluster-level variation. In our example, we control for the racial composition between neighborhoods to estimate the race effect within-neighborhood that is unbiased by any factor jointly associated with race and neighborhood. The estimated coefficient represents the contrast between black and white residents, holding the level of segregation (% black) constant. Note that the cluster mean should refer to the mean of the variable in a given sample and not the mean from an external source (e.g., Census data) as the degree of cluster-level confounding is dependent on the imbalance in the dataset at hand. Although the mean is the most common function chosen in the applied literature, it has been noted that the neighborhood-level exposure could conceivably be related to the outcome through some other function, which if known could be superior to the mean [30]. In practice, however, the true data generating function is seldom known with such specificity, and the

Table 2
Stata and SAS code for analytic models

Model no.	Type	Stata code	SAS code	
1	Conventional single-level with cluster-robust SEs	reg y_var race, vce(cl cluster_var) logit y_var race, vce(cl cluster_var)	proc surveyreg or surveylogistic; cluster cluster_var; model y_var = race; run;	
2a	FE: dummy variable	reg y_var race i.cluster_var logit y_var race i.cluster_var	proc surveyreg or surveylogistic; cluster cluster_var; class cluster_var; model y_var = race cluster_var; run;	
2b	FE: conditional ("de-meaned")	xtreg y_var race, i(cluster_var) fe xtlogit y_var race, i(cluster_var) fe	proc glm; absorb cluster_var; model y_var = race; run;	proc logistic; strata cluster_var; model y_var = race; run;

3a	RE: intercept	xtreg y_var race, i(cluster_var) mle
		xlogit y_var race, i(cluster_var) re
			proc glimmix method=quad;
			class cluster_var;
			model y_var = race/solution dist=(nor or bin)
			link=(id or logit);
			random intercept/subject=cluster_var;
			run;	

3b	RE: intercept + slope	xtmixed y_var race || cluster_var: race, mle
		xtmelogit y_var race || cluster_var: race
			proc glimmix method=quad;
			class cluster_var;
			model y_var = race/solution dist=(nor or bin)
			link=(id or logit);
			random intercept race/subject=cluster_var;
			run;	

4a	Hybrid FE: cluster-mean centering	egen race_cluster=mean(race), by(cluster_var)
		gen race_c=race-race_cluster
		xtreg y_var race_c, i(cluster_var) mle
		xtlogit y_var race_c, i(cluster_var)
			proc glimmix method=quad;
			class cluster_var;
			model y_var = race_c/solution dist=(nor or bin)
			link=(id or logit);
			random intercept/subject=cluster_var;
			run;	

4b	Hybrid FE: cluster-mean centering + cluster-mean adjustment	egen race_cluster=mean(race), by(cluster_var)
		gen race_c=race-race_cluster
		xtreg y_var race_c race_cluster, i(cluster_var) mle
		xtlogit y_var race_c race_cluster, i(cluster_var)
			proc glimmix method=quad;
			class cluster_var;
			model y_var = race_c race_cluster/solution
			dist=(nor or bin) link=(id or logit);
			random intercept/subject=cluster_var;
			run;	

4c	Hybrid FE: cluster-mean adjustment	egen race_cluster=mean(race), by(cluster_var)
		xtreg y_var race race_cluster, i(cluster_var) mle
		xtlogit y_var race race_cluster, i(cluster_var)
			proc glimmix method=quad;
			class cluster_var;
			model y_var = race race_cluster/solution
			dist=(nor or bin) link=(id or logit);
			random intercept/subject=cluster_var;
			run;	

FE = fixed effects; RE = random effects.

mean is therefore considered to be a sensible and robust option for accounting for between-cluster variation related to an exposure as it derives from the de-meaned FE approach (Model 2b).

Models that incorporate a cluster-mean exposure variable have been termed "hybrid fixed effect models" because they can achieve the FE effect estimates in a RE or marginal specification without removing all between-cluster variation as is done in FE by entering dummy variables, cluster-mean centering the outcome and covariates, or using conditional likelihood [7]. As such, they are more flexible in terms of including other cluster-level covariates, as well as estimating variance components and allowing random slopes in the RE specification.

There are several ways to incorporate the cluster mean into a model. First, a covariate of interest can simply be cluster-mean centered (Model 4a, $\text{Race}_{ij} - \overline{\text{Race}}_{.j}$). However, without explicitly incorporating the cluster-mean as a parameter, this approach will not adjust other covariates (level 1 or level 2) for any confounding attributable to residential segregation. Including the cluster-mean variable itself as an additional covariate will provide estimates of both within and between cluster effects of the variable (Model 4b, β_w, β_b). When the individual-level (level 1) covariate is cluster-mean centered, the estimated cluster-mean (level 2) parameter will correspond to the total between-cluster (ecological or aggregate) effect (e.g., the difference between neighborhoods with more vs fewer black residents). Thus, this approach can be used to evaluate ecological and atomistic fallacies [31,32], which would be committed if the between and within cluster effects are not equal and both the individual and aggregate variables have not been included. By cluster-mean centering the individual-level covariate, however, the association between cluster and race is removed, so the cluster-level variable has not been properly adjusted for compositional effects.

When the equation is rewritten without cluster-mean centering, the estimated cluster-mean coefficient will represent the difference in the between- and within-cluster effects or a true contextual effect (Model 4c, e.g., the effect of living in a more segregated neighborhood, regardless of whether you are black or white, $\beta_c = \beta_b - \beta_w$). When $\beta_b - \beta_w = 0$, there is no possibility of cluster-level confounding bias in the original random intercept specification [33]. Thus, a test of the cluster-mean covariate in Equation 4c can be considered a test for bias or at least a test for violations of the RE assumption that the between and within effects are equivalent [11]. Note that when multiple level-1 confounders are examined they may all require cluster-mean adjustment or centering to produce an exposure effect estimate equivalent to that obtained in a FE model. In other words, valid within-cluster effects of covariates also need to be estimated if they are confounders of the exposure.

The cluster-mean adjusted approach can also be conducted in a marginal model such as GEE or even conventional regression models with cluster-robust SEs, if variance components are not of interest and marginal inference in nonlinear models is desired. A GEE model with exchangeable correlation structure will produce effect estimates consistent with a RE model for linear model forms [34]. In the logistic case, the GEE model will generally produce coefficients with smaller absolute magnitude than a RE model as a result of the noncollapsibility of odds ratios across neighborhoods [13,15,34]. Noncollapsibility of some measures such as the odds ratio refers to the nonequivalence between the average neighborhood-specific exposure effect versus the average individual exposure effect across total population [35,36]. The magnitude of the discrepancy depends on the variability in the outcome across clusters; if there is a large degree of variability or high intracluster correlation (ICC), the coefficients from GEE models will be substantially smaller than those from RE models. Some authors have argued in favor of the GEE model given that the RE specification requires unverifiable assumptions on the distribution of an unobserved variable [36]. With cross-sectional data, the cluster-specific interpretation of between-cluster effect estimates in random effects models also relies on nonidentifiable comparisons because changes in a contextual variable within a given cluster have not been measured.

Evaluating results of various approaches with an applied example

In this data example from two North Carolina counties, all singleton births to non-Hispanic black and white women were evaluated and only those with geocodable addresses and complete birth weight, gestational age, and demographic information (96.3%) were included in analyses (n = 31,489). There were a total of 390 neighborhoods represented with an average of ~80 births per neighborhood. Only 9.5% of the neighborhoods were completely segregated with only black or white births, corresponding to 3.4% of the study population. More than 85% of the population lived in neighborhoods with at least five births to women of each race, which suggests some generalizable support for within-neighborhood inference.

The degree of residential segregation or clustering of race within neighborhood is moderate with an ICC of 0.56. This implies roughly equal variation in race between as within neighborhoods. Thus, if an adverse neighborhood characteristic is associated with segregation, there will be positive confounding or an inflation of disparities attributable to contextual factors.

The results of unadjusted single-level models indicated that Black women deliver infants an average of 4.2 days earlier than white women and the odds of delivering before 37 weeks are about twice that of white women (Tables 3 and 4). The results of RE models, which account for unobserved heterogeneity in the outcome across neighborhoods by assuming no correlation between race and the random intercepts, are virtually identical to conventional regression models. This could indicate one of two things: 1) there is little contextual confounding or 2) there is contextual confounding present that violated the assumption of independence between race and the RE. Because the gold standard FE results are substantially different from the conventional and RE models, there is clear evidence of contextual confounding. For continuous gestational age, the FE race coefficient is approximately 22% lower than the conventional or RE race coefficient. As expected, there is no difference between the dummy variable and the de-meaned approaches (subtraction of neighborhood means for predictor and outcome). For PTB, the FE odds ratio (OR) for race is approximately 40% lower than the OR from conventional or RE models (i.e., (2.0 − 1.6)/(2.0 − 1.0) = 0.4)). The dummy variable

Table 3
Model comparisons for gestational age in weeks

	N	β (Race) (SE)	βw† (SE)	βb‡ (SE)	βc§ (SE)	σ	τ_0	τ_1
Single-level Unadjusted*	31489	−0.61 (0.03)				2.14		
Random intercept	31489		−0.59 (0.03)			2.13	0.14	
Random slope	31489		−0.61 (0.03)			2.13	0.02	0.34
GEE (exchangeable)	31489	−0.59 (0.03)						
Fixed effects								
Dummy variable	31489		−0.47 (0.03)					
Conditional (“demeaned”)	31489		−0.47 (0.03)			2.13	0.40	
Hybrid fixed effects								
Cluster-mean centering								
Random intercept	31489		−0.47 (0.03)			2.13	0.29	
Random slope	31489		−0.37 (0.05)			2.13	0.09	0.54
GEE (exchangeable)	31489		−0.47 (0.03)					
Conventional OLS regression*	31489		−0.47 (0.04)					
Cluster-mean centering + cluster-mean adjustment								
Random intercept	31489		−0.47 (0.03)	−0.85 (0.05)		2.13	0.12	
Random slope	31489		−0.50 (0.04)	−0.82 (0.05)		2.13	0.03	0.30
GEE (exchangeable)	31489		−0.47 (0.03)	−0.84 (0.05)				
Conventional OLS regression*	31489		−0.47 (0.04)	−0.83 (0.06)				
Cluster-mean adjustment								
Random intercept	31489		−0.47 (0.03)		−0.38 (0.06)	2.13	0.12	
Random slope	31489		−0.50 (0.04)		−0.31 (0.06)	2.13	0.03	0.30
GEE (exchangeable)	31489		−0.47 (0.03)		−0.37 (0.06)			
Conventional OLS regression*	31489		−0.47 (0.04)		−0.36 (0.07)			

B = between; c = contextual; GEE = generalized estimating equation; OLS = ordinary least squares; σ = standard deviation of individual-level error; SE = standard error; τ_0 = standard deviation of the random intercepts; τ_1 = standard deviation of the random slopes; w = within.
* Cluster-robust standard errors.
† βw refers to cluster-specific or within-neighborhood effect of race.
‡ βb refers to the between-neighborhood or ecological effect of race (cluster-mean coefficient); βb = βw + βc.
§ βc refers to the contextual effect of race (cluster-mean coefficient); βc = βb − βw.

Table 4
Model comparisons for PTB

	N	OR (Race) (SE)	ORw† (SE)	ORb‡ (SE)	ORc§ (SE)	τ_0	τ_1
Single-level unadjusted*	31,489	2.03 (0.09)					
Random intercept	31,489		1.99 (0.08)			0.13	
Random slope	31,489		1.99 (0.09)			0.11	0.12
GEE (exchangeable)	31,489	2.00 (0.08)					
Fixed effects							
Dummy variable	31,157		1.61 (0.08)				
Conditional	31,157		1.60 (0.08)				
Hybrid fixed effects							
Cluster-mean centering							
Random intercept	31,489		1.61 (0.08)			0.33	
Random slope	31,489		1.52 (0.10)			0.23	0.39
GEE (exchangeable)	31,489		1.59 (0.08)				
Conventional logit*	31,489		1.61 (0.08)				
Cluster-mean centering + cluster-mean adjustment							
Random intercept	31,489		1.60 (0.08)	2.87 (0.18)		0.08	
Random slope	31,489		1.60 (0.08)	2.87 (0.18)		0.08	0.00
GEE (exchangeable)	31,489		1.60 (0.08)	2.86 (0.18)			
Conventional logit*	31,489		1.60 (0.08)	2.86 (0.19)			
Cluster-mean adjustment							
Random intercept	31,489		1.60 (0.08)		1.79 (0.15)	0.08	
Random slope	31,489		1.60 (0.08)		1.79 (0.15)	0.08	0.00
GEE (exchangeable)	31,489		1.60 (0.08)		1.79 (0.15)		
Conventional logit*	31,489		1.60 (0.08)		1.78 (0.15)		

b = between; c = contextual; GEE = generalized estimating equation; OR = odds ratio; PTB = preterm birth; SE = standard error; w = within; τ_0 = standard deviation of the random intercepts; τ_1 = standard deviation of the random slopes.
* Cluster-robust standard errors.
† ORw refers to cluster-specific or within-neighborhood OR for race.
‡ ORb refers to the between-neighborhood or ecological OR for race (cluster-mean OR); ORb = $e^{(\beta w+\beta c)}$.
§ ORc refers to the contextual OR for race (cluster-mean OR); ORc = $e^{(\beta b+\beta w)}$.

approach is slightly biased upward due to the "incidental parameters problem." Also, 332 observations were dropped due to uniform outcomes (all preterm or all term) within neighborhoods, so odds ratios could not be estimated due to division by zero.

All three hybrid FE options (centering, centering with cluster-mean adjustment, cluster-mean adjustment) yielded disparity point estimates that were consistent with the conventional FE results. This is because they account for the correlation between race and the neighborhood intercept by incorporating the neighborhood mean of race either before or during model estimation. When the FE is obtained through cluster-mean centering, the level-2 variance increases compared to the original random intercept because it is not appropriately adjusted for differences in racial composition across neighborhoods—the variability across neighborhoods attributable to race was removed from the race variable so that the compositional portion of level-2 variance could not be controlled (Model 4a). Thus, the variance components from this model are identical to a random intercept model with no predictors. Adding the cluster mean accounts for neighborhood variability in the RE that is related to racial composition but the level-2 or between-neighborhood effect of residential segregation is still not adjusted for the level-1 effect of race since the neighborhood mean was subtracted from individual-level race prior to estimation (Model 4b). Therefore, the between-neighborhood effect will be purely ecological, capturing both individual and contextual effects. Because the estimated effects are substantially greater than the within-neighborhood or individual-level race effects, there is evidence that the between and within-neighborhood effects are not the same. The difference is the true contextual effect, which is obtained automatically when there is only cluster-mean adjustment and no level 1 cluster-mean centering (e.g., continuous gestational age: $-0.83 - (-0.47) = 0.36$, Model 4c). Because the contextual effect of residential segregation is not zero, there is evidence of contextual confounding of the individual-level race effect. In fact, the odds of PTB were increased by a similar amount for living in a completely black versus completely white neighborhood (OR 1.79) as for being black versus white as an individual characteristic (OR 1.60).

The neighborhood intercept variance is much greater for PTB than for gestational age. However, the random slope or neighborhood variation in the effect of race is much greater for gestational age. The addition of a random slope changes model parameters slightly so point estimates for the racial disparity are somewhat different from random intercept-only models. Point estimates for the average cluster-specific intercept and slope may change depending on the assumed covariance structure between the random intercepts and slopes, which we assumed to be zero (default in Stata and SAS). The marginal or population-average models (GEE or conventional regression with cluster robust variance), also provide identical neighborhood FE estimates of racial disparities in these data where the ICC is quite low ($<0.01\%$). The SEs were not meaningfully different between RE (hybrid) and conventional FE models, so efficiency may not be an important consideration when there is sufficient within-cluster variation and many observations per cluster ($\sim$80 in this example). Among hybrid FE models, there was also little difference in SEs across RE, GEE, and conventional models with cluster-robust SEs.

Summary and recommendations

In this North Carolina data example, as in most of the United States, there was considerable residential segregation (neighborhood clustering by race) that contributed to substantial confounding of the impact of race on gestational age at delivery and PTB. This positive correlation between race and neighborhood-based health risk is a clear violation of the RE model assumption of no correlation between predictors and random intercepts, which leads to upward bias in multilevel or RE models. This bias was corrected by FE or hybrid FE approaches—both of which accounted for all neighborhood variation in both the outcome and exposure. Although neighborhood variation in the outcomes was relatively small ($ICC < 0.01\%$), neighborhoods still explained a substantial portion of the racial disparities. This finding supports previous warnings that large neighborhood effects are not contingent on a high level of neighborhood variation or ICC [20,37].

RE models produce effect estimates that are a weighted average of the within-cluster and between-cluster effects. When there are differences in within and between cluster effects (i.e., contextual confounding), the degree of bias in RE models is dependent on the cluster size and the between- and within-cluster variance of both outcome and exposure [9,16,38]. Even in the presence of contextual confounding, the RE estimate will be weighted toward the unconfounded within-cluster effect when 1) the cluster size is large, 2) the within-cluster variance of the outcome is small relative to the between-cluster variance (i.e., high ICC), or 3) the between-cluster variance in exposure is low. In our example and most neighborhood and health studies, ICC's are generally modest (<10%) and there is often substantial clustering of exposures so that only an extremely large cluster size could compensate to provide a valid within-neighborhood effect. Because it is generally not known *a priori* whether a particular sample size is adequate to compensate for cluster-level confounding under a given ICC, the conventional RE model should always be compared to a FE or hybrid FE specification. We were able to locate only two examples in the literature where the RE model approximated the FE results in relation to health and social disparities and in both cases the average cluster size was quite large (10,000 to 2 million)[39,40].

Our results show that estimates comparable to conventional FE can be obtained through various hybrid strategies (cluster-mean centering and/or adjustment) and in various model types (RE, GEE, single-level regression). Compared with conventional FE, the hybrid approach allows estimation of neighborhood-level effects and may therefore be preferable when this is a goal. Among the models that implement the hybrid approach, the RE specification offers the most flexibility—capable of also estimating variance components and random slopes with a cluster-specific or conditional interpretation. GEE or cluster-adjusted conventional models are also attractive options if the marginal interpretation is satisfactory and variance components are not of interest. The marginal and conditional results are similar when the ICC is low as in our example. Among the hybrid approaches, we recommend the cluster-mean adjusted method (without centering) because it provides level 1 and level 2 effects adjusted for one another and valid variance components. This approach also adjusts other covariates for the neighborhood differences related to segregation. However, the addition of other level-2 covariates may be compromised with cluster-mean adjustment since other neighborhood covariates, like SES, may be too strongly correlated with racial composition to estimate unique effects [41]. This may necessitate separate models for each neighborhood covariate [24].

We also recommend comparing hybrid FE estimates to conventional FE models to ensure appropriate specification and valid estimates since it is not fail-proof; one or all of the variables in the model may require cluster-mean adjustment or centering and the cluster-mean must be derived from the sample, not from external estimates. Rabe-Hesketh et al. [9] advocate an approach of entering cluster-mean terms for all level-1 covariates in which the within and between-effects differ significantly at the $\alpha\%$ level, to ensure that level-1 estimates are comparable to FE estimates. It should be noted that there may be bias in level 1 effect estimates in both FE and hybrid FE models if important individual-level confounders are omitted (i.e., individual-level residuals are correlated with covariates) [42]. Similarly, there can be bias in specific contextual effect estimates in hybrid FE, RE, and GEE models if important covariates at either level are omitted.

Limitations and extensions

In our relatively simple example, we did not adjust for individual-level confounders that influence self-selection into neighborhoods, so our estimates of the neighborhood contribution to disparities (22% for gestational age and 40% for PTB) are upper bounds. Although it is difficult to account for the complex reciprocity between individual SES and neighborhoods in cross-sectional analyses, we attempted some adjustment for individual characteristics in a previously published empirical analysis [24]. We found that neighborhood SES was insufficient to account for all neighborhood differences related to segregation, suggesting the utility of the FE approach.

Although it is important to quantify the total contribution of neighborhoods to social disparities by race or class, this should not replace or deter efforts to measure and identify the actual neighborhood features that could be modified by interventions to reduce disparities and improve health. In fact, it would be most informative to compare specific neighborhood factors (e.g., crime, pollution, health care access/quality, etc) to the total neighborhood contribution to assess the relative priority for intervention to reduce disparities in a particular outcome.

We caution that comparable results among the different types of FE and hybrid FE models may not necessarily apply in other data sets with different ICCs and cluster sizes. For example, there may be more substantial gains in efficiency from a RE or hybrid FE model in cases with a large number of small clusters. And although the comparability of FE and hybrid FE results is generally robust, marginal logistic regression models yield effect estimates closer to the null with a discrepancy between marginal and cluster-specific estimates that rises with the ICC. However, predicted odds can readily be transformed to predicted probabilities, which can then be contrasted to form risk differences and risk ratios—collapsible measures that won't differ between marginal and cluster-specific specifications [43,44]. Working with contrasts of probabilities rather than contrasts of odds is especially advantageous in the situation of a common outcome in which the odds ratio overestimates the relative risk [45].

For survey-based data, there is not yet a consensus with respect to the handling of probability weights in multilevel models [46–48]. Some recent work examining logistic models of complex survey data concluded that the fully FE approach in a weighted ordinary logistic regression with cluster dummies or conditional logistic regression with scaled weights may be superior in cases of strong or moderate sampling bias [49,50]. The hybrid FE method in a RE model with original or scaled weights performed well only under the condition of weak or no sampling bias. Sampling bias can be evaluated by comparing weighted and unweighted estimates [48].

Finally, there is increasing awareness that we need to account for contextual inequalities throughout the life-course to fully explain racial disparities. These models would have to be extended to incorporate neighborhood-level FE with duration and/or timing of exposure controlled with interaction terms. To incorporate cluster by time/age interactions conventional FE models may be simpler, but future methodological developments could allow for cluster means to be calculated at particular time/age points or averaged across multiple time/age points in a hybrid approach.

Conclusions

In summary, we note that conventional and RE models can yield substantially biased effect estimates in the presence of confounding by neighborhood. FE and hybrid FE models can guarantee resolution of cluster level confounding, with the latter having greater flexibility in estimating level 2 effects and intercept/slope variability. Given the range of models that can be implemented in a hybrid approach (conventional, GEE, RE) and the frequent goal of accounting for contextual confounding, this approach should be used more often.

References

[1] LaVeist T, Thorpe Jr R, Bowen-Reid T, Jackson J, Gary T, Gaskin D, et al. Exploring health disparities in integrated communities: overview of the EHDIC study. J Urban Health 2008;85:11–21.
[2] Do DP, Frank R, Finch BK. Does SES explain more of the black/white health gap than we thought? Revisiting our approach toward understanding racial disparities in health. Soc Sci Med 2012;74:1385–93.
[3] Oakes JM, Johnson PJ. Propensity score matching for social epidemiology. In: Oakes JM, Kaufman JS, editors. Methods in social epidemiology. San Francisco, CA: Jossey-Bass; 2006. p. 370–92.
[4] Landrine H, Corral I. Separate and unequal: residential segregation and black health disparities. Ethn Dis 2009;19:179–84.
[5] Massey DS, Fischer MJ. Does rising income bring integration? New results for blacks, Hispanics, and Asians in 1990. Soc Sci Res 1999;28:316–26.
[6] Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Rep 2001;116:404–16.
[7] Allison PD. Fixed effects regression models. Los Angeles: SAGE; 2009.
[8] Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press; 2002.
[9] Rabe-Hesketh S, Skrondal A, Skrondal A. Multilevel and longitudinal modeling using Stata. College Station, TX: Stata Press Publication; 2008.
[10] Allison PD. A hybrid method. Fixed effects regression models. Los Angeles: SAGE; 2009. p. 23–25, 39–42.
[11] Rabe-Hesketh S, Skrondal A, Skrondal A. Between and within effects. Multilevel and longitudinal modeling using stata. College Station, TX: Stata Press Publication; 2008. p. 109–24.
[12] Begg MD, Parides MK. Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. Stat Med 2003;22:2591–602.
[13] Berlin JA, Kimmel SE, Ten Have TR, Sammel MD. An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angioplasty. Biometrics 1999;55:470–6.
[14] Desai M, Begg MD. A comparison of regression approaches for analyzing clustered data. Am J Public Health 2008;98:1425–9.
[15] Localio AR, Berlin JA, Ten Have TR. Confounding due to cluster in multicenter studies—causes and cures. Health Serv Outcomes Res Methodol 2002;3:195–210.
[16] Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. Biometrics 1998;54:638–45.
[17] Do DP, Finch BK, Basurto-Davila R, Bird C, Escarce J, Lurie N. Does place explain racial health disparities? Quantifying the contribution of residential context to the Black/white health gap in the United States. Soc Sci Med 2008;67:1258–68.
[18] Sastry N, Hussey JM. An investigation of racial and ethnic disparities in birth weight in Chicago neighborhoods. Demography 2003;40:701–25.
[19] Yang D, Howard G, Coffey CS, Roseman J. The confounding of race and geography: how much of the excess stroke mortality among African Americans is explained by geography? Neuroepidemiology 2004;23:118–22.
[20] Bingenheimer JB, Raudenbush SW. Statistical and substantive inferences in public health: issues in the application of multilevel models. Annu Rev Public Health 2004;25:53–77.
[21] Morenoff JD, House JS, Hansen BB, Williams DR, Kaplan GA, Hunte HE. Understanding social disparities in hypertension prevalence, awareness, treatment, and control: the role of neighborhood context. Soc Sci Med 2007;65:1853–66.
[22] Brumback BA, Dailey AB, Zheng HW. Adjusting for confounding by neighborhood using a proportional odds model and complex survey data. Am J Epidemiol 2012 Jun 1;175(11):1133–41.
[23] Clampet-Lundquist S, Massey DS. Neighborhood effects on economic self-sufficiency: a reconsideration of the moving to opportunity experiment. Am J Sociol 2008;114:107–43.
[24] Schempf AH, Kaufman JS, Messer LC, Mendola P. The neighborhood contribution to black-white perinatal disparities: an example from two North Carolina counties, 1999–2001. Am J Epidemiol 2011;174:744–52.
[25] Westreich D, Cole SR. Invited commentary: positivity in practice. Am J Epidemiol 2010;171:674–7. discussion 678–81.
[26] Raudenbush SW, Bryk AS. applications and data analysis methods, Hierarchical linear models. Thousand Oaks, CA: Sage Publications; 2002.
[27] Snijders TAB, Bosker RJ. Multilevel analysis: an introduction to basic and advanced multilevel modeling. Thousand Oaks, CA: Sage Publications; 1999.
[28] Mundlack Y. On the pooling of time series and cross section data. Econometrica 1978;46:69–85.
[29] Pan W. A note on the use of marginal likelihood and conditional likelihood in analyzing clustered data. Am Stat 2002;56:171–4.
[30] Brumback BA, Dailey AB, Brumback LC, Livingston MD, He Z. Adjusting for confounding by cluster using generalized linear mixed models. Stat Probability Lett 2010;80:1650–4.
[31] Diez Roux AV. A glossary for multilevel analysis. J Epidemiol Community Health 2002;56:588–94.
[32] Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. Soc Sci Med 1998;46:97–117.
[33] Firebaugh GA. Rule for inferring individual-level relationships from aggregate data. Am Sociol Rev 1978;43:557–72.
[34] Diggle P, Liang K-Y, Zeger SL. Analysis of longitudinal data. Oxford: Oxford University Press; 1994.
[35] Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? Stat Med 2009;28:221–39.
[36] Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Lippman SA, Jewell N, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology 2010;21:467–74.
[37] Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. J Epidemiol Community Health 2003;57:550–2.
[38] Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. J Roy Stat Soc Ser B 2006;68:859–72.
[39] Scribner RA, Theall KP, Simonsen NR, Mason KE, Yu Q. Misspecification of the effect of race in fixed effects models of health inequalities. Soc Sci Med 2009; 69:1584–91.
[40] Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: the perils of individualistic and ecologic fallacy. Int J Epidemiol 2009;38:342–60. author reply 370–3.
[41] Messer LC, Oakes JM, Mason S. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. Am J Epidemiol 2010;171:664–73.
[42] Firebaugh G. Commentary: 'Is the social world flat? W.S. Robinson and the ecologic fallacy'. Int J Epidemiol 2009;38:368–70. author reply 370–3.
[43] Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. J Clin Epidemiol 2007;60:874–82.
[44] Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. J Clin Epidemiol 2010;63:2–6.
[45] Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? BMJ 1998; 316:989–91.
[46] Carle AC. Fitting multilevel models in complex survey data with design weights: Recommendations. BMC Med Res Methodol 2009;9:49.
[47] Rabe-Hesketh S, Skrondal A. Multilevel modelling of complex survey data. J R Stat Soc A 2006;169(Part 4):805–27.
[48] Asparouhov T. General multi-level modeling with sampling weights. Commun Stat 2006;35:439–60.
[49] Brumback BA, Dailey AB, He Z, Brumback LC, Livingston MD. Efforts to adjust for confounding by neighborhood using complex survey data. Stat Med 2010; 29:1890–9.
[50] Brumback BA, He Z. Adjusting for confounding by neighborhood using complex survey data. Stat Med 2011;30:965–72.