Practice of Epidemiology

Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics

Mahasin S. Mujahid[1], Ana V. Diez Roux[1], Jeffrey D. Morenoff[2], and Trivellore Raghunathan[3]

[1] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI.
[2] Department of Sociology, College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI.
[3] Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI.

Most studies examining the relation between residential environment and health have used census-derived measures of neighborhood socioeconomic position (SEP). There is a need to identify specific features of neighborhoods relevant to disease risk, but few measures of these features exist, and their measurement properties are understudied. In this paper, the authors 1) develop measures (scales) of neighborhood environment that are important in cardiovascular disease risk, 2) assess the psychometric and ecometric properties of these measures, and 3) examine individual- and neighborhood-level predictors of these measures. In 2004, data on neighborhood conditions were collected from a telephone survey of 5,988 residents at three US study sites (Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York). Information collected covered seven dimensions of neighborhood environment (aesthetic quality, walking environment, availability of healthy foods, safety, violence, social cohesion, and activities with neighbors). Neighborhoods were defined as census tracts or census clusters. Cronbach's alpha coefficient ranged from 0.73 to 0.83, with test-retest reliabilities of 0.60–0.88. Intraneighborhood correlations were 0.28–0.51, and neighborhood reliabilities were 0.64–0.78 for census tracts for most scales. The neighborhood scales were strongly associated with neighborhood SEP but also provided information distinct from neighborhood SEP. These results illustrate a methodological approach for assessing the measurement properties of neighborhood-level constructs and show that these constructs can be measured reliably.

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

Abbreviations: ICC, intraneighborhood correlation coefficient; SEP, socioeconomic position.

Editor's note: An invited commentary on this article appears on page 868, and the authors' response is published on page 872.

In recent years, there has been increasing interest in the effects of neighborhood on health. Studies have shown that living in a socioeconomically disadvantaged neighborhood is associated with increased morbidity and mortality, inde- pendent of individual-level factors (1–7). Most of these stud- ies have used census-defined areas (census tracts/block groups) as proxies for neighborhoods and aggregate measures of so- cioeconomic position (SEP) as crude proxy measures for a variety of health-relevant features across which neighbor- hoods may differ. This body of work has become the foun- dation of the evolving "neighborhood effects" literature in epidemiology.

The use of aggregate SEP measures raises methodologi- cal questions regarding the ability of these studies to actually

Correspondence to Dr. Ana V. Diez Roux, Center for Social Epidemiology and Population Health, University of Michigan School of Public Health, 1214 South University Avenue, 2nd Floor, Ann Arbor, MI 48104 (e-mail: adiezrou@umich.edu).

estimate neighborhood effects "independently" of individual SEP (8–10). In addition, aggregate SEP measures may be poor proxies for the specific features of neighborhoods that are relevant, limiting the causal interpretation of any associations observed. One way to address these limitations is to investigate the specific causal mechanisms through which neighborhoods affect health outcomes. This requires moving beyond neighborhood SEP to the study of specific features of neighborhoods. Some of these features may be associated with commonly used measures of neighborhood SEP, but others may not. For example, both low-SEP neighborhoods and high-SEP neighborhoods could have poor social cohesion. There may also be important variation in health-relevant dimensions, even among neighborhoods of similar SEP.

The measurement of ecologic features of neighborhoods is much less developed in epidemiology than the measurement of individual-level variables. One option in measuring neighborhood-level properties is to use locational data available in administrative or commercial databases (e.g., locations of recreational facilities or food stores) (11, 12) and geographic information systems (13, 14). A second option is to employ systematic social observation in which raters visit neighborhoods in order to assess them on specific dimensions (15–18). However, these options may not be feasible or may not be suited to the assessment of certain constructs (e.g., social cohesion).

A third option is to measure characteristics of neighborhoods by asking each study participant to report on the conditions in his or her neighborhood (19–23). Although this approach is useful, it has two limitations. One limitation is that reporting bias may create spurious associations between self-reported neighborhood conditions and self-reported health outcomes (source bias) (24). For example, persons who are sedentary may rate their neighborhoods as worse with regard to recreational resources than their more active counterparts, irrespective of the actual conditions in the neighborhood. A second limitation is that the neighborhood-level constructs are measured on the basis of reports made by individuals, and although individual reports are undoubtedly influenced by objective reality, they are also influenced by personal factors and perceptions which may introduce measurement error. An alternative approach is to measure neighborhood conditions by incorporating information obtained from a separate sample of persons who reside in the same neighborhoods as the study participants. These persons can serve as informants of neighborhood conditions. Their responses can be aggregated to the neighborhood level and linked with the study population of interest in order to study relations between these neighborhood features and health outcomes.

In traditional psychometrics, the reliability of a scale is assessed on the basis of internal consistency and test-retest reliability. Assessing the measurement properties of ecologic measurements moves beyond an assessment of the psychometric properties to what has been termed "ecometrics" (24). Ecometrics is an extension of the two levels implicit in traditional psychometric assessment (scale item responses nested within individuals) because it introduces a third level: scale items nested within individuals who are nested within neighborhoods. It allows quantification not only of how consistently individuals respond to the different component items of a scale (the internal consistency measure of psychometrics) but also the extent to which residents of the same neighborhood rate their neighborhood similarly (24).

Few empirical studies have investigated the psychometric properties of survey measures of neighborhood constructs (16, 25, 26); still fewer have assessed their ecometric properties (24, 27). In our study, we had three primary objectives: 1) to develop neighborhood scales that represent features of neighborhoods potentially important for cardiovascular disease risk, 2) to assess the psychometric and ecometric properties of such scales, and 3) to examine how individual-level variables and neighborhood socioeconomic indicators are related to these scales. We focused on cardiovascular disease-related constructs because they are among the health outcomes most commonly examined in relation to neighborhood conditions (1, 2, 5, 7).

MATERIALS AND METHODS

Study population

Data were collected via a telephone survey of residents of selected census tracts in Baltimore City/County, Maryland; Forsyth County, North Carolina; and New York, New York, between January and August of 2004. These three areas were selected because they are the geographic areas from which participants in a cohort study of cardiovascular disease, the Multi-Ethnic Study of Atherosclerosis, were sampled (28). The main objective of the survey was to construct measures of neighborhood-level properties for these areas, using an independent sample of individuals as "informants," so that this information could be linked to Multi-Ethnic Study of Atherosclerosis participants in future analyses. Using random-digit dialing, we identified a sample of telephone numbers in the areas of interest. On the basis of prior work (17, 24, 29), we estimated a total desired sample size of 5,800 across the three sites, which would yield a mean number of 25 participants per neighborhood cluster (as defined below). One adult aged 18 years or older was randomly selected to participate within each sampled household. The survey was administered in English or Spanish. We surveyed 5,988 respondents (1,752 in Maryland, 1,616 in North Carolina, and 2,620 in New York). A sample of 120 persons (40 at each site) was reinterviewed 2–3 weeks after the initial interview for assessment of test-retest reliability. The final response rate, calculated using American Association for Public Opinion Research criteria (30), was 46.5 percent. The American Association for Public Opinion Research response rate for the test-retest reliability study was 80.0 percent.

Study questionnaire

The telephone questionnaire ascertained information on neighborhood-level dimensions relevant to cardiovascular disease. In responding to the questionnaire, participants were asked to refer to the area approximately 1 mile (1.6 km)

around their home. On the basis of a conceptual model (31) and prior work (25), seven neighborhood dimensions were assessed, including aesthetic quality (six items), walking environment (10 items), availability of healthy foods (four items), safety (three items), violence (four items), social cohesion (four items), and activities with neighbors (five items). Scale items were drawn from published work whenever possible (17, 20, 32–35). For most scales, responses for each item ranged from 1 to 5 (1 = strongly agree, 2 = agree, 3 = neutral (neither agree nor disagree), 4 = disagree, and 5 = strongly disagree). Responses for the scales on violence and activities with neighbors ranged from 1 to 4 (1 = often, 2 = sometimes, 3 = rarely, and 4 = never). For each scale, a score was estimated by taking the average across all items within the scale. Only respondents with complete information for all items within a scale were assigned a scale score. Because our neighborhood constructs were identified a priori on the basis of a conceptual framework, we retained the initial seven scales for analyses. Some items were dropped from three of our neighborhood scales because dropping these items improved the scales' internal consistency (Cronbach's alpha coefficient increased by 0.06–0.12). The items in each scale (including the items dropped) are shown in table 1.

Definition of neighborhood

Two alternate definitions of neighborhoods (census tracts and neighborhood clusters) were investigated. Census tracts are subdivisions of counties containing an average of 4,000 persons (36). Respondents in this study represented 576 census tracts (208 in Maryland, 71 in North Carolina, and 297 in New York), with a median of eight participants per tract (range: 1–62 participants). Neighborhood clusters are clusters of spatially contiguous block groups. For the New York and Maryland sites, the clusters contained 8,000–12,000 persons and encompassed approximately 10 census block groups (or 2–3 census tracts). Because North Carolina is less densely populated and the block groups are much larger in geographic size, clusters for North Carolina were defined to contain four block groups, on average. Clusters were constructed by aggregating spatially contiguous block groups with similar sociodemographic and housing characteristics using spatially constrained clustering (37, 38) and Boundary SEER software developed by TerraSeer, Inc. (Crystal Lake, Illinois; www.terraseer.com). This approach combines block groups based on spatial contiguity and minimization of the within-cluster sums of squares for the variables of interest (in our case, factor scores derived from demographic, socioeconomic, and housing characteristics) (37, 38). Clusters created using this data-driven approach were refined using natural boundaries, highways, and local knowledge. A total of 161 clusters were included in these analyses (51 in Maryland, 53 in North Carolina, and 57 in New York), with a median of 26 participants per cluster (range: 2–322 participants).

Statistical analysis

Data were weighted in all analyses to account for the sampling design and to correct for nonresponse. The psy-chometric properties of each scale were assessed using Cronbach's alpha coefficient and the 2-week test-retest reliability (39).

We assessed the ecometric properties of the neighborhood scales using three-level multilevel models (40). The level 1 model (item responses within individuals) was defined as:

$$\text{Level 1: } Y_{ijk} = b_{0jk} + e_{ijk},$$
$$e_{ijk} \sim N(0, \sigma^2),$$

where Y_{ijk} represents the ith response of person j in neighborhood k and b_{0jk} is the estimated mean scale score for person j in neighborhood k. The errors of measurement e_{ijk} of item i for person j in neighborhood k are assumed to be normally distributed with variance σ^2. In the level 2 model (persons within neighborhoods), the estimated mean scale score for person j in neighborhood k is modeled as a function of a neighborhood mean and a person-specific deviation:

$$\text{Level 2: } b_{0jk} = \lambda_{00k} + \alpha_{0jk},$$
$$\alpha_{0jk} \sim N(0, \tau_b),$$

where λ_{00k} represents the mean value for the neighborhood scale in neighborhood k and α_{0jk} represents a random effect for person j in neighborhood k. The random effect α_{0jk} is assumed to be normally distributed with mean 0 and variance τ_b. τ_b is the within-neighborhood variance, which quantifies the variability in the person-specific score within neighborhoods.

The level 3 model (neighborhoods) models the neighborhood-specific means as a function of an overall mean and a neighborhood-specific deviation:

$$\text{Level 3: } \lambda_{00k} = \gamma_{000} + U_{00k},$$
$$U_{00k} \sim N(0, \tau_\eta),$$

where γ_{000} represents the mean value for the neighborhood-level measures across neighborhoods and U_{00k} represents a random neighborhood effect for each neighborhood k. The random effect U_{00k} is assumed to be normally distributed with mean 0 and variance τ_η. τ_η is the between-neighborhood variance and represents the variability in neighborhood mean score across neighborhoods.

Using the above model, we calculated the intraneighborhood correlation coefficient (ICC) and the reliability of the neighborhood-level measure. The ICC quantifies the percentage of variability in the scale score that lies between neighborhoods (24). It is calculated as the ratio of the variance between neighborhoods divided by the sum of between- and within-neighborhood variance components. The ICC ranges from 0 to 1, with a higher value indicating greater agreement between respondents within a neighborhood. The neighborhood-level reliability of the neighborhood score (λ_{00k}) (24, 41) is a function of the ICC as well as the number of participants in each neighborhood (n_{jk}). It is calculated as the ratio of the "true" score variance to the observed score variance in the sample neighborhood mean, with values ranging from 0 to 1. The reliability will be high (close to 1) when: 1) the neighborhood means vary substantially across neighborhoods (holding constant the sample size per group) or 2) the sample size per neighborhood is large.

TABLE 1.   Neighborhood scale items included in a telephone survey on neighborhood conditions administered at three US study sites* (*n* = 5,988), 2004

Aesthetic quality

   1. There is a lot of trash and litter on the street in my neighborhood.†

   2. There is a lot of noise in my neighborhood.†

   3. In my neighborhood the buildings and homes are well-maintained.

   4. The buildings and houses in my neighborhood are interesting.

   5. My neighborhood is attractive.

   6. There are interesting things to do in my neighborhood.‡

Walking environment

   1. My neighborhood offers many opportunities to be physically active.

   2. Local sports clubs and other facilities in my neighborhood offer many opportunities to get exercise.

   3. It is pleasant to walk in my neighborhood.

   4. The trees in my neighborhood provide enough shade.

   5. In my neighborhood it is easy to walk places.

   6. I often see other people walking in my neighborhood.

   7. I often see other people exercising (for example, jogging, bicycling, playing sports) in my neighborhood.

   8. My neighborhood has heavy traffic.‡

   9. There are busy roads to cross when out for walks in my neighborhood.‡

  10. In my neighborhood it is easy to walk places.‡

Availability of healthy foods

   1. A large selection of fresh fruits and vegetables is available in my neighborhood.

   2. The fresh fruits and vegetables in my neighborhood are of high quality.

   3. A large selection of low-fat products is available in my neighborhood.

   4. There are many opportunities to purchase fast foods in my neighborhood.‡

Safety

   1. I feel safe walking in my neighborhood, day or night.

   2. Violence is not a problem in my neighborhood.

   3. My neighborhood is safe from crime.

Violence

  During the past 6 months, how often:

     1. …was there a fight in your neighborhood in which a weapon was used?

     2. …were there gang fights in your neighborhood?

     3. …was there a sexual assault or rape in your neighborhood?

     4. …was there a robbery or mugging in your neighborhood?

Social cohesion

   1. People around here are willing to help their neighbors.

   2. People in my neighborhood generally get along with each other.

   3. People in my neighborhood can be trusted.

   4. People in my neighborhood share the same values.

Activities with neighbors

   1. About how often do you and people in your neighborhood do favors for each other? By favors, we mean such things as watching each other's children, helping with shopping, lending garden or house tools, and other small acts of kindness.

   2. When a neighbor is not at home or on vacation, how often do you and other neighbors watch over their property?

   3. How often do you and other people in the neighborhood ask each other for advice about personal things such as child-rearing or job openings?

   4. How often do you and people in your neighborhood have parties or other get-togethers where other people in the neighborhood are invited?

   5. How often do you and other people in your neighborhood visit in each other's homes or speak with each other on the street?

---

* Selected census tracts in Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York.

† Reverse-coded.

‡ Item was dropped to increase the internal consistency of the scale.

$$\mathrm{ICC} = \frac{\tau_\eta}{\tau_\eta + \tau_b};$$

$$\mathrm{Reliability}(\lambda_{00k}) = \frac{\tau_\eta}{\tau_\eta + \left\{ \displaystyle\sum_1^J [\tau_b + \sigma^2/n_{jk}]^{-1} \right\}^{-1}}.$$

We also explored individual- and neighborhood-level predictors of the neighborhood scales. We investigated six individual-level covariates: age, gender, race/ethnicity, education, income, and duration of residence in the neighborhood. For this analysis, we estimated within-neighborhood effects of each of the six individual-level covariates by including them in the level 2 model as variables centered around the group (neighborhood) mean (42). This approach minimizes any confounding of individual-level effects by omitted neighborhood-level variables. We also used a three-level model (without any centering) to investigate associations of census tract percentage of poverty derived from the 2000 US Census with the scales after adjusting for measured individual-level factors. We estimated the percentage of between-neighborhood variability in scale scores explained by the poverty indicator using the formula

$$\frac{\tau_{00}(\mathrm{model\,X}) - \tau_{00}(\mathrm{model\,Y})}{\tau_{00}(\mathrm{model\,X})},$$

where $\tau_{00}(\mathrm{model\,X})$ is the between-neighborhood variance from the model with only level 2 covariates included and $\tau_{00}(\mathrm{model\,Y})$ is the between-neighborhood variance from the full model with both level 2 and level 3 covariates included.

To investigate the robustness of our results to violation of model assumptions at level 1, we also constructed three-level logistic models (with dichotomized responses) and ordinal logistic models. Decomposition of variance in these models is not straightforward, because the level 1 variance depends on the covariates and because the level 1 variance and the levels 2 and 3 variances are on different scales (43, 44). In the estimation of neighborhood reliabilities from these models, we used the standard approximation to the level 1 variance used in HLM software (Scientific Software International, Lincolnwood, Illinois). In a preliminary examination of the convergent validity of the scales, we examined the correlations between the observed neighborhood means for each domain. Our expectation was that positive aspects of neighborhoods (walkability, good aesthetic quality, safety, low violence, access to healthy foods) would tend to cluster.

RESULTS

Fifty-four percent of the study sample was female. The mean age was 43.8 years (standard deviation, 17.1), and the mean duration of residence in the neighborhood was 12.9 years (standard deviation, 13.6). The sample was diverse in terms of socioeconomic characteristics and race/ethnicity and was approximately representative of the areas from which it was drawn, although respondents were slightly more likely than the total population to be in the higher educational

TABLE 2. Selected characteristics of respondents to a telephone survey on neighborhood conditions ($n = 5,988$) compared with those of the total population of the census tracts* from which the survey sample was drawn, 2004

Characteristic	Survey		Geographic areas in the sampling frame† (%)
	No. of subjects	Weighted %	
Study site			
Baltimore, Maryland	1,746	29.3	N/A‡
Forsyth County, North Carolina	1,615	26.7	N/A
New York, New York	2,627	44.0	N/A
Age (years)			
<65	5,014	87.5	88.4
≥65	974	12.6	11.6
Race/ethnicity			
White	3,140	34.7	33.5
African-American	1,711	29.8	33.5
Hispanic	788	25.9	28.0
Asian	127	4.1	2.8
Other	183	4.9	1.0
Unknown	39	0.6	
Education			
Less than high school diploma	735	17.1	29.9
High school graduation/ some college	2,536	42.5	42.6
College graduation or more	2,704	40.2	27.5
Unknown	13	0.2	
Annual income			
$0–$49,999	2,991	53.3	66.0
≥$50,000	2,287	34.4	34.0
Unknown	710	12.3	

* Selected census tracts in Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York.
† Derived from the 2000 US Census.
‡ N/A, not applicable.

categories (table 2). Descriptive statistics for the scales are shown in table 3. Cronbach's alpha coefficient ranged from 0.73 (walking environment) to 0.83 (violence). Test-retest reliability coefficients (for the subset ($n = 120$)) were also high, ranging from 0.60 (walking environment) to 0.88 (safety).

The ecometric properties of the scales are shown in table 4. Neighborhood reliabilities were 0.64 or more for census tracts and 0.78 or more for census clusters. The "activities with neighbors" scale showed reliabilities substantially lower than those for the other scales (0.28 for tracts and 0.46 for clusters). The ICCs were also low for the "activities with neighbors" scale (only 0.06 for tracts and 0.05 for clusters) but were substantially higher for the other scales (0.28–0.51 for tracts and 0.22–0.45 for clusters). Similar results for ICC and neighborhood reliability were obtained

TABLE 3.   Descriptive statistics for seven scales on neighborhood conditions assessed by telephone survey at three US study sites* (*n* = 5,988), 2004†

Scale	No. of subjects	No. of items in scale	Range of scores	Mean score	Standard deviation	Cronbach's alpha	Test-retest correlation‡	95% confidence interval
Aesthetic quality	5,879	5	1–5	3.3	0.8	0.75	0.83	0.77, 0.88
Walking environment	5,732	7	1–5	3.6	0.7	0.73	0.60	0.47, 0.71
Availability of healthy foods	5,774	3	1–5	3.4	1.0	0.78	0.69	0.57, 0.77
Safety	5,803	3	1–5	3.2	1.0	0.77	0.88	0.83, 0.91
Violence	4,942	4	1–4	1.8	0.8	0.83	0.72	0.62, 0.80
Social cohesion	5,436	4	1–4	2.6	0.8	0.74	0.65	0.53, 0.74
Activities with neighbors	5,477	5	1–5	3.4	0.8	0.78	0.73	0.63, 0.80

* Selected census tracts in Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York.

† The percentage of respondents with missing data was less than 5% of the full sample (*n* = 5,988) for aesthetic quality, walking environment, availability of healthy foods, and safety and less than 10% for social cohesion and activities with neighbors. The violence scale had 17.5% missing data. Missing data were generally due to the participant's indicating that he/she did not know whether the condition applied to his/her neighborhood. This was especially common for the violence scale, because participants were asked to report on specific events (see table 1). Participants excluded because of missing data on any scale were slightly more likely to be older, female, non-White, and from New York.

‡ Test-retest correlation for a sample of 120 participants in the test-retest reliability study.

when logistic or ordinal logistic regression was used. For example, for aesthetic quality, ICCs for census tracts were 0.51, 0.48, and 0.48 and neighborhood reliabilities were 0.78, 0.77, and 0.73 for three-level normal, ordinal logistic, and logistic models, respectively. Site-specific models (not shown) revealed no clear site differences in ICC or reliability, although North Carolina tended to consistently have the lowest values.

Correlations between the seven neighborhood scales (not shown) generally indicated good convergent validity. For example, high correlations in the expected direction at the census tract level (with the score being estimated by the mean for all respondents in the tract) were observed for safety and violence (–0.68), aesthetic quality and safety (0.72), and social cohesion and safety (0.72). The "activities with neighbors" scale was positively correlated with social cohesion (0.43) but was largely uncorrelated with the other measures.

There was some evidence that reports of neighborhood characteristics varied according to the individual-level characteristics of the respondents (table 5). Older participants were significantly more likely to report better aesthetic quality, better availability of healthy foods, more safety and social cohesion, and less violence. Black participants were more likely to report higher levels of aesthetic quality, walking environment, and safety and to report lower levels of violence. Hispanic participants reported higher levels of aesthetic quality, walking environment, healthy foods, and safety and lower levels of violence. Higher-income persons reported higher aesthetic quality, walkability, safety, social cohesion, and activities with neighbors.

After controlling for the individual-level characteristics of respondents, neighborhood poverty was significantly associated with more violence, with poorer aesthetic quality, walking environment, and availability of healthy foods, and with less safety and social cohesion (table 6). Neighborhood percentage of poverty explained a large amount of the variability across neighborhoods in aesthetic quality, safety, violence, and social cohesion (54.6–67.1 percent) and moderate amounts of differences in walkability and availability of healthy foods (15.8–26.5 percent). Neighborhood poverty was not associated with the "activities with neighbors" scale.

DISCUSSION

The purpose of our study was to examine the measurement properties of scales designed to assess selected neighborhood-level characteristics. Our neighborhood scales had good psychometric properties with high internal consistency (range of Cronbach's alpha: 0.73–0.83) and test-retest reliability (range: 0.60–0.88). A pilot study conducted by Echeverria et al. (25) reported slightly higher Cronbach's alpha coefficients (range: 0.77–0.94) and test-retest reliabilities (range: 0.78–0.91) for several similar neighborhood scales. However, the pilot study was conducted in a small volunteer sample of persons residing in only one location.

Despite much recent interest in neighborhood health effects, few studies have assessed the ecometric properties of neighborhood measures (24, 27). Assessment of these properties is important for new data collection and for the interpretation of study results. With the exception of the "activities with neighbors" scale, our measures showed good ecometric properties, with neighborhood reliabilities ranging from 0.76 to 0.88 (for neighborhoods defined as census tracts) and ICCs ranging from 0.24 to 0.46 (for neighborhoods defined as census clusters). There is a trade-off between the reliability and the ICC of our neighborhood scales depending on the size of the area used to define a neighborhood. Smaller areas ("census tracts") have higher ICCs than larger geographic areas ("census clusters"). Larger geographic areas are likely to contain more heterogeneity in the characteristics being assessed, leading to less agreement among participants within a neighborhood. Similar to the ICC, the reliability of the neighborhood measures is a function of the between- and within-neighborhood variances. However, it is also positively related to the number of individuals within each neighborhood. On average,

TABLE 4. Variance components, intraneighborhood correlation coefficients, and neighborhood-level reliabilities for seven scales on neighborhood conditions assessed by telephone survey at three US study sites* (n = 5,988), 2004

| | Scale | | | | | | | | | | | | |
| | Aesthetic quality | | Walking environment | | Availability of healthy foods | | Safety | | Violence | | Social cohesion | | Activities with neighbors | |
	Census tracts	Census clusters	Census tracts	Census clusters	Census tracts	Census clusters	Census tracts	Census clusters	Census tracts	Census clusters	Census tracts	Census clusters	Census tracts	Census clusters
Within-person variance	0.81	0.81	0.85	0.85	0.64	0.64	0.64	0.64	0.40	0.40	0.59	0.59	0.65	0.65
Within-neighborhood variance	0.25	0.28	0.18	0.20	0.52	0.55	0.38	0.41	0.31	0.33	0.28	0.29	0.39	0.39
Between-neighborhood variance	0.26	0.23	0.14	0.10	0.21	0.16	0.35	0.35	0.18	0.18	0.14	0.15	0.03	0.02
Intraneighborhood correlation	0.51	0.45	0.43	0.33	0.28	0.22	0.48	0.45	0.37	0.35	0.34	0.34	0.06	0.05
Neighborhood reliability†	0.78	0.89	0.73	0.83	0.64	0.78	0.77	0.89	0.72	0.86	0.68	0.84	0.28	0.46

* Selected census tracts in Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York.
† Estimates for neighborhood reliability correspond to the average reliability across all neighborhoods.

neighborhood clusters will have more respondents because of their larger geographic size; hence, reliabilities are higher for clusters than for census tracts, despite the higher ICCs of the latter. The lower ICCs for larger areas are also consistent with the lower ICCs observed for the North Carolina site (not shown), where census tracts are substantially larger in size and hence may be more heterogeneous. The ecometric properties of neighborhood measures may vary depending on the size of the geographic areas assessed and the geographic variability in the construct of interest. Unfortunately, sample size limitations precluded detailed investigation of this regional variability in our data.

Raudenbush and Sampson (24) reported similar reliabilities (range: 0.74–0.89) and ICCs (range: 0.13–0.39) for neighborhood measures of perceived violence, neighborhood decline, social cohesion, social disorder, and social control with neighborhoods defined similarly to our census clusters. Our decision to construct neighborhood clusters and sample 25–30 participants per cluster was based in part on these results, which indicated that an intraneighborhood sample size of 25–30 will maximize the neighborhood reliability (24). However, in our analyses, both ICCs and reliabilities did not differ substantially for census tracts and census clusters. Thus, in our data, the loss of reliability resulting from the use of census tracts with smaller sample sizes is offset by the greater homogeneity of these smaller areas in terms of the characteristics of interest and by the large number of "neighborhoods" available for analysis when census tracts are used. Having a larger number of neighborhoods has the added advantage of increasing power to detect neighborhood health effects (45).

Census tracts and census clusters obviously do not match up exactly with the geographic areas respondents were asked to refer to in the survey (approximately 1 mile (1.6 km) around their homes). Moreover, it is possible that participants responded with regard to the area they intuitively thought of as their "neighborhood," despite the survey instructions. It is also unrealistic to think that the dimensions we measured vary significantly across the arbitrarily defined geographic barriers of census tracts or clusters. Despite these complexities, our data show that an important part of the variability in neighborhood scores is between our arbitrarily defined "neighborhoods," supporting the utility of our instrument in the measurement of true area-level constructs. However, there is also evidence of variation in responses within neighborhoods. Part of this may be due to the arbitrary geographic definition of "neighborhoods" that we used. Other sources of within-neighborhood variability include variations induced by the necessarily subjective nature of the reports, as well as simple measurement error. The presence of this within-neighborhood variability is an argument for averaging over respondents or raters in estimating the "true" neighborhood characteristic. The relatively high neighborhood reliability estimates indicate that the mean is a reasonable estimate for the true neighborhood score, although there is still room for improvement, especially in areas with small sample sizes.

Of the scales we examined, the "activities with neighbors" scale clearly had poor ecometric properties, despite having good psychometric properties. There was marked

TABLE 5.   Mean difference in neighborhood scale scores according to individual-level factors† at three US study sites‡ (n = 5,988), 2004

Individual-level predictor variable	Scale						
	Aesthetic quality	Walking environment	Availability of healthy foods	Safety	Violence	Social cohesion	Activities with neighbors
Age (per 10 years)	0.07 (0.01)**,§	0.00 (0.01)	0.06 (0.01)**	0.03 (0.01)**	−0.06 (0.01)**	0.06 (0.01)**	−0.01 (0.01)
Gender							
Male¶							
Female	0.04 (0.02)	0.02 (0.02)	0.07 (0.03)*	−0.09 (0.03)**	−0.04 (0.02)*	0.02 (0.02)	−0.01 (0.02)
Race/ethnicity							
White¶							
Black	0.27 (0.04)**	0.12 (0.03)**	0.04 (0.04)	0.25 (0.04)**	−0.18 (0.03)**	0.04 (0.03)	−0.04 (0.03)
Hispanic	0.24 (0.05)**	0.08 (0.04)*	0.15 (0.05)*	0.15 (0.05)*	−0.19 (0.04)**	−0.01 (0.04)	−0.09 (0.05)
Asian	0.04 (0.06)	0.03 (0.05)	0.06 (0.09)	0.20 (0.07)*	−0.19 (0.06)**	−0.01 (0.06)	−0.17 (0.08)*
Other	0.18 (0.07)**	0.02 (0.06)	0.00 (0.08)	−0.08 (0.08)	−0.14 (0.06)*	−0.13 (0.06)*	−0.14 (0.07)*
Education							
Less than high school diploma¶							
High school diploma	0.04 (0.04)	−0.03 (0.04)	0.01 (0.05)	0.06 (0.05)	−0.01 (0.04)	−0.03 (0.05)	−0.01 (0.04)
Some college	0.01 (0.04)	−0.02 (0.04)	0.01 (0.05)	0.03 (0.05)	0.00 (0.04)	0.02 (0.04)	−0.01 (0.04)
College graduation or more	0.01 (0.04)	−0.05 (0.04)	−0.08 (0.05)	0.07 (0.05)	0.04 (0.04)	−0.03 (0.05)	0.00 (0.04)
Annual income							
<$25,000¶							
$25,000–$49,999	0.02 (0.03)	0.03 (0.03)	−0.08 (0.04)	0.06 (0.04)	−0.05 (0.03)	0.09 (0.03)*	0.04 (0.03)
≥$50,000	0.13 (0.03)**	0.14 (0.03)**	0.01 (0.04)	0.17 (0.04)**	−0.06 (0.03)	0.22 (0.03)**	0.19 (0.03)**
Time (per 10 years)	−0.01 (0.01)	0.01 (0.01)	0.02 (0.01)*	−0.03 (0.01)*	0.04 (0.01)**	0.02 (0.01)	0.07 (0.01)**

* $p < 0.05$; ** $p < 0.001$.
† Derived from a three-level model including all individual-level variables centered around the group (neighborhood) mean (40).
‡ Selected census tracts in Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York.
§ Numbers in parentheses, standard error.
¶ Reference category.

TABLE 6.   Mean difference in neighborhood scale scores according to neighborhood (census tract) poverty† at three US study sites‡ (n = 5,988), 2004

Variable	Scale						
	Aesthetic quality	Walking environment	Availability of healthy foods	Safety	Violence	Social cohesion	Activities with neighbors
% poverty§ (neighborhood-level predictor)	−0.28 (0.01)*,¶	−0.13 (0.01)*	−0.15 (0.02)*	−0.30 (0.01)*	0.22 (0.01)*	−0.17 (0.01)*	−0.02 (0.01)*
Variance							
Within persons	0.80	0.86	0.64	0.62	0.40	0.58	0.63
Within neighborhoods	0.24	0.18	0.50	0.38	0.29	0.27	0.38
Between neighborhoods	0.08	0.07	0.15	0.10	0.08	0.03	0.02
Percentage of between-neighborhood variance explained	67.1	26.5	15.8	65.4	54.6	60.7	1.7

* $p < 0.001$.
† Derived from three-level models including neighborhood poverty and all of the individual-level variables shown in the table. None of the variables were centered around the group (neighborhood) mean. Therefore, the poverty effect is adjusted for differences in neighborhood composition.
‡ Selected census tracts in Baltimore, Maryland; Forsyth County, North Carolina; and New York, New York.
§ Estimates correspond to a 10% increase in the percentage of census tracts below the federal poverty level.
¶ Numbers in parentheses, standard error.

Am J Epidemiol 2007;165:858–867

heterogeneity in the ways in which residents of the same "neighborhood" responded to these items. This indicates that this measure may be tapping into an individual-level construct as opposed to a true neighborhood-level construct.

A preliminary indication of the convergent validity of our measures is that they correlated with each other in the expected directions. We also saw strong associations between neighborhood poverty and all but one of the neighborhood scales. Neighborhood poverty also explained large amounts of the between-neighborhood variability in our scales. These findings were expected and are consistent with the use of aggregate measures of neighborhood SEP as proxies for specific neighborhood conditions potentially relevant to cardiovascular disease risk. However, there is also evidence of neighborhood variations in these constructs which are not captured by neighborhood SEP. Thus, the direct measurement of these features provides information which is correlated with, but also distinct from, neighborhood SEP.

In using area-level aggregates of survey responses to characterize neighborhoods, respondents in each neighborhood are viewed as informants of the conditions in their area. To the extent that people's perceptions reflect reality, the averaging of responses across multiple persons within a neighborhood reduces measurement error due to individual subjectivity. In our data, there was some evidence of within-area differences in reported neighborhood characteristics associated with individual-level characteristics, including race/ethnicity and income. Other studies have also documented variation in neighborhood constructs based on individual characteristics (17, 46). If there is systematic variability in the way that respondents rate their neighborhood based on sociodemographic characteristics, these variables can be incorporated into the level 2 model in order to derive adjusted estimates of the neighborhood-level construct of interest. However, the presence of associations of the scales with individual-level variables does not necessarily imply systematic differences in the ways in which respondents rate their neighborhoods. It is possible that reporting differences reflect real within-area differences. For example, income groups may be spatially clustered within census tracts or clusters, with some groups living in areas with poorer environmental conditions or near the boundaries of the tract and in close proximity to areas with poor environmental conditions.

One limitation of our data is the wide range in the number of participants in each area. The number of survey respondents ranged from 1 to 62 in each census tract and from 2 to 322 in each census cluster. For some neighborhoods, measures are based on the responses of a few participants or, in extreme cases, only one participant. One approach to dealing with this problem is to construct empirical Bayes estimates which borrow strength across neighborhoods and shrink estimates for neighborhoods with few observations towards the overall mean (40, 47, 48). Future research is needed to examine the consequences of using simple means or empirical Bayes estimates as predictors of health outcomes.

Despite repeated calls for the study of the environment as a main effect and in interaction with individual-level factors in epidemiology, the assessment of ecologic settings remains in its infancy. In this paper, we have demonstrated the feasibility of measuring constructs that vary over geographic areas using survey data and have shown data supporting the validity and reliability of these measures. Improving the measurement of environmental and group-level factors is a prerequisite for investigating their causal effects.

REFERENCES

1. Diez-Roux AV, Nieto FJ, Muntaner C, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. Am J Epidemiol 1997;146:48–63.
2. Smith G, Hart C, Watt G, et al. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: The Renfrew and Paisley Study. J Epidemiol Community Health 1998;52:399–405.
3. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health 2001;55:111–22.
4. Bosma H, van de Mheen HD, Borsboom GJ, et al. Neighborhood socioeconomic status and all-cause mortality. Am J Epidemiol 2001;153:363–71.
5. Diez Roux AV, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. N Engl J Med 2001;345:99–106.
6. Borrell LN, Diez Roux AV, Rose K, et al. Neighbourhood characteristics and mortality in the Atherosclerosis Risk in Communities Study. Int J Epidemiol 2004;33:398–407.
7. Sundquist K, Winkleby M, Ahlen H, et al. Neighborhood socioeconomic environment and incidence of coronary heart disease: a follow-up study of 25,319 women and men in Sweden. Am J Epidemiol 2004;159:655–62.
8. O'Campo P. Invited commentary: advancing theory and methods for multilevel models of residential neighborhoods and health. Am J Epidemiol 2003;157:9–13.
9. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. Soc Sci Med 2004;58:1929–52.
10. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. Soc Sci Med 2004;58:1953–60.
11. Giles-Corti B, Donovan RJ. The relative influence of individual, social and physical environment determinants of physical activity. Soc Sci Med 2002;54:1793–812.
12. Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: The Atherosclerosis Risk in Communities Study. Am J Public Health 2002;92:1761–7.
13. Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. Epidemiol Rev 1999;21:143–61.

14. Richards TB, Croner CM, Rushton G, et al. Geographic information systems and public health: mapping the future. Public Health Rep 1999;114:359–60.
15. Morenoff JD. Neighborhood mechanisms and the spatial dynamics of birth weight. AJS 2003;108:976–1017.
16. Pikora TJ, Bull FC, Jamrozik K, et al. Developing a reliable audit instrument to measure the physical environment for physical activity. Am J Prev Med 2002;23:187–94.
17. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 1997;277:918–24.
18. Buka SL, Brennan RT, Rich-Edwards JW, et al. Neighborhood support and the birth weight of urban infants. Am J Epidemiol 2003;157:1–8.
19. Troped PJ, Saunders RP, Pate RR, et al. Associations between self-reported and objective physical environmental factors and use of a community rail-trail. Prev Med 2001; 32:191–200.
20. Stahl T, Rutten A, Nutbeam D, et al. The importance of the social environment for physically active lifestyle—results from an international study. Soc Sci Med 2001;52:1–10.
21. Balfour JL, Kaplan GA. Neighborhood environment and loss of physical function in older adults: evidence from the Alameda County Study. Am J Epidemiol 2002;155: 507–15.
22. Wilcox S, Bopp M, Oberrecht L, et al. Psychosocial and perceived environmental correlates of physical activity in rural and older African American and white women. J Gerontol B Psychol Sci Soc Sci 2003;58:P329–37.
23. Gomez JE, Johnson BA, Selva M, et al. Violent crime and outdoor physical activity among inner-city youth. Prev Med 2004;39:876–81.
24. Raudenbush SW, Sampson RJ. Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. Sociol Methodol 1999;29:1–41.
25. Echeverria SE, Diez-Roux AV, Link BG. Reliability of self-reported neighborhood characteristics. J Urban Health 2004; 81:682–701.
26. Giles-Corti B, Timperio A, Cutt H, et al. Development of a reliable measure of walking within and outside the local neighborhood: RESIDE's Neighborhood Physical Activity Questionnaire. Prev Med 2006;42:455–9.
27. Gauvin L, Richard L, Craig CL, et al. From walkability to active living potential: an ''ecometric'' validation study. Am J Prev Med 2005;28(suppl 2):126–33.
28. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol 2002;156:871–81.
29. Raudenbush SW, Sampson RJ. Assessing direct and indirect effects in multilevel designs with latent variables. Sociol Methods Res 1999;28:123–53.
30. American Association for Public Opinion Research. Standard definitions: final definitions of case codes and outcome rates for surveys. 3rd ed. Lenexa, KS: American Association for Public Opinion Research, 2004.
31. Diez Roux AV. Residential environments and cardiovascular risk. J Urban Health 2003;80:569–89.
32. King AC, Castro C, Wilcox S, et al. Personal and environmental factors associated with physical inactivity among different racial-ethnic groups of U.S. middle-aged and older-aged women. Health Psychol 2000;19:354–64.
33. Booth ML, Owen N, Bauman A, et al. Social-cognitive and perceived environment influences associated with physical activity in older Australians. Prev Med 2000;31:15–22.
34. Ball K, Bauman A, Leslie E, et al. Perceived environmental aesthetics and convenience and company are associated with walking for exercise among Australian adults. Prev Med 2001;33:434–40.
35. Giles-Corti B, Donovan RJ. Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. Prev Med 2002;35:601–11.
36. Bureau of the Census, US Department of Commerce. 1980 census of population: classified index of industries and occupations. Washington, DC: US GPO, 1982.
37. Legendre P. Constrained clustering. In: Legendre P, Legendre L, eds. Developments in numerical ecology. (NATO ASI series, series G: ecological sciences, vol 14). Berlin, Germany: Springer Publishing Company, 1988:289–307.
38. Legendre P, Fortin MJ. Spatial pattern and ecological analysis. Vegetatio 1989;80:107–38.
39. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;2:420–8.
40. Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc, 2002.
41. Raudenbush SW, Bryk AS. Three-level models. In: Hierarchical linear models: applications and data analysis methods. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc, 2002: 228–51.
42. Bingenheimer JB, Raudenbush SW. Statistical and substantive inferences in public health: issues in the application of multilevel models. Annu Rev Public Health 2004;25:53–77.
43. Goldstein H, Browne W, Rasbash J. Partitioning variation in multilevel models. Understanding Stat 2002;1:223–31.
44. Raudenbush SW, Bryk AS. Hierarchical generalized linear models. In: Hierarchical linear models: applications and data analysis methods. 2nd ed. Thousand Oaks, CA: Sage Publications, Inc, 2002:291–335.
45. Snijders TAB, Bosker RJ. Multilevel analysis: an introduction to basic and advanced multilevel modeling. Thousand Oaks, CA: Sage Publications, Inc, 1999.
46. Subramanian SV, Lochner KA, Kawachi I. Neighborhood differences in social capital: a compositional artifact or a contextual construct? Health Place 2003;9:33–44.
47. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43:671–81.
48. Thomas N, Longford NT, Rolph JE. Empirical Bayes methods for estimating hospital-specific mortality rates. Stat Med 1994;13:889–903.