EXTENDED ABSTRACT

A census data based simulation experiment on the effects of modeling neighborhood effects on a health outcome at "the wrong" scale of the census geography

Claudia Nau, Department of Sociology, Pennsylvania State University

INTRODUCTION

This paper presents a simulation experiment based on census information from the Summary File 1 for Los Angeles County. The primary objective is to assess how predictions of the effects of neighborhood characteristics on a health outcome change when their effect is modeled at one scale of the census geography when the actual mechanism is operating at another. The second objective is to assess how the effects of miss-representation of the neighborhood scale vary by the strength of the "true" underlying effect.

Census units, particularly census tracts have been the most common neighborhood proxy in neighborhood effects research. Much of the popularity of census tracts, however, is grounded in data scarcity. In many data sets the only contextual information that is available is the respondent's membership in a census unit, such as the census tract, block group or less commonly, the census block. Therefore, the use of census units as neighborhood proxies remains an important practice in the neighborhood effects literature.

Defining neighborhoods as discrete units tacitly assumes that the risk exposure is uniform within each unit and that it is adequately captured by the size (scale) and shape of the census units. Spielman and Yoo (2009) show in a simulation study that estimating neighborhood effects at a scale that is too small or too large relative to the spatial dimension of the neighborhood effect can lead to over or underestimate of the effect.

Because of the continuing importance of census units this paper uses actual census information combined with a simulation approach to understand how predictions of neighborhood effects on a health outcome change when "neighborhood" is modeled at one scale of the census geography when the actual neighborhood effect is operating at another. Neighborhood effects, that is, the effect of a particular neighborhood characteristic on a health outcome, are simulated at three levels: the block, block group, and tract. Within each level, a set of scenarios is created by letting the strength of the neighborhood effect vary. Hierarchical linear models are then used to assess (1) how the predictions of the neighborhood effect differ from the data generating effect (2) how the effect of misrepresenting the neighborhood scale varies by strength of the underlying neighborhood effect.

DATA AND METHOD

The study area is Los Angeles County for which neighborhood information is derived from the decennial census of 2000. [1] For this preliminary analysis I will use only one

[1] LA has also been chosen because it is the site of one of the major large scale neighborhood studies, the Los Angeles Family and Neighborhood Survey, which will be used in further analysis and comparisons to this simulation study.

neighborhood level predictor: The percentage of the population that self identifies as being Hispanic. Table 1 shows the percentage of Hispanics in the total population and the mean and standard deviation of the neighborhood percentage of Hispanics at each level of the census geography. Percent Hispanics has been chosen for this preliminary analysis because this neighborhood variable offers the highest between neighborhood variance at each neighborhood level out of all socio-demographic predictors available in the SF1.

I rely on information of the summary file 1 (SF1) because I consider the census geography at the tract, block group and block level. For the latter, only information from the SF1 is released. The summary file 1 provides limited information on the socio-economic characteristics of a neighborhood. It provides however, the racial composition and information that can serve as indicators of neighborhood socio-economic status, such as the percent of female headed households or vacant housing units per census unit. The percent of female headed households will be added to the analysis because it presents a different profile in it's between neighborhood variances at the block, block group and tract level compared to the percent Hispanic population.

The outcome for which the simulation experiment will be conducted is Body Mass Index. To set up the simulation I borrow information from the multilevel analysis of Inagemi et al. (2006) who use the Los Angeles Family and Neighborhood Survey (LA.FANS) in a study of contextual effects on body mass index. I assume that all relevant individual level characteristics are controlled for and borrow the starting values for my simulation from their multilevel model that controls for important individual level confounders. These values are: intercept $b_0 = 24.96 = 25$, neighborhood level variance $u_j = 0.17$, and individual level variance $e_{ij} = 23$. The value of the neighborhood effect b_1 measures the effect of a 1% change in the percent Hispanic of the neighborhood population on the BMI. This b_1 , the neighborhood effect coefficient, will be varied to simulate different scenarios.

The simulation results are preliminary. I am using a simple additive method to generate individual BMI values. For BMI simulations with block level neighborhood effects the simulation is conducted as follows: First, 50 individuals are assigned to each census block, for each block a neighborhood error term is generated using $u_j \sim N(0, 0.17)$. Then, individual level error terms are drawn from a normal distribution $e_{ij} \sim N(0, 24)$ and BMI values are calculated using $Y_{ij} = b_0 + X_1 * b_1 + u_j + e_{ij}$, where X_1 is the vector of the block level percentages of the Hispanic population. For BMI simulations with neighborhood effects at the tract and block group level the overall number of individuals is kept constant but individuals are regrouped into the respective census unit and assigned a u_j . The vector of X_1 becomes the percentage of the Hispanic population at the tract and block group level, respectively.

This simulation strategy leaves the covariance of the level one and level two error terms non-zero. The prediction of the coefficients of the neighborhood effects is not affected. The estimation of the variance components and standard errors might be affected, however. Therefore, they will not be interpreted at this point. Below I explain that the parameters used in the simulation reproduce well in the respective multilevel model. In addition, the $cov(u_j,e_{ij})$ of the simulated error terms is small (in the scenarios presented in Table 1 they are 0.00081 for the Block level simulation and -0.00126 for the tract level). Therefore, I suggest it is defensible to use this simulation approach for this exploratory analysis. For the final version I will use the "simul" function in the actuar R-package that allows to simulate hierarchical data (Goulet and Pouliot 2008).

RESULTS

Table 1 compares two sets of models, each set consists of three types of models that use (1) the block, (2) the block group and (3) the census tract as level two neighborhood units. Set (A) has been estimated on individual level BMI outcomes that have been generated using block level percentages of the Hispanic population and a block level error term. Set (B) is analogously estimated on BMI values simulated as being influenced by fixed and random effects operating at the tract level. The Block-level model (A.1) and the Tract level model (B.3), reproduce the parameters used for the simulation (listed in the first column before each set of models) well. In Model Set (A), the estimates of the regression coefficients for percent Hispanic varies little, independent of which census scale the neighborhood is modeled (the coefficient in the Block and the Tract model is 0.97 compared to the simulation parameter of 1). This consistency is unlike the differences shown in Model Set (B). Model Set (B) has been estimated on BMI values generated with neighborhood effects at the tract level. The tract level neighborhood effects do not stand up well to the disaggregation at the block level. Blocks, which have a higher between neighborhood variability in the percent Hispanic than tracts (Table 1), are estimated to have a notably lower "neighborhood effect" (0.0755 compared to 1.0211 in the tract level model). In model set (A) and model set (B) the coefficients of the tract and block group model are comparable. It is not surprising that tract and block group models behave similarly given that on average only three block groups form one tract and given the fact that neighboring block groups are likely to be similar in terms of their racial composition (a fact that is ignored in this and all HLM analyses).

Graph 1 shows the three types of models (block, block group and tract level models) estimated on BMIs that are simulated using block level (upper panel) and tract level (lower panel) neighborhood effects of varying sizes. Each scenario is depicted at a tick mark of the x-axis. In the first scenario, the neighborhood effect is set to be 0.1, in the second it is set to 1, etc. For each model the point symbol plots the ratio of the coefficient estimate from the respective model over the "true value", the size of the parameter that was used in the simulation (ie. 0.1, 1, 10 or 100). Thus, each panel compares how well each type of model performs in reproducing the respective neighborhood effect and, if and how this performance changes with changing neighborhood effect sizes.

Focusing first on simulations with neighborhood effects of 1 and bigger, I find support for the conclusions from table 1: If the effect operates at the block level, neighborhood models at higher levels of the census geography reproduce the neighborhood effect well (within a range of 3-4% of the "true" parameter). If the neighborhood effect operates at the tract level, block groups reproduce the effect with deviations of less than 5%. Neighborhood effects estimated at the block level when the actual effect operates at the tract level underestimate the neighborhood effect notably, by up to 25%.

Estimating models on BMI simulations where the neighborhood effect was set to 0.01 would not produce statistically significant neighborhood effects in the hierarchical linear models. For Models run on BMI simulations with simulated neighborhood effects of .1 the t-values indicated statistical significance while the respective block or tract level model would poorly reproduce the simulation parameters . The fact that the standard error is small while the coefficient does not reproduce the underlying parameter might be due to the simulation method.

CONCLUSION AND DISCUSSION

The goal of this paper is to produce evidence on how much the prediction of the effect of neighborhood characteristics can be influenced by modeling the effect at a too big or too small scale of the census geography. BMI as health outcome is of particular interest because, as most health outcomes, the variance explained at the neighborhood level as well as the effects of particular neighborhood characteristics tend to be small and therefore potentially more susceptible to the effects of mis-measurement of the neighborhood scale. I choose to perform this simulation study using census information because it is the most available and most widely used source of neighborhood boundaries and contextual information.

Results from this preliminary analysis indicate that block group level and tract level models yield similar neighborhood effects. Modeling small scale effects at large scales of the census geography affects the prediction of neighborhood effects very little. Modeling tract level mechanisms at the block level, on the contrary, leads one to underestimate the effect notably. Effect size above a certain threshold does not affect the relative performance of each model type. This threshold will be explored further for the final paper. In addition, the final paper will be based on a proper multilevel simulation and include interpretations of changes in the variance components. One more neighborhood characteristics, the percent of female headed households, will be added to the analysis.

TABLES AND FIGURE

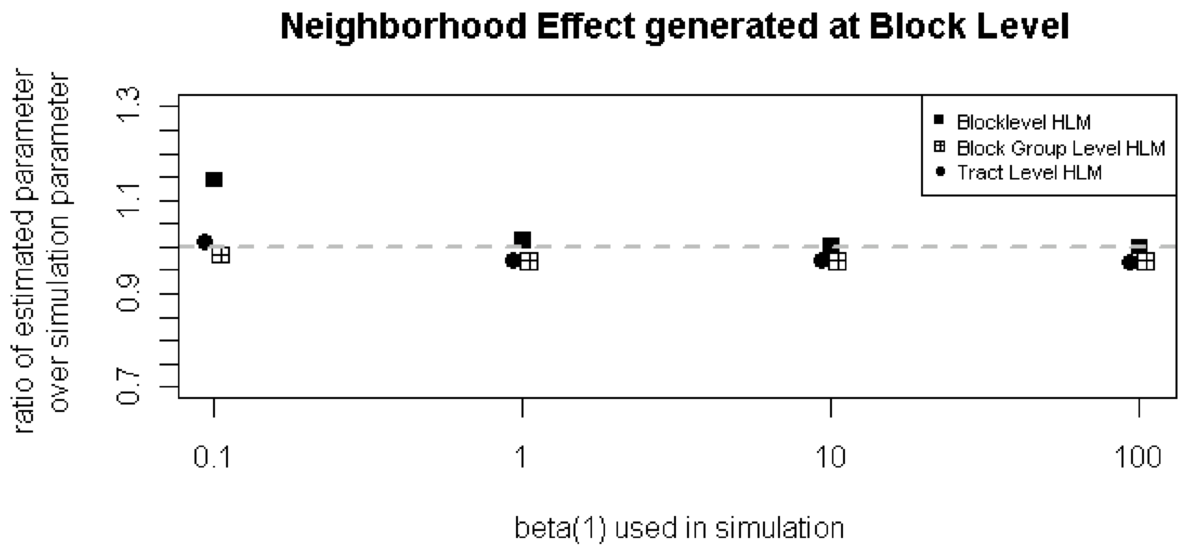
Table 1. Proportion of the Population that self-identifies as Hispanic in Los Angles County, and the average proportion of the Hispanic population by census unit (standard deviation in parentheses).

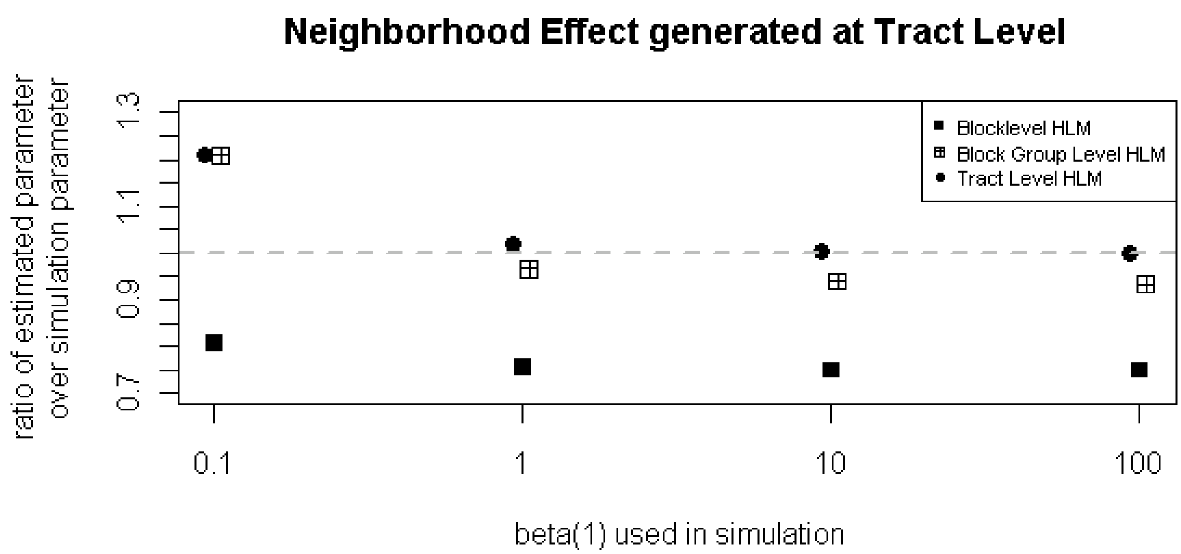
	Hispanic Population
All LA County	44.56%
Block Level average	36.6% (31.58)
Block Group average	42.09% (29.77)
Tract Level average	43.58% (29.49)

Table 2: Comparison of Hierarchical Linear Models estimated on simulated individual level values of Body Mass Index with neighborhood effects simulated at A. the Block level and B the Tract Level.

	A. Neighborhood Effect generated at Block level				**B. Neighborhood Effect generated at Tract Level**			
		A(1)	*A(2)*	*A(3)*		*B(1)*	*B(2)*	*B(3)*
	Simulation Parameters	***Block Level***	**Block Group**	**Tract**	**Simulation Parameters**	**Block Level**	**Block Group**	***Tract***
beta 0	**26**	***25.9947***	25.9989	25.994	**26**	26.1045	26.0116	***25.9848***
		(0.0045)	(0.0051)	(0.0053)		(0.0047)	(0.0105)	***(0.0173)***
beta 1	**1**	***1.0142***	0.9704	0.9732	**1**	0.0755	0.9664	***1.0211***
		(0.0094)	(0.0106)	(0.011)		(0.0097)	(0.0207)	***(0.0333)***
var(uj)	**0.17**	***0.1668***	0.0135	0.0048	**0.17**	0.1978	0.18	***0.1748***
var(eij)	**24**	***23.9981***	24.1782	24.1902	**24**	23.985	23.9856	***23.9855***
AIC		***21652072***	21659189	21659874		21653468	21638582	***21633784***
n individuals	**3595600**	***3595600***	3595600	3595600	**3595600**	3595600	3595600	***3595600***
N Neighborhoods	**71912**	***71912***	6311	2049	**2049**	71912	6311	***2049***

Graph 1: Comparisson of the success of block, block group and tract level hierachical linear models in predicting the underlying neighborhood effect when BMI is simulated to be affected at the block level (upper panel) and the tract level (lower panel)





REFERENCES

Goulet, V, and LP Pouliot. 2008. "Simulation of compound hierarchical models in R." *North American Actuarial Journal* 12:401-412.

Inagami, S, DA Cohen, BK Finch, and SM Asch. 2006. "You Are Where You Shop Grocery Store Locations, Weight, and Neighborhoods." *American Journal of Preventive Medicine* 31:10-17.

Spielman, SE, and E Yoo. 2009. "The spatial dimensions of neighborhood effects." *Social Science & Medicine* 68:1098-1105.