Published in final edited form as:

Soc Sci Med. 2009 December ; 69(11): 1584–1591. doi:10.1016/j.socscimed.2009.08.010.

Misspecification of the Effect of Race in Fixed Effects Models of Health Inequalities

Richard Allen Scribner, Katherine P Theall, PhD, Neal R Simonsen, PhD, Karen E Mason, MSPH, and Qingzhao Yu, PhD

Louisiana State University Health Sciences Center, School of Public Health, New Orleans, LA
UNITED STATES

Abstract

The purpose of this study is to characterize the different results obtained when analyzing health inequalities data in which individuals are nested within their neighborhoods and a single level model is used to characterize risk rather than a multilevel model. The inability of single level models to characterize between neighborhood variance in risk may affect the level of risk attributed to black race if blacks are differentially distributed in high risk neighborhoods. The research replicates in Los Angeles an approach applied by a different group of researchers in Massachusetts (Subramanian, S. V., Chen, J. T., Rehkopf, D. H., Waterman, P. D., & Krieger, N. 2005). Single level and multilevel models were used to analyze Los Angeles County, California, US. all-cause mortality data for the years 1989-1991, modeled as 29,936 cells (deaths and population denominators cross-tabulated by age, gender, and race/ethnicity) nested within 1,552 census tracts. Overall blacks had 1.27 times the risk of mortality compared to whites. However, multilevel models demonstrated considerable between census tract variance in mortality for both blacks and whites which was partially explained by neighborhood poverty. Comparing the results of equivalent single level and multilevel models, the mortality odds ratio for blacks compared to the white reference group reversed itself, indicating greater risk for blacks in the single level model and lower risk in the multilevel model. Adding an area based socioeconomic measure (ABSM) to the single level model reduced but did not remove the discrepancy. Predictions of mortality risk for the interaction of race and age group demonstrate that all single level models exaggerated the mortality risk associated with black race. We conclude that characterizing health inequalities in mortality for blacks using single level models, which do not account for the cross level interaction created by the greater likelihood of black residence in neighborhoods where the risk of mortality is greater regardless of race, can exaggerate the risk of mortality attributable to the individual level effects of black race.

Keywords

mortality; Race/ethnicity; social class; health disparities; public health; neighborhoods; USA

© 2009 Elsevier Ltd. All rights reserved.

Corresponding Author: Richard Allen Scribner, Louisiana State Univer Health Sciences Center New Orleans, LA UNITED STATES rscrib@lsuhsc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

BACKGROUND

The absence of a generally accepted measure of social class to monitor health inequalities, similar to what is available in Europe, has led to a situation in the US whereby health inequalities are monitored in terms of measures of race/ethnicity without regard to class (Krieger, Chen, Waterman, Rehkopf, & Subramanian, 2003a; Krieger, Williams, & Moss, 1997). According to a number of researchers this situation in the US has led to bias, whereby race-based explanations (e.g., genetic vulnerability) and approaches (e.g., equal access to health care services) to health inequalities are favored over class-based explanations (e.g., social disadvantage) and approaches (e.g., public policy to address social disadvantage) (Braveman, Egerter, Cubbin, & Marchi, 2004; Kawachi, Daniels, & Robinson, 2005; Krieger, Chen, Waterman, Rehkopf, & Subramanian, 2005). As a result, there have been significant efforts to develop measures of social class in the US to monitor health inequalities. The largest of these efforts is the Public Health Disparities Geocoding Project led by Krieger and colleagues (Krieger et al., 2003a, 2005; Krieger, Chen, Waterman, Soobader, Subramanian, & Carson, 2002). The project advocates the use of area based socioeconomic measures (ABSMs) at the census tract level as a substitute for an individual level measure of social class (Krieger et al., 2002).

Since the introduction of ABSMs numerous studies have incorporated them as a proxy for social class in monitoring health inequalities without much concern regarding the implications of analyzing data that has a multilevel structure (e.g., individuals nested within neighborhoods), using single level models rather than multilevel models. Single level models using ABSMs include covariates estimating individual and area level attributes, but the data are fit with a single intercept and slope. Multilevel models, on the other hand, typically refer to models with (at a minimum) a random intercept and may include random slopes (Raudenbush & Bryk, 2002). A model with a random intercept allows each area level unit (e.g., neighborhood) to have its own intercept which corresponds to the mean level of the outcome in that neighborhood. Therefore the coefficient for the random intercept captures the variation in the outcome that is due to differences in the mean or reference level of the outcome across neighborhoods. A multilevel model can include random coefficients for each of the covariates in the model as well. Since the coefficient for each covariate is an estimate of the relation between the covariate and the outcome, these are referred to as random slope models (Raudenbush & Bryk, 2002). A multilevel model where a covariate is assigned a random slope indicates that the relation between that covariate and the outcome will be estimated for each of the level two units (e.g., neighborhoods). Therefore there is a fixed effect corresponding to the mean of all the slopes across the level two units and a random effect corresponding to the variance of all the level two slopes around the mean slope.

One common type of study used to estimate health inequalities uses national survey data in which respondents' residence information is georeferenced to link with US census data to generate a set of ABSMs. This is known as a single level study with level two covariates. For example, numerous studies have analyzed the National Health Interview Survey data using ABSMs generated by georeferencing the census tract of residence for participants to characterize their social class and examine its potential effect over and above the individual level effects of income and education (Cubbin, LeClere, & Smith, 2000; LeClere, Rogers, & Peters, 1998; LeClere, Rogers, & Peters, 1997; Winkleby & Cubbin, 2003). Despite the multilevel data structure, these studies typically use single level models (i.e., single intercept and single slope) that include census tract level covariates as ABSMs (Bond Huie, Hummer, & Rogers, 2002; Cubbin et al., 2000; LeClere et al., 1998; LeClere et al., 1997; Winkleby & Cubbin, 2003). Presumably they are analyzed this way because there are too few individuals nested within the census tracts to permit multilevel modeling.

Although Krieger and colleagues maintain that the use of an ABSM in individual level or ecologic models (i.e., single level models) is not compromised by the fact that it is a multilevel construct, there are potential problems with the interpretation of the covariates in these studies (Krieger, Chen, Waterman, Soobader, Subramanian, & Carson, 2003b). Race/ethnicity is especially problematic due to the differential distribution of the population into census tracts according to race/ethnicity. Moreover, when there is a differential grouping of individuals into census tracts according to race/ethnicity and that grouping independently exposes individuals of that race/ethnicity to an increased risk due to contextual risk factors related to social disadvantage, the model is misspecified (T. Laveist, Thorpe, Bowen-Reid, Jackson, Gary, Gaskin et al., 2007). For example, suppose there is no health inequality between blacks and whites due to individual effects but a large inequality due to the greater likelihood that blacks tend to live in socially disadvantaged census tracts. In this situation an individual level analysis of the entire population (i.e., single level model including neighborhood level covariates) would indicate a strong risk-enhancing effect for black race/ethnicity even though the same analysis repeated within each of the respective census tracts would reveal no effect for black race/ethnicity. Including an ABSM in the analysis of the entire population would reduce but not necessarily eliminate the effect of the grouping of blacks in socially disadvantaged census tracts. Multilevel modeling reduces the possibility of this type of misspecification by providing an intercept for each census tract (i.e., a random intercept model) as explained above. The coefficient for the random intercept captures the increased risk experienced by blacks due to being differentially distributed in socially disadvantaged neighborhoods.

This contrast between single level models with level two covariates and multilevel models can be demonstrated graphically. For example, achieving the best least squares fit in a single level model at the individual level requires the association between race and the dependent variable (i.e. the slope) to pass through a single intercept (see Figure 1). By contrast, in a multilevel model each census tract has its own intercept, therefore the slope for the association for race and the dependent variable passes through the mean level of the outcome for each census tract, resulting in two variance estimations—one for the average variance of the dependent variable around each census tract mean (i.e., the within census tract variance) and one for the variance of all the census tract means around the overall mean (i.e., between census tract means). This is called a random intercept model. This is especially important for health inequalities research because in the USA, blacks are a minority that is disproportionately distributed in neighborhoods where the overall mortality is high and contextual factors may be associated with this increased mortality.

Despite calls for multilevel analyses of health inequalities (Soobader, Cubbin, Gee, Rosenbaum, & Laurenson, 2006), few studies have employed a multilevel design to directly assess the mortality risk between blacks and whites in terms of within-neighborhood and between neighborhood risk (Pickett & Pearl, 2001; Subramanian, Chen, Rehkopf, Waterman, & Krieger, 2005; Yen & Kaplan, 1999). The primary problem is the limited number of data sources with a sufficient number of outcomes nested within an areal unit of analysis that approximates neighborhoods. A solution to this data limitation has recently been provided by Subramanian and colleagues (Subramanian et al., 2005). They introduce a new application for a multilevel methodology that has been previously described but never applied to US mortality data. Subramanian and colleagues' results regarding inequality in the risk of mortality between blacks and whites in Massachusetts using this approach are in contrast with those for the single level analyses that include ABSMs (Subramanian et al., 2005). In their final random effects model, the mortality odds ratios for the main effect of black race as well as for all interactions involving black race and age are less than one compared with the reference group, suggesting that black race is not associated with increased mortality once the between census tract variance in mortality is taken into account.

By contrast, in single level models the mortality odds ratios for black race/ethnicity typically never go below 1.00, even with the inclusion of ABSMs in the model (LeClere et al., 1997). This difference in estimated effects suggests that the effect of black race at the individual level is exaggerated in single level models even in the presence of a neighborhood level ABSM term due to the fact that the effect for black race in single level models is conflated with the between census tract differences in mortality, where much of black risk resides due their disproportionate residence in high risk neighborhoods.

Since Robinson's seminal paper (1950) on ecologic fallacy few studies have challenged the validity of inferences made from single level models using individual level data (Alker, 1969; Subramanian, Jones, Kaddour, & Krieger, 2009). Subramanian and colleagues (2009) recently revisited the Robinson analysis using the multilevel approach employed in this analysis and found that "ignoring ecologies [led Robinson] to a severely incomplete, if not misleading knowledge. (Subramanian et al., 2009)."

The current study attempts to replicate the findings of Subramanian and colleagues (2005) using the same methodology on a different population and to compare the results of single level and multilevel approaches as they apply to estimates of the effect for black race. The sample was drawn from Los Angeles County, California which has nearly twice the black population and is more segregated than Massachusetts (Massey, 2001), thus potentially magnifying the neighborhood effect that might be conflated with estimates of the effect for black race at the individual level.

METHOD

The Los Angeles County Department of Health Services provided mortality records for all deaths occurring in the county for the years 1989, 1990, and 1991. The Los Angeles County Department of Health Services provided each record with age, gender, and race/ethnicity identifiers as well as census tract of residence at the time of death. The geocode rate for the mortality data reported by the county was 85.1%. Denominator data consisting of population counts for each gender, age group, and race/ethnicity cell were obtained from the 1990 U.S. census. The 1990 US census was also accessed to generate the ABSM percent households in poverty at the census tract level. Death records missing demographic data were omitted. The final dataset consisted of 129,120 premature death records corresponding to 17,024,709 person years distributed across 1,552 Los Angeles County census tracts.

Data Structure

The data were analyzed using a method described by Subramanian and colleagues (Subramanian et al., 2005; Subramanian, Duncan, & Jones, 2001) that permits multilevel analysis of mortality data obtained from vital records using census data as a denominator. The mortality data and census data are divided into twenty cells per census tract, defined by gender, race, and age group (see Table 1). The total number of deaths in each cell over the three study years was used as the numerator, while the census tract population for that gender, race, and age group category for 1990 tripled was used as the denominator (i.e., person years). The cells were then nested within each census tract across Los Angeles County. Structurally, the model is identical to models with individuals at level 1 only, and the census data provides the population at risk (Subramanian et al., 2005; Subramanian et al., 2001).

The outcome then is defined as a proportion: the number of deaths in the cell over the total population (in person-years) for the cell. The individual level predictors correspond to the five age categories (0-14 years, 15-24 years, 25-44 years, 45-64 years and 65 years and older), the two gender categories (male and female) and the two race/ethnicity categories

(white and black). Because the 1990 Census does not provide populations for race by ethnicity, age, and sex combined, the white category utilized in this study includes both Hispanics and non-Hispanics. In any event, 99.4% of deaths defined as Hispanic were also categorized as white.

Census tracts were characterized using the same measure used by Subramanian and colleagues (2005), percent households in poverty. The variable was divided into four categories (0-4.99%, 5-9.9%, 10-19.9%, and greater than 20%). This is the same variable recommended by Kreiger and colleagues (2003a) for use as an area based sociodemographic measure (ABSM) at the census tract level to assess the effect of social class.

The following exclusions were applied to the Los Angeles County tract level data: Tracts with < 500 population in 1990 and/or suspect based on examining trends in population from 1990-1999 (n=9), all non-residential tracts in the City of Industry (n=6), tracts in cities with only one tract (n=11), tracts in unincorporated cities with no data available at the city level (n=59), and tracts with 0 households (n=3). In the cell level data, all cells with 0 populations were omitted (n=1050 from all deaths file), as were cells with proportion deaths > 1SD (n=584). Almost the entire latter group had only 1 or 2 individuals in the cell.

While the choice of scale to characterize neighborhood represents a contentious issue in the literature we feel justified in using census tract as the basis to characterize neighborhood in this analysis for two reasons. First, Krieger and colleagues (2003) have studied the issue of scale with regard to including area based socioeconomic measures (ABSM) and concluded that census tract and block group are the most meaningful level for characterizing social class effects related to neighborhood of residence. Second, the study by Subramanian (2006), which is replicated in this analysis, uses census tract as the primary areal unit for their analysis.

Statistical Analysis

Data were analyzed using both single level models with and without a census tract level covariate included and the corresponding multilevel models using random intercepts and random slopes. The modeling progressed through two stages for both the single level and multilevel models: 1) a basic model with individual level variables and their interaction terms included and 2) a complete model with individual level variables and their interaction terms as well as the census tract level variable percent households in poverty and its interaction with the race variable. The multilevel models included a random intercept and a random slope for race. In the complete model the percent households in poverty variable is included at the census tract level as well as the cross level interaction between race and census tract poverty. These models are equivalent to the model used by Subramanian and colleagues (2005). The complete multilevel model used was the following:

$$\text{logit } (\pi_{ijk}) = \beta_{0jk} + \beta_1(\text{Male})_{jk} + \beta_2(\text{Black})_{jk} + \beta_3(15 - 24)_{jk} + \beta_4(25 - 44)_{jk} + \beta_5(45 - 64)_{jk} + \beta_6(>65)_{jk} + \beta_7(\text{Male} * \text{Black})_{jk} + \beta_8(\text{Male} * 15 - 24)_{jk} + \beta_9(\text{Male} * 24 - 44)_{jk} + \beta_{10}(\text{Male} * 45 - 64)_{jk} + \beta_{11}(\text{Male} * >64)_{jk} + \beta_{12}(\text{Black} * 15 - 24)_{jk} + \beta_{13}(\text{Black} * 25 - 44)_{jk} + \beta_{14}(\text{Black} * 45 - 64)_{jk} + \beta_{15}(\text{Black} * >64)_{jk} + \beta_{16}(\text{Black} * 15 - 24)_{jk} + \beta_{17}(5 - 10\% \text{Poverty})_k + \beta_{18}(10 - 20\% \text{Poverty})_k + \beta_{19}(20\% \text{Poverty})_k + \beta_{20}(\text{Black} * 5 - 10\% \text{Poverty})_{jk} + \beta_{21}(\text{Black} * 10 - 20\% \text{Poverty})_{jk} + \beta_{22}(\text{Black} * >20\% \text{Poverty})_{jk}$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \mu_{0jk}$$

$$\beta_{2k} = \beta_2 + \nu_{2k}$$

$$\begin{pmatrix} \nu_{0k} \\ \nu_{2k} \\ \mu_{0jk} \end{pmatrix} \approx N(0, \Omega_\nu): \quad \Omega_\nu = \begin{pmatrix} \sigma^2_{\nu 0} & & \\ \sigma_{\nu 0 \nu 2} & \sigma^2_{\nu 2} & \\ & & \sigma^2_{\mu 0} \end{pmatrix}$$

where logit (π_{ijk}), the log odds of mortality for individuals in cell j in census tract k , is a function of β_{0jk} , the log odds of mortality for the reference group, plus the effects of the relevant cell and census tract level predictors and interactions as well as the residual effects at the cell level for individuals in cell j within census tract k , μ_{0jk} , and residual effects at the census tract level for cells within census tract k , ν_{0k} . Finally, the residual effect of census tract k 's deviation from the mean slope of the effect of black race on mortality is added, ν_{2k} . The variances at the census tract level are assumed to be normally distributed with a mean of 0 and a covariance matrix Ω_ν . The elements of the covariance matrix are the between census tract variance for the intercept, $\sigma^2_{\nu 0}$, the between census tract variance in the slopes for Black race across census tracts, $\sigma^2_{\nu 2}$, and the covariance between the two variances, $\sigma_{\nu 0 \nu 2}$. Finally, the variance at the cell level is assumed to be normally distributed with a mean of 0 and a variance of Ω_μ which is equal to the variance of cell level random effects, $\sigma^2_{\mu 0}$.

It should be noted that these models addressed the over-dispersion commonly found in binomial modeling, which was likely with the current data due to the large number of zero cells. We used an additive approach to address over-dispersion by “adding” an additional random term to the model. Several researchers have described this approach of including an additional “pseudo” level at level 1 so that each level 2 unit has exactly one level 1 unit (Browne, Subramanian, Jones, & Goldstein, 2005; Goldstein, 2003). According to Browne and colleagues (2005) this proportion could be thought of as consisting of two components, the fitting of 0-1 responses (i.e., a function of the denominator) at level 1 and the fitting of the proportions at the “pseudo” level, also at level 1. Including a random intercept at the census tract level results in the three level binomial logit link model presented above. All models were analyzed using MLwiN (Rabash, Browne, & Goldstein, 2000). The random coefficient models employed the Iterative Generalized Least Squares (IGLS) procedure in the MLwiN program to obtain parameter estimates, which is equivalent to the generalized least squares procedure in SAS (Singer & Willett, 2003). Odds ratios and 95% confidence intervals are reported and were used to examine the variation in premature mortality based on individual and census tract level predictors.

RESULTS

Table 1 describes the data according to the twenty cell definitions. The crude mortality rates demonstrate that for every age, race, and sex category the likelihood of mortality is greater for blacks than for whites. Calculating an age adjusted odds ratio (aOR) for this data results in an increased risk for blacks (aOR = 1.27).

We now present the two strategies for analyzing individual level data with neighborhood level covariates. We first present the results of single level models where all of the data is analyzed at the individual level including the census tract level variable, which in this case is the percent households in poverty categorized into four categories (0-4.99%, 5-9.9%, 10-19.9%, and greater than 20%). The census tract level covariate in this model is typically referred to in the literature as an area based socioeconomic measure (ABSM) and has been proposed as a substitute for European measures of social class to monitor social inequalities (Krieger et al., 2003a, 2005). Next we present the results of multilevel models where the individual level data is nested within its respective census tract and the census tract level variable, percent households in poverty, is included at the census tract level.

The first set of results are from single level models with and without adding the ABSM. In Table 2, Model 1 only the individual variables and their interactions are included. These results are typical of most analyses that have been used to characterize the health inequality associated with black race. The risk of mortality is 19% higher for blacks than for whites (aOR= 1.19, 95% CI=1.11-1.28) (Table 2, Model 1) in the reference group. In the presence of the numerous interactions we can better characterize this relation by predicting the risk of mortality for the interaction between age group and race and presenting it graphically (See Figure 2). In Figure 2, graph A for every age category except the oldest age category, where there is a crossover, blacks are at greater risk of mortality compared with whites. The black-white mortality crossover has been observed for over 25 years in mortality studies and refers to the age at which black mortality rates fall below those of whites, typically above the age of 75 (Corti, Guralnik, Ferrucci, Izmirlian, Leveille, Pahor et al., 1999; Wing, Manton, Stallard, Hames, & Tryoler, 1985). Explanations for the crossover fall into two general categories; 1) loss of frailer subpopulations at younger ages among blacks (i.e., the heterogeneity explanation) and 2) greater misreporting of age by blacks leading to underestimation of mortality (i.e., the data quality explanation) (Lynch & Brown, 2003).

In the next single level model the ABSM and its interaction with black race are added. The effect for blacks is no longer significant (aOR= 0.89, 95% CI=0.80-1.00) (Table 2, Model 2) after accounting for the strong effect of census tract poverty and its cross level interaction. Because this effect is observed in a single level model it is not clear whether the effect of the ABSM is acting as a marker for socioeconomic status at the individual level or due to similar social and environmental exposures for those living in neighborhoods characterized by increasing levels of poverty. For example, Krieger and colleagues (2003b) suggest the effect of an ABSM like percent households in poverty have effects independent of individual level measures of socioeconomic position. In addition, there is a significant interaction between census tract poverty and black race indicating the effect of poverty is greatest among black residing in the poorest neighborhoods (aOR= 1.21, 95% CI= 1.12-1.32). Again, predicting the results graphically demonstrates that in the single level model including the ABSM does not completely remove the health inequality in mortality for blacks compared with whites (Figure 2, graph B).

The multilevel models are essentially equivalent to the single level models with the exception of random intercepts at the cell and census tract levels as well as a random slope for race at the census tract level. In the first multilevel model the main effects for all the individual level predictors and their interaction do not vary that much from the equivalent single level model except for the effect of race. In this model allowing the intercepts to vary randomly dramatically reduces the effect of black race (aOR= 0.75, 95% CI=0.70-0.79) (Table 2, Model 1) compared with the reference group. This effect for black race is essentially a comparison of mortality risk within each census tract average across all of the study census tracts.

In the next multilevel model percent households in poverty and its interaction with black race is added at the census tract level. To better characterize the effect of black race in the presence of multiple interaction terms, we predicted the risk of mortality for the interaction between age group and race in the two single level models (Table 3, Model 1 and 2) as well as in the complete multilevel model (Table 2, Model 2). As can be seen in Figure 2, in the single level model without an ABSM the odds of mortality for blacks compared to whites are higher in nearly all age groups. Adding the ABSM to the single level model reduced the discrepancy between blacks and whites, yet blacks still demonstrated a greater risk of mortality compared to whites for most age groups except for a crossover in the over 65 age group. In the multilevel model there is no difference in the risk of mortality between blacks and whites except for the crossover among the oldest age group, where risk of mortality is

lower for blacks than for whites. Again, we can better characterize this relation by predicting the risk of mortality for the interaction between age group and race and presenting it graphically (See Figure 2, Graph C). The difference in the multilevel model compared with the single level models is that the between census tract difference in mortality has been partitioned out. Consequently the multilevel model is essentially a comparison of the risk of mortality for blacks and whites residing within the same census tract.

The analyses using multilevel modeling replicate the study by Subramanian and colleagues (2005). The results generally parallel those observed by Subramanian and colleagues (2005) with some notable exceptions. First, the age effects are less marked in the Los Angeles County sample, thereby blunting the black-white mortality crossover occurring in the oldest age group. Next, there is a notable difference for the effect of race. In the Massachusetts analysis the risk of mortality for blacks is higher compared to the white reference category in the model with only individual level predictors (aOR= 1.30, 95% 1.08-1.56) and only falls below that of the white reference category when census tract level predictors are added to the model (aOR= 0.72, 95% 0.53-0.98). In addition the magnitude of effect for census tract poverty is different. While the trend associated with increasing poverty was in the same direction for the two studies, the effect was twice as strong in Los Angeles County (e.g., aOR = 2.88 for > 20% poverty in Los Angeles versus aOR = 1.42 in Massachusetts). Also of note was the lack of an effect for the cross level interaction between census tract level poverty and black race in the Los Angeles County study compared with the Massachusetts study. In terms of the random effects, the between census tract level variance in mortality was large for both whites (i.e., $\sigma^2_{v0} = 0.469$) and blacks (i.e., $\sigma^2_{v0} + 2(\sigma_{v0v2}) + \sigma^2_{v2} = 0.374$) in Los Angeles County (Note: the estimate for blacks is the sum of the between census tract variance for whites (σ^2_{v0}), the differential variance associated with blacks (σ^2_{v2}) and two times the covariance between these two variances (σ_{v0v2})). This is also in contrast to the finding in the Massachusetts study that there was significant heterogeneity (i.e., little between census tract variance for whites and a large amount of between census tract variance for blacks). This effect may explain why the protective effect of black race emerges in the Los Angeles County sample as soon as between census tract mortality is allowed to vary compared with single level models. It also may explain why there is a relatively small main effect for the cross level interaction between black race and census tract poverty, because both races have significant between census tract level variance. In addition, the inclusion of census tract poverty and the cross level interaction tend to explain more of the between census tract level variance in mortality for whites (0.47 to 0.23) than for blacks (0.37 to 0.34).

There are several possible explanations for these differences based on the nature of the underlying populations. First, the Massachusetts mortality data includes both urban and rural samples whereas the Los Angeles County data is almost entirely urban. Given that rural residents do not share the mortality profiles of urban dwellers with regard to social disadvantage, which is believed to be related to weaknesses in standard social disadvantages measures when applied to rural samples due to relatively small population counts and greater population heterogeneity in those samples and not necessarily due to the absence of an effect (Haynes & Gale, 2000), there may have been blunting of the between census tract variance for whites assuming that whites are disproportionately represented in the rural areas. Second, the percentage of blacks in Massachusetts (5.5%) was nearly half that of Los Angeles County (9.9%) in 1990. Given that residential segregation is driven by white avoidance of blacks, and avoidance behavior by whites tends to begin when the percentage of blacks approaches 10% (Massey, 2001), it is likely that residential segregation is a bigger issue in Los Angeles County cities (n=92) compared with all cities across the state of Massachusetts. In fact, measures of segregation are far lower even for major cities in Massachusetts (e.g., Boston) compared to major cities in Los Angeles County (e.g., Los

Angeles, Long Beach) (Massey, 2001), reflecting a pattern of urban development in southern California that was driven by racial considerations (Pulido, 2000). This would suggest that a mortality gradient at the neighborhood level has developed in Los Angeles County, which would explain why accounting for between census tract variance in the Los Angeles County data essentially removed any fixed effect for race at the individual level. In contrast, for Massachusetts most of the between census tract level variance in mortality risk is either very small, as it is for whites ($\sigma^2_{v0} = 0.085$), or in the case of blacks ($\sigma^2_{v0} + 2(\sigma_{v0v2}) + \sigma^2_{v2} = 0.524$) is primarily attributable to the large variance in the relationship between race/ethnicity and mortality across Massachusetts census tracts ($\sigma^2_{v2} = 0.337$) (Subramanian et al., 2005). Again, this could be due to the blunting of the relationship between disadvantage and mortality in rural areas.

To explore this possibility we repeated the analysis for census tracts in Los Angeles County cities where less than 8% of the population was black and for cities where greater than 8% of the population was black. We found that the analysis of cities with less than 8% black yielded results similar to those seen for Massachusetts, whereas in cities with a higher proportion of blacks had virtually no black-white within census tract difference in mortality. This suggests that in Los Angeles County cities unaffected by residential segregation, race effects associated with lower levels of income and education do play a role in explaining the increased odds of mortality for blacks in those cities.

DISCUSSION

The results indicate that, using the same data, dramatically different conclusions could be reached depending on whether a single level model or a multilevel model was used to analyze the Los Angeles County mortality data. In the single level models blacks are at greater risk of mortality in most age groups except in the elderly, where the well documented cross over in mortality occurs (Berkman, Singer, & Manton, 1989; Corti et al., 1999; Wing et al., 1985). Inclusion of census tract ABSMs in the single level model reduced but did not eliminate the predicted effect of black race that is typically seen in single level studies of national survey data (e.g., NHIS) which include ABSMs. In contrast, the predictions of the multilevel model show no difference in the risk of mortality for blacks and whites for most age groups, as well as crossovers, where risk in blacks is lower than whites, at both the oldest and youngest age groups. The fact that the multilevel models account for the between census tract variance in mortality indicates that black are more likely to reside in high poverty neighborhoods with elevated mortality rates for both blacks and whites. Consequently, single level models, even when they do include ABSMs, can exaggerate the effect of black race on mortality risk (see Figure 1).

Furthermore, the effect of adding census tract level ABSM percent households in poverty also varied between the single level and multilevel models. In the single level model, the addition of the ABSM reduced the effect of black race/ethnicity so that blacks had a near-equivalent odds of mortality compared to whites in the reference group (Table 2, Model 3). This is commonly seen in the literature where ABSMs are included in individual level models and reduce but do not necessarily eliminate the individual effect of race, which has made interpretation of the effect in the literature equivocal. While most of these studies argue that the association is evidence of a contextual effect at the neighborhood level (Diez-Roux, Nieto, Muntaner, Tyroler, Comstock, Shahar et al., 1997; Krieger et al., 2005; Krieger et al., 2002; Mansfield, Wilson, Kobrinski, & Mitchell, 1999; Waitzman & Smith, 1998), others argue the census tract level variables are proxies for unmeasured individual level predictors and/or reflect the movement of individuals in and out of neighborhoods based on health status, a structural confounding effect (Mechanic, 2005; Oakes, 2004; Pickett & Pearl, 2001; Sampson, Morenoff, & Gannon-Rowley, 2002). In the multilevel model the

interpretation is more straightforward; the fact that the census tract variables only affected the between tract level variance is consistent with a contextual effect, although structural confounding cannot be ruled out, and the fact that the race/ethnicity effect was unaffected suggests that unmeasured individual level confounders associated with neighborhood poverty are an unlikely source of the observed results. Our results are consistent with the research literature which proposes health inequalities are socially determined.

Acknowledgments

This study was supported by grant R01 AA013810-01A2 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). The views presented in this paper are those of the authors and do not represent those of the funding agencies

REFERENCES

- Alker, HA. A typology of ecologic fallacies. In: Dogan, M.; Rokkan, S., editors. *Quantitative Ecological Analysis* pp. 69-86. Massachusetts Insititue of Technology; Cambridge, Massachusetts: 1969.
- Berkman L, Singer B, Manton K. Black/white differences in health status and mortality among the elderly. *Demography* 1989;26(4):661-678. [PubMed: 2583323]
- Bond Huie SA, Hummer RA, Rogers RG. Individual and contextual risks of death among race and ethnic groups in the United States. *J Health Soc Behav* 2002;43(3):359-381. [PubMed: 12467258]
- Braveman PA, Egerter SA, Cubbin C, Marchi KS. An approach to studying social disparities in health and health care. *Am J Public Health* 2004;94(12):2139-2148. [PubMed: 15569966]
- Browne WJ, Subramanian SV, Jones K, Goldstein H. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society* 2005;168(3):599-613.
- Corti MC, Guralnik JM, Ferrucci L, Izmirlian G, Leveille SG, Pahor M, et al. Evidence for a black-white crossover in all-cause and coronary heart disease mortality in an older population: the North Carolina EPSE. *Am J Public Health* 1999;89(3):308-314. [PubMed: 10076478]
- Cubbin C, LeClere FB, Smith GS. Socioeconomic status and injury mortality: individual and neighbourhood determinants. *J Epidemiol Community Health* 2000;54(7):517-524. [PubMed: 10846194]
- Diez-Roux AV, Nieto FJ, Muntaner C, Tyroler H, Comstock GW, Shahar E, et al. Neighborhood Environments and Coronary Heart Disease: A Multilevel Analysis. *American Journal of Epidemiology* 1997;Vol. 146(No. 1):48-63. [PubMed: 9215223]
- Geronimus AT. To mitigate, resist, or undo: addressing structural influences on the health of urban populations. *Am J Public Health* 2000;90(6):867-872. [PubMed: 10846503]
- Goldstein, H. *Multilevel Statistical Methods*. Arnold Publishers; London, UK: 2003.
- Haynes R, Gale S. Deprivation and poor health in rural areas: inequalities hidden by averages. *Health Place* 2000;6(4):275-285. [PubMed: 11027953]
- Kawachi I, Daniels N, Robinson DE. Health disparities by race and class: why both matter. *Health Aff (Millwood)* 2005;24(2):343-352. [PubMed: 15757918]
- Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am J Public Health* 2003a;93(10):1655-1671. [PubMed: 14534218]
- Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health* 2005;95(2):312-323. [PubMed: 15671470]
- Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. 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. [PubMed: 12196317]

- Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *J Epidemiol Community Health* 2003b;57(3):186–199. [PubMed: 12594195]
- Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health* 1997;18:341–378. [PubMed: 9143723]
- Laveist T, Thorpe R Jr. 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*. 2007
- LaVeist TA, Wallace JM Jr. Health risk and inequitable distribution of liquor stores in African American neighborhood. *Soc Sci Med* 2000;51(4):613–617. [PubMed: 10868674]
- LeClere FB, Rogers RG, Peters K. Neighborhood social context and racial differences in women's heart disease mortality. *J Health Soc Behav* 1998;39(2):91–107. [PubMed: 9642901]
- LeClere FB, Rogers RG, Peters KD. Ethnicity and mortality in the US: Individual and community correlates. *Social Forces* 1997;76(1):169–198.
- Lynch SM, Brown S. Black-white differences in mortality compression and deceleration and the mortality crossover reconsidered. *Research on Aging* 2003;25(5):456–483. B.
- Mansfield CJ, Wilson JL, Kobrinski EJ, Mitchell J. Premature mortality in the United States: the roles of geographic area, socioeconomic status, household type, and availability of medical care. *Am J Public Health* 1999;89(6):893–898. [PubMed: 10358681]
- Massey, DS., editor. Residential Segregation and Neighborhood Conditions US Metropolitan Areas. National Academy Press; Washington, DC: 2001.
- Mechanic D. Policy challenges in addressing racial disparities and improving population health. *Health Aff (Millwood)* 2005;24(2):335–338. [PubMed: 15757916]
- Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 2004;58(10):1929–1952. [PubMed: 15020009]
- Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001;55(2):111–122. [PubMed: 11154250]
- Pulido L. Rethinking environmental racism white privilege and urban development in southern California. *Annals of the Association of American Geographers* 2000;90(1):12–40.
- Rabash, J.; Browne, WJ.; Goldstein, H. A User's Guide to MLwiN. London, UK: 2000.
- Raudenbush, S.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. Sage; Thousand Oaks, CA: 2002.
- Robinson WS. Ecological correlations and the behavior of individuals. *American Sociological Review* 1950;15:351–357.
- Sampson RJ, Morenoff JD, Gannon-Rowley T. Assessing “Neighborhood Effects”: Social processes and new directions in research. *Annu Rev Sociol* 2002;28:443–478.
- Singer, JD.; Willett, JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press; New York, NY: 2003.
- Soobader M, Cubbin C, Gee GC, Rosenbaum A, Laurenson J. Levels of analysis for the study of environmental health disparities. *Environ Res* 2006;102(2):172–180. [PubMed: 16781704]
- Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. *Am J Public Health* 2005;95(2):260–265. [PubMed: 15671462]
- Subramanian SV, Duncan C, Jones K. Multilevel perspectives on modeling census data. *Environment and Planning* 2001;33:399–417.
- Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: The perils of individualistic and ecologic fallacy. *Int J Epidemiol*. 2009
- Waitzman NJ, Smith KR. Phantom of the area: poverty-area residence and mortality in the United States. *Am J Public Health* 1998;88(6):973–976. [PubMed: 9618634]
- Williams DR. Race, socioeconomic status, and health. The added effects of racism and discrimination. *Ann N Y Acad Sci* 1999;896:173–188. [PubMed: 10681897]

- Williams DR, Jackson PB. Social sources of racial disparities in health. *Health Aff (Millwood)* 2005;24(2):325–334. [PubMed: 15757915]
- Wing S, Manton KG, Stallard E, Hames CG, Tryoler HA. The black/white mortality crossover: investigation in a community-based study. *J Gerontol* 1985;40(1):78–84. [PubMed: 3965564]
- Winkleby MA, Cubbin C. Influence of individual and neighbourhood socioeconomic status on mortality among black, Mexican-American, and white women and men in the United States. *J Epidemiol Community Health* 2003;57(6):444–452. [PubMed: 12775792]
- Yen IH, Kaplan GA. Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study. *Am J Epidemiol* 1999;149(10):898–907. [PubMed: 10342798]

Figure 1.

Least squares regression lines for the association between mortality and race for data from two neighborhoods (i.e., circle and square) using single level model with a single intercept and a random intercept model with each neighborhood having its own intercept. Residents of the circle neighborhood are predominantly black residing in high mortality neighborhoods, residents of the square neighborhood are predominantly white residing in a low mortality neighborhood. The single level model slope indicates whites have lower risk of mortality while the average slope from the random intercept model indicate blacks have lower mortality if you account for the fact they tend to be grouped in a bad neighborhood.

Figure 2.

Graphs demonstrating the predicted risk of mortality with standard errors (dashed lines) by age group and race for single level and multilevel models: Graph A) single level model with no ABSMs (Table 2, Model 1), Graph B) single level model with census tract level ABSM (Table 2, Model 2), and Graph C) multilevel model with census tract level ABSM (Table 3, Model 2).

Table 1

Description of data used for multilevel analysis in terms of cell definitions

	No. of Cells (% of Total)	No. of Deaths	Person Years	Mortality Rate per 100,000
Individual Level	29936	129120	17024706	
Age 0-14				
White male	1552	1819	1357995	134
Black male	1507	564	358389	157
White female	1551	1351	1291500	105
Black female	1507	465	350607	133
Age 15-24				
White male	1552	2086	1038381	201
Black male	1472	685	219369	312
White female	1552	436	941448	46
Black female	1481	166	231822	72
Age 25-44				
White male	1552	7266	2548551	285
Black male	1541	2130	454032	469
White female	1552	2204	2367933	93
Black female	1539	831	524859	158
Age 45-64				
White male	1552	11130	1337199	832
Black male	1503	3135	220887	1419
White female	1551	6989	1393455	502
Black female	1476	2080	264408	787
Age 65 and over				
White male	1548	33165	752466	4408
Black male	1157	4661	92115	5060
White female	1547	42094	1132122	3718
Black female	1244	5863	147168	3984

Table 2

Adjusted Odds Ratios and 95% Confidence Intervals for individual and census tract level predictors of mortality in single level models.

Fixed Effects	Model 1	Model 2
Individual Level Predictors		
Age		
0-14	1.00	1.00
15-24	0.43 (0.39, 0.48)	0.43 (0.39, 0.48)
25-44	0.872 (0.82, 0.93)	0.90 (0.85, 0.96)
45-64	4.65 (4.40, 4.90)	5.0 (4.74, 5.27)
>65 yrs	33.37 (31.70, 35.13)	35.43 (33.66, 37.30)
Gender		
Female	1.00	1.00
Male	1.25 (1.18, 1.33)	1.00
Race/Ethnicity		
White	1.00	1.00
Black	1.19 (1.11, 1.28)	0.89 (0.80, 1.00)
Individual Interaction Effects		
Gender/Race		
Male Black	1.02 (0.98, 1.06)	1.03 (0.99, 1.06)
Gender/Age		
Male		
0-15 yrs	1.00	1.00
15-24 yrs	3.45 (3.09, 3.84)	3.40 (3.05, 3.79)
25-44 yrs	2.42 (2.25, 2.60)	2.39 (2.22, 2.57)
45-64 yrs	1.35 (1.26, 1.44)	1.35 (1.26, 1.44)
>65 yrs	0.87 (0.81, 0.92)	0.90 (0.84, 0.95)
Race/Age		
Black		
0-15 yrs	1.00	1.00
15-24 yrs	1.27 (1.15, 1.42)	1.32 (1.18, 1.47)
25-44 yrs	1.37 (1.27, 1.49)	1.38 (1.27, 1.50)
45-64 yrs	1.36 (1.26, 1.46)	1.30 (1.21, 1.41)
>65 yrs	0.87 (0.81, 0.93)	0.80 (0.74, 0.86)
Tract Level Predictors		
% Households in Poverty		
0-4.9%		1.00
5-9.9%		1.29 (1.27, 1.32)
10-19.9%		1.47 (1.45, 1.50)
20-100%		1.86 (1.82, 1.90)

Fixed Effects	Model 1	Model 2
Cross-level Interactions (Race and CT Poverty)		
White, 0-4.9%		1.00
Black, 5-9.9%		0.97 (0.88, 1.07)
Black, 10-19.9%		1.07 (0.98, 1.17)
Black, 20-100%		1.21 (1.12, 1.32)

Note: Odds ratios for interaction terms are the unique (additional) effect of a particular variable and do not include the main effect

Table 3

Adjusted Odds Ratios and 95% Confidence intervals for individual and census tract level predictors of mortality in multilevel models with a random intercept and a random slope for the black covariate

Fixed Effects	Model 1	Model 2
Individual Level Predictors		
Age		
0-14	1.00	1.00
15-24	0.45 (0.41, 0.50)	0.45 (0.41, 0.50)
25-44	0.95 (0.89, 1.01)	0.96 (0.90, 1.02)
45-64	5.30 (5.02, 5.60)	5.37 (5.07, 5.69)
>65 yrs	37.77 (35.88, 39.77)	39.16 (37.12, 41.31)
Gender		
Female	1.00	1.00
Male	1.26 (1.18, 1.34)	1.26 (1.22, 1.30)
Race/Ethnicity		
White	1.00	1.00
Black	0.75 (0.70, 0.79)	0.76 (0.65, 0.88)
Individual Interaction Effects		
Gender/Race		
Male Black	1.02 (0.98, 10.07)	1.03 (0.98, 1.07)
Gender/Age		
Male		
15-24 yrs	3.40 (3.05, 3.79)	3.39 (3.04, 3.78)
25-44 yrs	2.40 (2.23, 2.59)	2.38 (2.21, 2.57)
45-64 yrs	1.37 (1.27, 1.46)	1.36 (1.24, 1.49)
>65 yrs	0.93 (0.87, 0.99)	0.93 (0.86, 0.99)
Race/Age		
Black		
15-24 yrs	1.29 (1.15, 1.46)	1.29 (1.15, 1.45)
25-44 yrs	1.36 (1.24, 1.50)	1.36 (1.24, 1.49)
45-64 yrs	1.16 (1.06, 1.28)	1.16 (1.06, 1.27)
>65 yrs	0.67 (0.62, 0.74)	0.66 (0.61, 0.72)
Tract Level Predictors		
% Households in Poverty		
0-4.9%		1.00
5-9.9%		1.33 (1.23, 1.43)
10-19.9%		1.70 (1.58, 1.83)
20-100%		2.88 (2.67, 3.11)
Cross-level Interactions (Race and CT Poverty)		

Fixed Effects	Model 1	Model 2
White, 0-4.9%		1.00
Black, 5-9.9%		1.05 (0.90, 1.22)
Black, 10-19.9%		1.07 (0.93, 1.23)
Black, 20-100%		0.87 (0.76, 1.00)
Random Effects	Estimate (SE)	Estimate (SE)
Constant/constant (σ^2_{v0})	0.469(0.018)	0.232(0.010)
Constant/black (σ_{v0v2})	-0.104(0.014)	-0.033(0.009)
Black/Black (σ^2_{v2})	0.113(0.016)	0.051(0.012)

Note: Odds ratios for interaction terms are the unique (additional) effect of a particular variable and do not include the main effect.