Neighborhood social characteristics and chronic disease outcomes: does the geographic scale of neighborhood matter?

Malia Jones

Prepared for consideration for PAA 2013

Short Abstract

Empirical research on the relationship between characteristics of the local environment and health outcomes depends upon a valid specification of geospatial scale of neighborhoods. In this paper, I compare results of operationalizing neighborhood at four spatial scales derived from US census geography. I compare how neighborhood social characteristics, operationalized at each scale, predict individual-level hypertension and obesity in a sample of Los Angeles County adults (n for individuals=900, n for neighborhoods=65).

Descriptive results show that neighborhood characteristics are similar across scales, but that census tracts do not accurately or consistently capture the scale at which meaningful variation in neighborhood characteristics occur. In multivariate models, the main effects of neighborhood poverty on chronic disease are not driven by the selection of census tracts as the areal unit for neighborhoods. Creating alternative neighborhood scales based on census geography does not appear to perform substantially and consistently better and involves considerable time investment.

Extended Abstract

Introduction

Although many studies have used census tracts to operationalize neighborhoods, there is little theoretical and methodological agreement about how to specify the appropriate scale for neighborhood. Although neighborhoods are clearly spatial in nature, it is not obvious how big a "neighborhood" is. Galster (2001) suggests that scale should be dependent on the level at which meaningful between-group variation exists—i.e., it depends on the scale at which the exposure and the outcome operate. Failing to specify the appropriate geospatial scale can weaken the observed effects, leading to type I error and contributing to inconclusive or conflicting empirical results. Thus, specifying an appropriate measurement approach for the neighborhood is of critical importance to our ability to understand neighborhood effects in empirical models.

In this paper, I describe methods and results from testing the sensitivity of a model of neighborhood effects on health outcomes using different specifications of geographic scale of neighborhoods. I compare results of operationalizing neighborhood at several spatial scales derived from US census geography, each using the neighborhood of residence as the core. The geospatial scales I compare are: Census block-groups, tracts, and first-order and second-order contiguous tracts (defined using two different weighting approaches). I compare how social characteristics, operationalized at each of these four scales, perform in a model of neighborhood-level exposure to stress on individual-level chronic disease outcomes.

Methods

Data and sample

I use a representative sample of adults in Los Angeles County from Waves 1 and 2 of the Los Angeles Family and Neighborhood Study (L.A.FANS). My analytic sample contains 900 adults who lived in 65 Los Angeles census tracts at the time of the first wave of data collection, 2000-2001. I include sample weights in multivariate models, making the results representative of the population of Los Angeles County in 2000 who remained in the County in 2006-2008.

Variables

I use two main chronic disease outcomes, which were selected based on their salience as outcomes of the stress process in adults. Being overweight is measured with body mass index (BMI) and waist circumference, and blood pressure is measured with systolic blood pressure (the top number in a blood pressure reading).

The focal independent variables are neighborhood social characteristics at approximately the time of the Wave 1 survey. I use five key indicators of neighborhood social environment, which have been linked in previous research to social and physical disorder: percent poverty, percent black, percent Latino, percent unemployment, and percent of housing that is renter-occupied. To facilitate interpretation of results in multivariate models, the neighborhood measures have been rescaled so that 1 unit represents the difference between the 25th percentile and the 75th percentile of that characteristic in the analytic sample—for example, a neighborhood that is at the low end of the distribution for poverty versus a neighborhood at the high end of the distribution. In this paper, I focus on the geospatial scale over which these neighborhood social characteristics are averaged, and examine whether variation in the scale makes any difference in the conclusions.

I control for demographic and socioeconomic status variables, including respondent age, education, gender, family income, and race/ethnicity/immigration status reported at Wave 1. I also control for time elapsed between data waves, change in family income, and change in education status. I include an interaction term for individual-level family income with neighborhood-level percent poverty.

My outcomes are measured at Wave 2 of the L.A.FANS survey, and I also control for the same health measure at Wave 1 of the survey, making the models effectively longitudinal in nature.

Creating measures for super-tracts

For each of the five selected neighborhood social characteristics, I calculated the area-weighted mean and population-weighted mean for two areas that are larger than census tracts, which I refer to as 1st-order and 2nd-order super-tracts. First-order super-tracts include the index tract (where the respondent lives) and 1st-order contiguous tracts—those tracts that share a border or a corner with the index tract itself.[1] Then I calculated the area-weighted mean and population-weighted mean across the index tract, 1st-order tracts, and 2nd-order tracts—those tracts which share a border or a corner with one of the 1st-order contiguous tracts.

Area-weighting allows a tract with a larger area to be more influential in the final super-tract value, while population-weighting allows a tract with higher population to have more influence in the final super-tract value. My theoretical framework suggests that people are exposed to disorder in the *area* in which they live, regardless of the number of people who live there. But since disorder is a social ill, there is an argument to be made that a disorderly and densely populated area could be a stronger exposure than a disorderly and sparsely populated area. Since it is not clear which measure is better here, I present results for both.

This process produced area-weighted mean and population-weighted mean measures for each of the five key census characteristics at the 1st-order and 2nd-order super-tract level, for a total of four sets of super-tract measures.

Creating measures for block-groups

I used a different approach to calculate the characteristics of areas smaller than the census tract, using census block-groups. The L.A.FANS sampling strategy involved selecting tracts from within three poverty strata, then to select *blocks* from the selected tracts; respondents are clustered at tract and block levels. However, the Census Bureau does not provide any social characteristics data at the block level—the smallest geographic unit for which data are publically available is the *block group*.

Fortunately, blocks are always nested within block groups. I obtained block-group level data for all block-groups in Los Angeles County from the 2000 decennial census SF3 (U.S. Census Bureau 2000), and derived the five measures of neighborhood social environment. I applied the block-group level data to the census blocks.[2,3] I performed these data processing tasks in ESRI ArcView 10.0, using the 2000 TIGER/Line shapefiles for census block-groups and blocks (U.S. Census Bureau 2000; ESRI (Environmental Systems Resource Institute) 2011). Each of several census blocks sharing a block group got the same values for social characteristics—there is no within-group variation below the block group level, even though my data have a row for every block.[4]

Analytic approach

My analytic approach is to substitute each of these neighborhood measures in a multivariate OLS regression model of health outcomes. My hypothesis is that the coefficients for super-tract characteristics are approximately the same as for census characteristics measured at the census tract

[1] Because census areas are created based largely on population, there was high correlation between the area-weighted mean and the population-weighted mean measures for the same areas.
[2] I used a polygon-to-polygon spatial join of the second type, where each census block is given the attributes of the block group in which it falls.
[3] For all the geospatial processing in this study, I projected my maps and shapefiles to NAD 1983 State Plane California VI, FIPS code 0406, in U.S. feet.
[4] As part of this task I also referenced 2000 census block geography to 1990 census block geography.

level and that the total variance explained will be similar across models, i.e., that the geographic scale at which I operationalize neighborhood is not driving the main results. I assess coefficients for individual variables and measures of overall model fit. Analysis was performed using Stata 11.0 (StataCorp. 2011).

Results and Discussion

Descriptive results

The 65 L.A.FANS census tracts (and ~300 L.A.FANS blocks) have a median of seven 1^{st}-order contiguous tracts and a median of sixteen 2^{nd}-order contiguous tracts. There is wide variation in the mean areas for the various measures. In rural parts of the county, they are much larger than in more densely settled urban parts of the county. For example, in downtown Los Angeles, the sample mean tract size is ~0.5 square miles; 1^{st}-order super tracts are mean 3 sq. mi, and 2^{nd}-order super tracts are mean 7 sq. mi. In a more rural part of the county, the mean tract is nearly 100 square miles; 1^{st}-order super tracts are 300 sq. mi and 2^{nd}-order super tracts are 700 sq. miles.

For each person, the area comprising the neighborhood is partially shared across the six operational definitions of neighborhoods and so we would expect to see similarity among the areas. If tracts are very like their neighboring tracts, we might expect to see quite a bit of similarity—block and tract descriptions would resemble the super-tract descriptions if the scale of segregation is large, or if there is little variation on the characteristic in the county as a whole. However, if tracts are unlike adjacent tracts, we would expect to see differences in the descriptions of the areas when more tracts are aggregated.

My results show that the average social characteristics are quite similar across the neighborhood scales. For example, the mean population in poverty is 22% in block groups, 23% in tracts, and 19-22% for the larger super-tract areas. The only characteristic that shows substantial variation is the proportion unemployed, which is higher in the blocks than in the other five neighborhood definitions. The area-weighted mean measures generally have the lowest values for the neighborhood characteristics. This is because they give more weight to less densely populated places, which are also likely to be populated by more advantaged people—they tend to be more affluent, white, and homeowners. Some of the differences as we move from block-group to tract to super tract are also attributable to regression to the mean—the larger areas are closer to the mean for the county simply because they capture more people. This appears to be most marked for population density and proportion Latino.

A somewhat different story emerges when we look at the average characteristics for various neighborhood scales by region of the County. Here I use the eight Los Angeles County Service Planning Areas (SPAs). There are dramatic differences in in social characteristics from region to region of the County. For example, the proportion in poverty is 37% for the L.A.FANS respondents living in South LA and only 6% for the L.A.FANS respondents living in West LA. The proportion of renter-occupied housing ranges from 24% in Antelope Valley to 78% in Metro LA. Because L.A.FANS oversampled poor and very poor census tracts, this pattern is somewhat exaggerated but it is generally similar to the regional pattern in the County as a whole.[5]

In most SPAs, the larger areas captured by the 1^{st}- and 2^{nd}-order contiguity means diminish the exposure to disadvantage slightly. For example, in the San Fernando Valley, block-group level poverty is 19%, tract-level poverty is 18%, the first-order super tract poverty is 16%, and the second-order super tract is 13% for the area-weighted mean and 16% for the population-weighted mean. It appears that

---

[5] Note that although the L.A.FANS sampling strategy ignored SPAs, the nonpoor, poor, and very poor tracts were not sprinkled at random across the larger landscape of the county. The approach of stratifying all tracts by poverty status and sampling within strata led to an actual sample pattern of poor and very poor sampled tracts coming from the poor and very poor SPAs, and nonpoor tracts coming from the less poor SPAs.

adjacent tracts are generally a little less poor than index tracts. This fact is in part an artifact of oversampling very poor tracts—we see mechanical regression to the mean when we start with very poor tracts and then add more tracts into an average. However, the reverse is true in affluent West LA—the larger area introduces more poverty exposure, indicating that neighboring tracts are generally more poor than index tracts. In Metro and South LA, the poorest SPAs, the larger area has little effect, suggesting that tracts in these SPAs are surrounded by similar tracts. That is to say, there appears to be little difference between an index tract and its adjacent tracts in South LA and Metro, but there is a large difference between the index tract and its adjacent tracts in West LA, and a moderate difference in other parts of the city.

Galster (2001) argues that the appropriate measure for the effect of neighborhoods is the scale at which there are observable, meaningful between-group differences on exposures. Otherwise, how can we distinguish the effect of one "neighborhood" from another? These descriptive statistics suggest that census tracts do not capture the scale at which segregation really operates on five key social characteristics. Furthermore, census tracts do not capture the scale of segregation *consistently* on these five social characteristics.

Multivariate Results

No pattern emerges from the multivariate models to suggest that operationalizing social characteristics of neighborhoods at the block-group, 1st-order, or 2nd-order super tracts would perform better than census tracts. My hypothesis—that the main effects of neighborhood poverty on chronic disease outcomes are not driven purely by the selection of census tracts as the areal unit for neighborhoods—is supported. Creating area-weighted or population-weighted means based on census tract geography does not appear to perform substantially and consistently better and involves considerable time investment. Census data are limited at the block-group level, and again involve considerable time. Block-group models are also difficult because the small number of respondents per block-group makes multilevel modeling unstable.

More broadly speaking, these various "neighborhoods" are so similar to one another in terms of average social characteristics that the choice of scale—among the six nested scales I created—has little effect on the results.

One limitation of this study is that all the geographic scales I test are based on census geography. Census geography is not designed for the purpose of measuring how neighborhoods influence chronic disease outcomes, and is not necessarily the best match for the question I have asked. I show that the effect of averaging across larger areas varies by LA County Service Planning Area, providing evidence that census tracts are *not* the scale at which meaningful between-group variation on these social characteristics occurs, at least not consistently. As Galster (2001) suggests, it may be more informative to seek out the scale at which these social characteristics actually vary.

Empirical approaches to identifying the scale at which there exists meaningful between-group variation in disorder have not been well-explored to date. In most regions of Los Angeles County, residential segregation operates at a fairly large scale; neighboring census tracts tend to resemble each other--in terms of income, race and ethnicity, age, and other characteristics--across broad territories, especially in poor, densely populated neighborhoods. An approach to defining neighborhoods which analyzes the scale of segregation to discover meaningful "neighborhoods" of exposure might yield more reliable results, and this is a key area of future research.

References

Ailshire, J., M. Bader, et al. (2010). Do Neighborhood Effects Depend on the Defintion of the Neighborhood? National Meeting of the Population Associaton of America. Dallas, TX.

ESRI (Environmental Systems Resource Institute) (2011). "ArcMap 10.0. ESRI, Redlands, California. ."

Galster, G. (2001). "On the Nature of Neighborhoods." Urban Studies **38**(12): 2111-2124.

StataCorp. (2011). "Stata Statistical Software: Release 11. College Station, TX: StataCorp LP. ."

U.S. Census Bureau. (2000). "Census TIGER/Line Data." Retrieved January 13, 2012, from http://arcdata.esri.com/data/tiger2000/tiger_download.cfm.

U.S. Census Bureau (2000). Decennial Census, SF-3 file.