Invited Commentary

Invited Commentary: Beyond the Metrics for Measuring Neighborhood Effects

Lynne C. Messer1,2

1 Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC.
2 National Health and Environmental Effects Research Laboratory, Human Studies Division, Environmental Protection Agency, Research Triangle Park, NC.

Received for publication October 24, 2006; accepted for publication November 7, 2006.

The “neighborhood effects” literature, as currently constructed in epidemiology, would benefit from critical attention to the following four issues: 1) use of appropriate measurement tools and methods for neighborhood environments; 2) theoretical or conceptual guidance as to the aspects of residential environments most salient to human health (both within and across health endpoints); 3) the scale on which investigators measure neighborhood environments to best correspond to meaningful neighborhood boundaries and relevant neighborhood exposures; and 4) those selection and structural features that confound investigators’ capacity to draw causal inferences from neighborhood environments to human health. In this issue of the Journal, Mujahid et al. (8) report on the psychometric and ecometric results of their neighborhood scale testing. By providing a more rigorous approach to testing of neighborhood-attribute measurement tools, they make an important contribution to the literature. Also noteworthy is their explicit use of a conceptual model, which may facilitate the development of more meaningful area-level measures. Unfortunately, while Mujahid et al. had the capacity to consider spatially relevant measures, based on prior research and their conceptual model, they relied exclusively on aggregated census units. Hopefully, the scale issue will be addressed more thoroughly in future work. Clearly, Mujahid et al. have made progress in addressing how, and to a lesser extent what, to measure when researchers estimate neighborhood effects. But equally clearly, important work remains to be done.

censuses; data collection; epidemiologic methods; psychometrics; residence characteristics; social class; social environment

Neighborhoods matter to the people who live in them. Residents report feeling safer, experiencing fewer crime-related problems, and being more likely to exercise social control when they are close to home (1, 2). And neighborhoods appear to matter to the health of the people who live in them. While the definitions of “good” and “bad” neighborhoods have changed over time, the literature shows that people who live in “good” neighborhoods have better health than those who live in “bad” ones (3–7). What is unclear is how, and how much, residential context independently shapes health risks.

In order to advance our understanding of neighborhood effects, some fundamental issues need to be addressed. 1) How do we best measure neighborhood characteristics? 2) What aspects of residential environments are salient to human health (and are they the same for different health endpoints)? 3) Where, or on what scale, should we measure residential environments and the boundaries of neighborhoods? And finally, 4) What confounds our capacity to draw inferences from the neighborhood effects literature? Addressing any one of these questions will not guarantee a ready answer to the question of how neighborhoods influence human health, but failing to grapple with these issues will ensure that progress remains limited.

HOW TO MEASURE? HYBRIDS AND MULTIPLE METHODS

In this issue of the Journal, Mujahid et al. (8) make their strongest contribution to the neighborhood effects literature.
by addressing how to measure residential characteristics. They report on a psychometric and econometric analysis of seven neighborhood features (aesthetic quality, walking environment, healthy food availability, safety, violence, social cohesion, and activities with neighbors) which they propose serve as mechanisms to explain potential associations between neighborhoods and cardiovascular health. Questionnaires regarding these features were administered to approximately 6,000 residents of selected census tracts in the three geographic areas (Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York) included in the Multi-Ethnic Study of Atherosclerosis. The paper is part of a larger project (9, 10) and describes an area-assessment hybrid measure that aggregates nonparticipant perceptions to produce area-level scales. Relying on residential informants, rather than study participants, allowed the authors to effectively circumvent the problem of same-source bias (11), and averaging subjective responses was likely to reduce the magnitude of subjective bias in the data. The scale reliability assessment was distinctive (10–12) and was a particular strength of this analysis, but the validity of the scales remains to be demonstrated through use in another population. The use of scales like these will undoubtedly complement other types of neighborhood data, and triangulation of neighborhood-level information (including directly observed, administrative source, and survey data) is strongly encouraged.

WHAT TO MEASURE? THEORY AND MECHANISMS

Despite the attention devoted to the topic in recent years, the neighborhood effects literature is relatively underdeveloped. Four years ago, O’Campo (13) identified the lack of theory as one of the most pressing issues standing in the way of progress in multilevel research, particularly regarding the mechanisms by which neighborhood environments affect health risks (12); little has changed in the ensuing years. The general lack of theoretical attention in the literature makes the paper by Mujahid et al. (8) all the more remarkable, because the authors were explicit in their conceptual model and the underlying empirical work (9). However, while impressive, the paper promises more, from a theoretical perspective, than it delivers. While Mujahid et al. noted the need to identify specific neighborhood features relevant to disease risk and implied that their seven domains demonstrated such mechanistic relevance, the prior theoretical work they cite (9, 10) does not unambiguously support this contention. In previous work (9), Diez Roux reported that three residential characteristics—facility accessibility, activity opportunities, and aesthetic qualities—were empirically associated with physical activity, while the availability of healthy food was associated with diet. Her review noted the limitations of the existing empirical work and concluded that the current evidence linking specific neighborhood features with health behaviors or outcomes was weak, conflicting, and cross-sectional. Probably, the next steps undertaken by these investigators will involve assessing the association between these residential characteristics and cardiovascular outcomes. Assuming that the seven neighborhood characteristics proposed by Mujahid et al. (8) demonstrate utility in association with cardiovascular health, which seems reasonable, intermediate information—linking the distal (neighborhood factors) to the proximal (individual health and related behaviors or states)—will be required before we can understand the “mechanisms” by which neighborhood influences health. For instance, improving the “quality of walking environment” may not change the prevalence of an associated health outcome if residents with different walking environments don’t actually behave differently. Ultimately, linking distal and proximal neighborhood factors with health outcomes is the direction the field, not just these investigators, will need to go in to explain how residential features influence health and disease.

WHERE TO MEASURE? NEIGHBORHOODS AND SCALE

The geographic definition of “neighborhood” varies in the literature, but in the United States it is usually based on a census unit aggregation. In light of the need to consider within- and between-neighborhood heterogeneity, no consensus as to the preferred level of aggregation for best approximating “neighborhood” has emerged. Previous work has found some census units more predictive of particular health outcomes than others (14), but the geographic unit that maximizes predictive utility may not be the one that best corresponds to one’s theory of causation. Mujahid et al. (8) included three units of spatial aggregation in their analysis; in replying to questionnaire items, respondents were asked to consider a 1-mile (1.6-km) radius around their house, and the investigators analyzed data at both the census tract and cluster levels. (It is not clear from the text which measure, census tracts or census clusters, best conformed to this spatial referent.) The good news is that Mujahid et al. were able to demonstrate the robustness of their statistical tests across various geographic aggregations. The bad news is that they had the opportunity to meaningfully contribute to the conversation about neighborhoods and scale and failed to do so in two different ways. First, their conceptual model could have been used to suggest an optimal spatial level for selected neighborhood features to exert their respective influence(s). For instance, we would expect “activities with neighbors” to operate on a smaller scale than “walking environment” and would expect the “availability of healthy foods” to be meaningful only if people shopped in their neighborhoods. Instead of applying their conceptual model, Mujahid et al. appear to have used census tracts or clusters throughout the analyses. Second, within each site, they aggregated the census tracts to produce clusters. Given the sociodemographic characteristics and relative sizes of the three geographic areas in the study, this aggregation certainly resulted in different levels of residential heterogeneity. Census clusters in Forsyth County, North Carolina, simply do not mean the same thing as census clusters in New York City. Mujahid et al. noted that the scales were consistently less reliable in North Carolina than in the other two areas (8, p. 862), which resulted from the smaller numbers and greater within-cluster heterogeneity.
However, the meaning of what constitutes “neighborhood” in these three areas is neither known nor comparable (e.g., a cluster in Maryland might be 1 square mile (1.6 km²) in size, but in North Carolina it might be 10 square miles (16 km²)). How these aggregations coincided with the respondent referent is also not clear. Hopefully, future work will expand on the investigators’ hard-won insight as to the scale(s) at which neighborhoods influence cardiovascular disease.

Unfortunately, Mujahid et al. are not alone in failing to deal with this question adequately—the neighborhood effects literature is plagued by a lack of attention to scale. The geographic levels at which researchers report their statistical findings largely reflect data availability (census, county, etc.) and numerical considerations (outcome and exposure counts per geographic unit) rather than meaningful units of community and neighborhood. In taking a one-dimensional, not conceptually or theoretically informed, approach to aggregating data, we risk encountering the modifiable unit area problem (15), which arises from the imposition of artificial geographic units on continuous spatial phenomena. We can “create” different exposure effects by varying how exposure data are aggregated, and can only avoid this problem by applying substantive knowledge—theoretical, conceptual, or mechanistic—to inform our analysis at meaningful levels.

**CAN WE MAKE CAUSAL INFERENCES? SELECTION AND STRUCTURAL CONFOUNDING**

While the issues associated with causal inference and neighborhood observational studies have been noted elsewhere (16), it seems important to highlight a particular concern: the lack of attention paid to nonrandom selection into neighborhoods. Taking a structural approach, we can think of confounding as occurring when an exposure and an outcome share an underlying common cause (17, 18). Frequently, these underlying common causes, confounders, are neither known nor measured. It has long been recognized in sociology and economics that the geographic distribution of households is not random but arises from political, economic, historical, and social processes; these neighborhood-forming processes cause people to differentially locate and constitute the underlying common causes that produce selection bias. For instance, an educated couple may seek to live in a specific neighborhood because they want to ensure that their children have access to a good educational system. They will seek out neighborhoods with educational resources (e.g., schools and libraries), and the literature suggests that these educated persons are also more likely to exercise, eat healthy food, and seek appropriate medical care (19–23). Neighbors and neighborhoods with these characteristics will be attractive to others of the same ilk and means. Similarly, a South Asian family may locate in a specific area to live in proximity to other South Asian families. Specific dietary, recreational, or cultural practices that influence health may become prevalent in such areas. Consequently, health behaviors and outcomes will tend to cluster geographically, but not as a result of environmental or residential factors.

In most statistical models, noncomparable individual characteristics are routinely adjusted for (ethnicity less so). The causal inference problem arises because the more one controls for the factors that differentially select persons into certain neighborhoods (using age, income, religion, ethnic identity, employment, employment type, sexual orientation, preferred recreational activities, and so on), the less likely it is that an individual will reside in any neighborhood other than his/her own. Confounding that results from differential allocation of people to certain places, by virtue of social stratification or some other selection process, has been termed “structural confounding” (24, 25). Structural confounding ensures that people residing in different residential environments are not exchangeable, because even when matched according to traditional covariate patterns, they still differ with regard to the factors, those unmeasured confounders, which caused them to differentially select into specific neighborhoods. Nonexchangeability makes them noncomparable, and their noncomparability renders causal contrasts across neighborhoods virtually meaningless (26). For instance, Mujahid et al. (8) and their colleagues might want to compare cardiovascular disease outcomes in persons with good walking environments with those in persons with bad walking environments (adjusted for, let’s say, age, education, and disability), but they might really be comparing persons who seek a neighborhood near the lake with persons who require a neighborhood on the public transit line (again, adjusted for age, education, and disability). These two groups of people simply aren’t exchangeable, but their lack of comparability would not be known or adjusted for using traditional methods. Given that most statistical modeling approaches depend on exchangeability across contexts, structural confounding will place significant limitations on these investigators’—and all neighborhood effects investigators’—ability to make causal inferences from observational data. These limitations can be better understood and addressed only through the application of social and economic (choice) theory.

Everyone wants to live in a neighborhood with lots of amenities (clean air, good schools, access to shopping) and minimal disamenities (pollution, crime, noise). Each person’s capacity to locate in such a neighborhood is largely structured; we don’t all get to choose equally. It is almost certainly these structural factors—the unmeasured influences that facilitate our selection into certain environments and discourage our selection into others—that influence our health, over and above what researchers capture in neighborhood surveys or direct observation. Clearly, Mujahid et al. (8) have made progress in addressing how, and to a lesser extent what, to measure when we estimate neighborhood effects. But equally clearly, important work remains to be done.

**ACKNOWLEDGMENTS**

Dr. Lynne C. Messer was supported by a NHEERL (National Health and Environmental Effects Research Laboratory)–DESE (Department of Environmental Sciences
and Engineering) Cooperative Training in Environmental Sciences Research grant (EPA CT CR83323601).

Conflict of interest: none declared.

REFERENCES