Published in final edited form as:

J Epidemiol Community Health. 2012 April ; 66(4): 342–351. doi:10.1136/jech.2009.103325.

Reliability of perceived neighborhood conditions and the effects of measurement error on self-rated health across urban and rural neighborhoods

Sandi L. Pruitt, Ph.D., M.P.H.^{1,*}, Donna B. Jeffe, Ph.D.^{1,2}, Yan Yan, M.D., Ph.D.³, and Mario Schootman, Ph.D.^{1,2}

¹ Washington University School of Medicine, Department of Medicine, Division of Health Behavior Research, Saint Louis, MO U.S.A

² Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine, Saint Louis, MO U.S.A

³ Washington University School of Medicine, Departments of Surgery and Biostatistics, Saint Louis, MO U.S.A

Abstract

Background—Limited psychometric research has examined the reliability of self-reported measures of neighborhood conditions, the effect of measurement error on associations between neighborhood conditions and health, and potential differences in the reliabilities between neighborhood strata (urban vs. rural and low vs. high poverty).

We assessed overall and stratified reliability of self-reported perceived neighborhood conditions using 5 scales (*Social and Physical Disorder*, *Social Control*, *Social Cohesion*, *Fear*) and 4 single items (*Multidimensional Neighboring*). We also assessed measurement error-corrected associations of these conditions with self-rated health.

Methods—Using random-digit dialing, 367 women without breast cancer (matched controls from a larger study) were interviewed twice, 2–3 weeks apart. We assessed test-retest (intraclass correlation coefficients [ICC]/weighted kappa [k]) and internal consistency reliability (Cronbach's α). Differences in reliability across neighborhood strata were tested using bootstrap methods. Regression calibration corrected estimates for measurement error.

Results—All measures demonstrated satisfactory internal consistency ($\alpha \geq .70$) and either moderate (ICC/k = .41–.60) or substantial (ICC/k = .61–.80) test-retest reliability in the full sample. Internal consistency did not differ by neighborhood strata. Test-retest reliability was significantly lower among rural (vs. urban) residents for 2 scales (*Social Control*, *Physical Disorder*) and 2 *Multidimensional Neighboring* items; test-retest reliability was higher for *Physical Disorder* and lower for 1 item *Multidimensional Neighboring* item among the high (vs. low) poverty strata.

*Address for correspondence: Washington University School of Medicine, Department of Medicine, Division of Health Behavior Research, Box 8504, 4444 Forest Park Avenue, Saint Louis, MO 63108, Phone: 314-286-1921, Fax: 314-286-1919, spruitt@dom.wustl.edu.

Publisher's Disclaimer: LICENCE STATEMENT

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in JECH editions and any other BMJ PGL products to exploit all subsidiary rights, as set out in our licence (<http://jech.bmj.com/ifa/licence.pdf>).

COMPETING INTERESTS

Competing Interest: None declared.

After measurement error correction, the magnitude of associations between neighborhood conditions and self-rated health were larger, particularly in the rural population.

Conclusion—Research is needed to develop and test reliable measures of perceived neighborhood conditions relevant to the health of rural populations.

Keywords

Neighborhood; Psychometrics; Reliability and Validity; Urbanization; Rural Population

INTRODUCTION

Significant evidence spanning the multiple and diverse disciplines of public health, criminology, and urban sociology has demonstrated consistent associations between the social and physical environment of local residential neighborhoods and a range of individual outcomes.[1–7] In the public health and medical literature, research findings suggest relationships between neighborhood disadvantage and poorer physical and mental health.[4, 8] Several methods are available for the assessment of neighborhood conditions. Existing sources of administrative data such as U.S. Census data yield objective assessments. Systematic assessments of neighborhoods by independent, trained observers and self-reported perceived neighborhood conditions by neighborhood residents may yield more subjective assessments of neighborhood conditions. A growing body of research has focused on the unique role of residents' perceived neighborhood conditions such as social and physical disorder. Self-reported measures can capture conditions such as satisfaction and relations with neighbors and neighborhoods, perceived safety, and neighborhood problems, [6, 9, 10] which are not captured by more objective measures or by outsider observers. Several studies even have shown that, compared to other assessment methods, perceived neighborhood conditions demonstrate equal or even greater associations with self-rated health [11–13] and other self-reported health outcomes [14, 15]. Given the use of both self-reported exposure and outcome data, however, the extent to which these stronger associations represent same-source bias is unclear. Regardless, to date, research indicates that both perceived and more objectively measured neighborhood conditions play important, if somewhat differing, roles in health. The extent to which perceptions or reality are stronger predictors of health likely depends on the measured health outcomes or neighborhood conditions—for example, perceived fear or safety may be particularly important for outdoor exercise behaviors or for mental health.

Although increasingly used, a rather limited body of research has addressed the psychometric properties of perceived neighborhood conditions.[16–22] Accordingly, the reliability of several validated and widely used measures of perceived neighborhood conditions remains unexplored. These measures, including *Perceived Disorder* and *Collective Efficacy*, have been consistently linked with a variety of individual outcomes (e.g. self-rated health, violent crime, depression, mistrust, chronic conditions) across a multidisciplinary literature.[6, 14, 23–26].

Reliability is a necessary precursor for validity. Low reliability may affect the performance and interpretation of a measure, contribute to participant misclassification, and attenuate associations with other variables in epidemiological research. Specifically, low reliability may result in biased measures of association, reflecting the influence of measurement error rather than the underlying true association. However, little is known about the extent to which measurement error as a result of low reliability of perceived neighborhood conditions affects observed associations.

Many commonly measured perceived neighborhood conditions were initially developed in the urban sociology and/or criminology literature and while the scales are more widely used today, they are still predominantly used in higher poverty, urban populations.[4, 6, 27] While there is some evidence that residents from suburban or less densely populated areas may not view neighborhood conditions the same as respondents from inner-city or more densely populated metro areas,[28, 29] the reliability and relevance of commonly used perceived neighborhood condition measures in rural populations remains largely unknown. In the event that measured conditions are less relevant in rural residents' lives (e.g., the problem of graffiti), rural residents may be more likely to guess on these items. Guessing may result in greater measurement error in rural areas, leading to lower internal consistency and test-retest reliability, and possibly attenuating associations between neighborhood conditions and outcome variables.

The limited data on the reliability of perceived neighborhood conditions, especially in rural areas, and the lack of assessment of the potential role of measurement error in these measures, represents a major limitation in the neighborhood effects literature. As existing measures continue to be integrated into a wider literature and new measures of self-reported conditions are developed, it will be important to ensure they are reliable in both urban and rural populations as well as in low vs. high poverty areas. Therefore, we focused on the reliability assessment of more general social and physical neighborhood characteristics using some of the most widely used and broadly relevant measures available. We assessed overall and stratified (urban vs. rural and low vs. high poverty) test-retest and internal consistency reliability of perceived neighborhood conditions and we assessed the role of measurement error in associations of these conditions with self-rated health among urban/rural and low/high poverty residents.

METHODS

Study sample

Data were collected via telephone interview as part of a study of Missouri residents examining the effect of adverse neighborhood conditions on quality of life among women with breast cancer and an age-race-geography-matched, random-digit-dialed sample of women without breast cancer. The latter group of women without breast cancer was interviewed for this test-retest study. The Washington University School of Medicine institutional review board approved the study.

Data collection

All participants during the last 9 month period of the larger study who, during their first telephone interview, stated that they were interested in participating in a follow-up study (98%) were eligible to be re-interviewed for the retest study. The second interview took place 2–3 weeks after the first interview with a different interviewer. This time frame, often used in test-retest studies, is long enough to ensure that respondents are unlikely to remember their original answers, yet short enough that changes in perceptions or underlying neighborhood conditions are unlikely. Up to 15 call attempts, distributed across weekdays, weeknights, and weekends, were made.

The Time 1 interviews obtained information about sociodemographic characteristics; both interviews collected information about street address, length of residence, and perceived neighborhood conditions. Street addresses were geocoded and participants were linked to census tracts. Urban/rural status and poverty rate were included as indicators of objective neighborhood conditions and were measured at the census-tract level using 2000 U.S. Census data. Poverty rate was selected because it is a robust indicator of socioeconomic status

across levels of geography and time, has been associated with various health outcomes, and has relevance for policymakers.[30, 31] Area poverty rate was defined as the percent living in poverty in the residents' census tract and dichotomized at the median: $<9.14\%$ or $\geq 9.14\%$. Rural/urban status was defined using Rural Urban Continuum Area (RUCA) Codes of the residents' census tract. The RUCA distinguish metropolitan (codes 0–3) from non-metropolitan (codes 4–10) census tracts using measures of daily commuting connectivity, urbanization, and population size and density.[32]

We measured multiple perceived neighborhood conditions using 5 multi-item scales and 4 single item measures. All measures referred to participants' residential neighborhood; the exact boundaries of those neighborhoods were left up to the interpretation of the individual. Perceived *Neighborhood Disorder* was measured with the Ross-Mirowsky 15-item 4-point Likert-type scale consisting of two subscales, *Social Disorder* (n=9 items) and *Physical Disorder/Decay* (n=6 items). Neighborhood disorder captures the extent to which residents' perceive cues in their neighborhoods signaling the breakdown of social control.[10] Collective Efficacy was measured with two subscales, informal *Social Control*, measured with 5 5-point Likert-type items, and *Social Cohesion and Trust*, measured with 5 4-point Likert-type items. *Social Control* captures the resident's perceptions of the community's ability to control the behavior of inhabitants and *Social Cohesion* measures the resident's perceptions of the community's ability to organize effectively to promote positive and repel negative influences.[6] *Neighborhood Fear* was measured with the sum of 3 items capturing the number of days in the previous week (0–7) during which respondents feared violent or criminal activities or were afraid to leave their home.[7] Lastly, we used 4 of the 14 items of the Multidimensional Measure of Neighboring capturing dimensions of social life within neighborhoods. We selected 3 items from the "Neighborhood attachment" subscale and one of the two items capturing the "weak social ties" subscale because they capture unique aspects of neighboring not captured in the other, more widely studied scales used in this study. Items not selected demonstrated considerable conceptual overlap with other measured scales, particularly the *Social Cohesion* and *Social Disorder* scales. Three of the four items were measured with a 5-point Likert-type scale and one item used a 6-category response option to measure the number of neighbors with which one socializes.[27] The four items were not summed and used as a scale because no data exist to support the validity or internal consistency reliability of a scale using only these 4 items.

We measured one health outcome, self-rated health, a valid and reliable measure of health status used widely in the neighborhood effects literature.[23] Self-rated health was measured at Time 1 with the question, "How would you rate your health—would you say it is excellent, very good, good, fair, or poor?" Similar to other studies,[33–35] we contrasted persons who answered "fair" or "poor" to persons who answered "excellent," "very good", or "good."

Demographic and socio economic covariates were measured at Time 1 and include the following: age (25–44, 45–54, 55–64, 65–74, ≥ 75), race (non-Hispanic white, non-Hispanic black, or Hispanic/other/multiple), education ($\leq$ high school, some college, $\geq$ college graduate), household income ($\leq$ \$24,999, 25–34,999, 35–49,999, 50–74,999, $\geq$ 75,000, missing), marital status (married/living together vs. unmarried), employment status (employed, retired, unable to work/unemployed, or student/homemaker), insurance (yes/no), having a usual source of medical care (yes/no), smoking status (current, former, never), leisure time physical activity in the last 30 days (yes/no), home owner (yes/no), years living in current residence (<5 , 5–9.9, 10–19.9, 20–29.9, ≥ 30.0), and census tract-level percent living in poverty ($<$ median of 9.14% vs. $\geq$ median).

Statistical Analyses

Descriptive statistics and reliability analyses are provided for the total sample and by urban/rural and low/high poverty strata. For consistency, all perceived neighborhood conditions were coded such that higher scores indicate greater neighborhood disadvantage. All scales are scored as the mean of the sum of the items with the exception of the Fear scale which due to its skewed distribution, uses the sum of the items.

For multi-item scales, internal consistency reliability was assessed using Cronbach's α coefficient and test-retest reliability was assessed with intraclass correlation coefficients (ICC) and associated 95% confidence intervals from a 1-way analysis of variance.[36] The ICC is defined as the between-subject variation divided by sum of between-and within-subject (test-retest) variation[37] and represents the proportion of the total variation in a measure due to true differences between subjects rather than differences due to random error and variability between measurement times. The ICC ranges from 0–1, where 0 indicates that subject variability is negligible compared to total measurement error (retest variability + random error) and 1 indicates that total measurement error is negligible compared to the variability between subjects. To test differences between ICCs from the rural vs. urban and the low vs. high poverty strata, we used 1000 bootstrap samples to obtain the distribution of difference in two ICCs under the null hypothesis that the two ICCs are equal. Then we obtained the achieved significance level (ASL) as the p-value by comparing the observed difference in two ICCs with the null distribution.[38]

For the ordinal single-item variables, test-retest reliability was assessed using percent agreement and weighted kappa and 95% confidence intervals. The kappa statistic adjusts for the amount of agreement expected to occur by chance alone.[39] We report the linearly weighted kappa because it corrects for the magnitude of the difference between adjacent and nonadjacent ordinal categories and it is less sensitive to the range of responses than its alternative, the quadratic-weighted kappa.[40] We test the difference of the linearly weighted kappas between the urban/rural and low/high poverty strata with a chi-square heterogeneity test; p values <.05 indicate significant heterogeneity between the strata.[41, 42] To correct for the tendency of kappa to be highly dependent on the prevalence of the condition in the population, we also report the prevalence-and bias-adjusted kappa (PABAK).[43] PABAK assumes fifty percent prevalence of the condition and absence of any bias, thereby reflecting an ideal situation and ignoring the prevalence and bias present in the “real world”.

For both ICC and kappa, we judge the adequacy of test-retest reliability following the guidelines as suggested by Landis and Koch: poor: 0–0.20; fair: 0.21–0.40; moderate: 0.41–0.60; substantial: 0.61–0.80; and nearly perfect agreement: 0.81 to 1.0.[44] We adopt the convention of claiming satisfactory internal consistency when coefficient $\alpha \geq .70$.[45]

To estimate the effects of measurement error on observed associations, we used regression calibration (RC) to calculate the calibrated mean neighborhood conditions. We calculate both the naïve (uncorrected) association of mean (Time 1 and 2) neighborhood conditions and the calibrated mean (Time 1 and 2) neighborhood conditions with self-rated health. RC predicts and uses the “true” neighborhood characteristics for each subject to correct effect estimates. Neighborhood conditions are assumed to be measured with random additive error, estimated from test-retest replicated measures, effectively adjusting for test-retest reliability. Using a linear calibration function for replicated data, the calibrated mean for each participant can be calculated as: $X_i^* = \bar{X}_{tot} + \lambda(\bar{X}_j - \bar{X}_{tot})$, where $\bar{X}_{tot}$ is the grand mean of all observations, $\bar{X}_j^*$ is the mean of the replicate measurements for each participant, and λ is the reliability coefficient ICC.[46] We compare the mean calibrated (regression calibration)

odds ratio that corrects for measurement error to the naïve analysis(uncorrected for measurement error). Due to our small sample size, we include only those potential confounders that were associated ($p<.05$) with both the outcome (self-rated health) and the stratifying variable (urban/rural), including education (high school or less vs. some college or more) and income ($<\$25K$, $\geq\$25K$, or missing)in measurement error analyses.

Analyses were conducted with STATA/IC 10.1[47](descriptive analyses, regression calibration), R[48](alphas and ICCs, bootstrapping) and WINPEPI version 2.19[49] (kappas). All p-values are two-sided.

RESULTS

Descriptive statistics

The first interview was completed by 1166 women without breast cancer between August, 2007 and June, 2009. The CASRO response rate,[50] reflecting respondent cooperation and telephone sampling efficiency, was 69.1%. Of all eligible and interested participants in the test-retest study ($n=475$), 17.9% were unable to be reached within the 14–21 day window. Of the 390 women who were reached, 367(94.1%) completed the second interview for the test-retest study within the 2–3 week window (mean=19.1 days; median=20 days) after the initial interview. The average interview administration time was 34.1 minutes ($SD=7.7$) for the first interview and 18.4 ($SD=4.1$) minutes for the second interview (reflecting the abbreviated Time 2 survey). Descriptive statistics of the 367 women in the retest group (mean age 60.6 years, range: 25–95) are shown in Table 1. Urban residents ($n=268$) were more likely to be non-white, have higher education and income, better self-rated health, to have lived in their homes for a shorter period of time, and to live in census tracts with lower rates of poverty. Because poverty status of urban and rural residents varied so dramatically, we also report reliability analyses by neighborhood poverty status (above/below median of 9.14% poverty).

All items are listed in Table 2. Due to highly skewed responses, several items were collapsed: items measuring *Neighborhood Fear* were dichotomized (0 vs. ≥ 1 days) prior to summing to create the *Neighborhood Fear* scale and one item in the *Multidimensional Neighborhood* scale(MN4) was collapsed into 3-categories(≤ 3 , 4–6, > 6). Spearman correlations between all measured scales and items are shown in Table 3. Most scales and items were inter-correlated, including, as expected, the subscales of Collective Efficacy (*Social Cohesion* and *Social Control*) and Neighborhood Disorder (*Physical and Social Disorder*).

Internal consistency reliability

All measured scales, with the exception of the measurement of *Social Control* among the rural residents at both time points and among the low poverty residents at Time 1, and *Neighborhood Fear* at Time 1 in the total and rural samples, demonstrated satisfactory internal consistency reliability ($\alpha \geq .70$) (Table 4).

Internal consistency reliability was not statistically different across either of the compared strata(urban/rural or low/high poverty) at Time 1 (Table 4) or at Time 2 (data not shown). However, while point estimates for internal consistency reliability were very similar across urban/rural strata for two of the scales (*Social Cohesion* and *Fear*), the remaining scales (*Social Control*, *Physical Disorder*, *Social Disorder*) demonstrated higher reliability among the urban residents. Point estimates for internal consistency reliability were quite similar in the high and low poverty samples for most scales, with slightly greater reliability for *Physical Disorder* and *Fear(Time 1)* and *Social Control(Time 1)* among the high poverty sample.

Test-retest reliability

For all scales, ICCs among the total sample indicated substantial (>0.70) test-retest reliability. For all items, weighted kappas and PABAKs indicate moderate test-retest reliability in the total sample (Table 4).

While point estimates of test-retest reliability were higher in the urban (vs. rural) sample for every scale and for 3 of the 4 items, significant differences were evident only for the following scales and items: *Social Control*, *Physical Disorder*, *MNI*, and *MN4*. For each of these, the ICCs demonstrated substantial reliability among urban residents and only moderate test-retest reliability among rural residents. Among the high poverty sample, test-retest reliability was significantly higher for *Physical Disorder* and was significantly lower for *MN4*.

Correction for Measurement Error

Overall, 16.9% of women reported being in fair or poor health. Both the naïve (uncorrected) and regression calibration results showed that women were more likely to report being in fair or poor health when living in neighborhoods with worse perceived conditions although the confidence intervals for several of the odds ratios included the null (Table 5). In general, compared to regression calibration-corrected estimates, naïve (uncorrected) estimates of the association of mean neighborhood characteristics with self-rated health were biased toward the null (Table 5). For nearly all items and scales, the attenuation in naïve analyses, resulting from measurement error, was more substantial among rural residents than among the urban residents. The naïve estimates were even more attenuated when only Time 1 measures were used (data not shown), reflecting greater within-person error when using a single vs. replicate (mean) measures. In sensitivity analyses, we also ran both unadjusted models and models adjusting for the demographic variables of race and age (data not shown), however the results were not substantially different from the data presented in Table 5.

DISCUSSION

Overall, all measured perceived neighborhood conditions demonstrated satisfactory internal consistency and at a minimum, moderate or substantial test-retest reliability in the full sample. These results confirm the test-retest and internal consistency reliability of questions designed to capture perceived neighborhood conditions using a mixed urban and rural sample.

However, the reliability and utility of these measures among rural residents is less clear. While internal consistency reliability did not differ statistically by urban/rural status, point estimates were generally slightly lower in rural populations. And for several scales and items, test-retest reliability was significantly lower in rural populations; for others, the differences in reliabilities between rural and urban populations was not statistically significant.

Associations of perceived neighborhood conditions with self-rated health were generally attenuated in naïve analyses and after correction for measurement error, the odds ratios demonstrated stronger associations. Notably, naïve associations appeared to be slightly more attenuated among rural residents, reflecting greater measurement error in this group. This measurement error is caused in part by the lower reliability among rural residents that we demonstrated in several of the measured neighborhood conditions.

While we cannot draw definitive conclusions about the reasons for lower reliability, greater measurement error, and attenuated associations across different neighborhood strata, it is possible that some aspects of neighborhood conditions do not apply equally to the residents

in urban/rural or low/high poverty neighborhoods. For example, in the event that some referenced scenarios (e.g. graffiti) are uncommon occurrences or not as noticeable in rural neighborhoods, residents may guess more frequently when answering these items, making the scale less reliable and introducing greater measurement error at both time points among rural residents.

With regard to urban/rural differences, one might expect to find lower reliability for *Physical Disorder*, *Social Control*, and *MN4* among rural residents. The *Physical Disorder* scale includes items referencing graffiti, noise, vandalism, and abandoned buildings, all of which are likely to be less relevant for rural residents. Similarly, *Social Control* refers to the likelihood that neighbors “do something” about deviant behaviors unlikely in rural settings, including kids hanging out on a street corner or spray-painting graffiti on a local building. For *MN4*, (*How many of your closest neighbors do you typically stop and chat with when you run into them?*), the number of close neighbors likely differs by urban/rural status, as does the frequency of running into them. Accordingly, we observed significantly lower test-retest reliabilities for each of these variables, but we found no differences in internal consistency reliability. In contrast, equivalent reliability might be expected for scales and items such as *Social Disorder*, *Social Cohesion*, *Fear*, and *MN1-3* given their measurement of more universal phenomena (e.g. feelings and relationships with neighbors, etc.). Indeed, we did not find any significant differences in either type of reliability for these items with one exception – *MN1* (*I feel strongly attached to this residence*) had significantly lower test-retest reliability in the rural sample.

For the high/low poverty comparisons, one might expect higher reliability for *Social Disorder*, *Physical Disorder*, *Fear*, and *Social Control* in high-poverty samples, given that each of these scales refers to disorder and/or behaviors more common in high poverty areas. In contrast, one would not expect reliability differences for *Social Cohesion* or *MN1-4*. However, in general, while we found a trend toward greater reliability in the high poverty sample on these as well as other scales, only one comparison was significantly different in the expected direction: test-retest reliability was significantly higher for *Physical Disorder* in the high poverty sample. Surprisingly, the test-retest reliability for *MN4* (chatting with neighbors) was significantly *lower* among the high poverty sample.

While some comparisons demonstrated significantly different test-retest reliability, internal consistency reliability was not significantly different across any strata comparisons. This could reflect true equivalence of internal consistency across strata, with items in each scale “hanging together” in the same way in both groups. However, point estimates indicated some what lower internal consistency reliability in the rural sample, and the lack of significance may simply be a side effect of low power to detect true differences.

At this time, is not clear what type of perceived neighborhood conditions may be most relevant or reliable for rural populations. While our results suggest that the measures of *Social Disorder* and *Social Cohesion* may be equally reliable and therefore more universal phenomena, future research should explore alternative aspects of the physical environment and social control that may be more relevant among rural residents. Other perceived neighborhood conditions, including geographic and/or social isolation, access to social and/or health care services, the role of shared values and/or culture, and socio demographic homogeneity, should also be explored.

Importantly, we assess only the individual-level measurement properties of perceived neighborhood conditions. To assess the group-level measurement properties of these conditions, econometric analyses are needed. Econometric analyses integrate traditional psychometric analysis into multilevel frameworks, allowing for reliability assessment of a

measure at both the individual-and neighborhood-level. By calculating the random variation on a measure between residents living in a neighborhood, this approach can provide evidence for or against the use of aggregated individual-level responses as reliable indicators of mean neighborhood-level constructs.[9, 51] Our data were too sparse for the grouping and multilevel analysis by census tract or even by county that this approach requires: the participants in this study are distributed across 298 census tracts (mean: 1.2 individuals per tract, range: 1–5) and 88 counties (mean: 4.2 individuals per county, range: 1–107).

This study has several limitations. While matched to a population-based sample, our sample represents an older, relatively affluent, and predominantly white group of women, limiting the generalizability of our results. The primary study was not designed to test urban-rural differences which could have limited our power to detect differences. In particular, our small sample sizes, particularly in the rural sample, may have contributed to the wide confidence intervals seen in the measurement error analyses. The larger analytic standard errors calculated using regression calibration also contributed to the wide intervals. Additionally, our reliability analyses do not account for the differential characteristics of respondents in urban and rural areas; future research should strive to account for any such differences. Finally, the importance of reliability in perceived fear in the previous week (*Neighborhood Fear* scale), is less clear than for our other measures given the greater potential for change in this scale over a short timeframe.

Conclusion

Current understanding about the psychometrics of perceived neighborhood conditions is rather limited.[16–22] Little research has examined the measurement properties of perceived global neighborhood conditions across urban and rural neighborhoods and no studies, to our knowledge, have assessed the role of measurement error in these conditions. We addressed these gaps in the literature by examining multiple global measures of perceived neighborhood conditions and conclude that while all measures demonstrated satisfactory internal consistency and test-retest reliability in the full sample, the lower reliability and greater measurement error among rural residents disproportionately attenuated measures of effect in this population and may have other, unknown consequences.

A deeper understanding of both the psychometric and econometric properties of perceived neighborhood conditions is needed to advance the growing body of research linking neighborhoods and health. More research will be needed to understand the implications of conceptualizing and analyzing perceived neighborhood conditions as either individual-level or group-level constructs, or whether they should be incorporated at both levels using a multilevel framework as Kawachi et al. have advocated for the measurement of *Social Capital*. [52] Given recent research demonstrating differing patterns of associations between neighborhood context and mental health status across urban and rural neighborhoods [53], more attention to context in rural areas is warranted. Particular work is needed to develop and test valid and psychometrically and econometrically reliable measures relevant across diverse populations and neighborhoods as well as measures specifically designed to capture neighborhood conditions relevant to the health of rural populations.

Acknowledgments

We thank the Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine in St. Louis, Missouri, for the use of the Health Behavior, Communication and Outreach Core. This research was supported in part by grants from the National Cancer Institute (CA112159, CA91842), and preparation of the manuscript was supported in part by the first author's Alvin J. Siteman Cancer Center Prevention and Control Program postdoctoral fellowship and her career-development award from the Institute of Clinical and Translational Sciences at Washington University in St. Louis, funded by the National Institutes of Health/National

Center for Research Resources (KL2 RR024994). Its contents are solely the responsibility of the authors and do not necessarily represent the views of the funding agencies.

References

1. Chaix B, Leyland AH, Sabel CE, et al. Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmo, Sweden, in 2001. *Journal of Epidemiology and Community Health*. 2006; 60:427–35. [PubMed: 16614334]
2. Muntaner C, Eaton WW, Miech R, et al. Socioeconomic position and major mental disorders. *Epidemiol Rev*. 2004; 26:53–62. [PubMed: 15234947]
3. O'Campo P, Gielen AC, Faden RR, et al. Violence by male partners against women during the childbearing year: a contextual analysis. *Am J Public Health*. 1995; 85:1092–7. [PubMed: 7625502]
4. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health*. 2001; 55:111–22. [PubMed: 11154250]
5. Robert SA. Socioeconomic position and health: The independent contribution of community socioeconomic context. *Ann Rev Sociol*. 1999; 25:489–516.
6. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*. 1997; 277:918–24. [PubMed: 9252316]
7. Ross CE, Reynolds JR, Geis KJ. The contingent meaning for neighborhood stability for residents' psychological well-being. *American sociological Review*. 2000; 65:581–97.
8. Robert SA. Neighborhood socioeconomic context and adult health. The mediating role of individual health behaviors and psychosocial factors. *Ann N Y Acad Sci*. 1999; 896:465–8. [PubMed: 10681952]
9. Raudenbush SW, Sampson RJ. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology* 1999. 1999; 2929:1–41.
10. Ross CE, Mirowsky J. Disorder and decay: The concept and measurement of perceived neighborhood disorder. *Urban Affairs Review*. 1999; 34:412–32.
11. Sampson RJ, Raudenbush SW. Seeing disorder: Neighborhood stigma and the social construction of "broken windows". *Social Psychology Quarterly*. 2004; 67:319–42.
12. Weden MM, Carpiano RM, Robert SA. Subjective and objective neighborhood characteristics and adult health. *Soc Sci Med*. 2008; 66:1256–70. [PubMed: 18248865]
13. Wen M, Hawkey LC, Cacioppo JT. Objective and perceived neighborhood environment, individual SES and psychosocial factors, and self-rated health: an analysis of older adults in Cook County, Illinois. *Soc Sci Med*. 2006; 63:2575–90. [PubMed: 16905230]
14. Echeverria S, Diez-Roux AV, Shea S, et al. Associations of neighborhood problems and neighborhood social cohesion with mental health and health behaviors: the Multi-Ethnic Study of Atherosclerosis. *Health Place*. 2008; 14:853–65. [PubMed: 18328772]
15. Gary TL, Safford MM, Gerzoff RB, et al. Perception of neighborhood problems, health behaviors, and diabetes outcomes among adults with diabetes in managed care: the Translating Research Into Action for Diabetes (TRIAD) study. *Diabetes Care*. 2008; 31:273–8. [PubMed: 18000180]
16. Brownson RC, Chang JJ, Eyler AA, et al. Measuring the environment for friendliness toward physical activity: a comparison of the reliability of 3 questionnaires. *Am J Public Health*. 2004; 94:473–83. [PubMed: 14998817]
17. Echeverria SE, Diez-Roux AV, Link BG. Reliability of self-reported neighborhood characteristics. *J Urban Health*. 2004; 81:682–701. [PubMed: 15466849]
18. Kirtland KA, Porter DE, Addy CL, et al. Environmental measures of physical activity supports: perception versus reality. *Am J Prev Med*. 2003; 24:323–31. [PubMed: 12726870]
19. Mujahid MS, Diez Roux AV, Morenoff JD, et al. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am J Epidemiol*. 2007; 165:858–67. [PubMed: 17329713]
20. Saelens BE, Sallis JF, Black JB, et al. Neighborhood-based differences in physical activity: an environment scale evaluation. *Am J Public Health*. 2003; 93:1552–8. [PubMed: 12948979]

21. Sallis JF, Johnson MF, Calfas KJ, et al. Assessing perceived physical environmental variables that may influence physical activity. *Res Q Exerc Sport*. 1997; 68:345–51. [PubMed: 9421846]
22. Evenson KR, McGinn AP. Test-retest reliability of a questionnaire to assess physical environmental factors pertaining to physical activity. *Int J Behav Nutr Phys Act*. 2005; 2:7. [PubMed: 15958168]
23. Riva M, Gauvin L, Barnett TA. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *J Epidemiol Community Health*. 2007; 61:853–61. [PubMed: 17873220]
24. Ross CA, Mirowsky J. Neighborhood disadvantage, disorder, & health. *Journal of Health and Social Behavior*. 2001; 42:258–76. [PubMed: 11668773]
25. Ross CE. Neighborhood disadvantage and adult depression. *Journal of Health and Social Behavior*. 2000; 41:177–87.
26. Ross CE, Mirowsky J, Pribesh S. Powerlessness and the amplification of threat: Neighborhood disadvantage, disorder, and mistrust. *American sociological Review*. 2001; 66:568–91.
27. Skjaeveland O, Garling T, Maeland JG. A multidimensional measure of neighboring. *American Journal of Community Psychology*. 1996; 24:413–35.
28. Haney WGESK. Perception of neighborhoods by city and suburban residents. *Human Ecology*. 1978; 6:201–14.
29. Walton D, Murray SJ, Thomas JA. Relationships Between Population Density and the Perceived Quality of Neighbourhood. *Social Indicators Research*. 2008; 89:405–20.
30. Krieger N, Chen JT, Waterman PD, et al. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. *Am J Epidemiol*. 2002; 156:471–82. [PubMed: 12196317]
31. Singh, GK.; Miller, BA.; Hankey, BF., et al. Area socioeconomic variations in US cancer incidence, mortality, stage, treatment, and survival, 1975–1999. Bethesda, MD: National Cancer Institute; 2003. Report No.: NIH Publication No. 03–0000
32. USDA Economic Research Service. Measuring rurality: Rural-urban commuting area codes 2005 [8–17–2009]. Available from: <http://www.ers.usda.gov/briefing/rurality/ruralurbancommutingareas/>
33. Finch BK, Hummer RA, Reindl M, et al. Validity of self-rated health among Latino(a)s. *Am J Epidemiol*. 2002; 155:755–9. [PubMed: 11943694]
34. Power C, Matthews S, Manor O. Inequalities in self rated health in the 1958 birth cohort: lifetime social circumstances or social mobility? *BMJ*. 1996; 313:449–53. [PubMed: 8776310]
35. Shetterly SM, Baxter J, Mason LD, et al. Self-rated health among Hispanic vs non-Hispanic white adults: the San Luis Valley Health and Aging Study. *Am J Public Health*. 1996; 86:1798–801. [PubMed: 9003141]
36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979; 86:420–8. [PubMed: 18839484]
37. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996; 1:30–46.
38. Efron, B.; Tibshirani, RJ. An introduction to the bootstrap. Chapter 16-hypothesis testing with the bootstrap. Boca Raton, FL: Chapman & Hall/CRC; 1993.
39. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
40. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology*. 1996; 7:199–202. [PubMed: 8834562]
41. Abramson, JH. WINPEPI Programs: PAIRSetc Manual Version 2.19. 2009.
42. Fleiss, JL.; Levin, B.; Paik, MC. Statistical methods for rates and proportions. 3. Hoboken, NJ: John Wiley & Sons; 2003.
43. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993; 46:423–9. [PubMed: 8501467]

44. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
45. Nunnally, JC. *Psychometric theory*. New York, NY: McGraw-Hill, Inc; 1994.
46. Carroll, RJ.; Ruppert, D.; Stefanski, LA. *Monographs on Statistics and Applied Probability*. Vol. 63. Chapman & Hall; 1995. Measurement error in non-linear models.
47. Stata Statistical Software/IC: Release 10.1. 2009.
48. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria: <http://www.R-project.org>
49. Abramson JH. WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. *Epidemiol Perspect Innov*. 2004; 1:6. [PubMed: 15606913]
50. COASR (Council of American Survey Research Organizations). On the definition of response rates: A special report from the CASTRO Task Force on Completion Rates. Port Jefferson, NY: 1982.
51. Raudenbush, SW. *Neighborhoods and Health*. Vol. 5. Oxford, England: Oxford University Press; 2003. The quantitative assessment of neighborhood social environments; p. 112-131.
52. Kawachi I, Kim D, Coutts A, et al. Commentary: Reconciling the three accounts of social capital. *Int J Epidemiol*. 2004; 33:682–90. discussion 700–4. [PubMed: 15282222]
53. Peterson LE, Tsai AC, Petterson S, et al. Rural-urban comparison of contextual associations with self-reported mental health status. *Health Place*. 2009; 15:125–32. [PubMed: 18434234]

Table 1

Descriptive Characteristics of Women in the Retest Sample (n=367).

	Total n=367		Urban n=238		Rural n=129		Chi-square test for urban/rural difference
	N	%	N	%	N	%	
Race/ethnicity							12.0 (2) p=.002
Non-Hispanic White	328	89.6	206	86.6	122	95.3	
Non-Hispanic Black	21	5.7	21	8.8	0	0	
Hispanic, other, multiple	17	4.6	11	4.6	6	4.7	
Age							7.1 (4) p=.131
25-44	44	12.0	33	13.9	11	8.5	
45-54	65	17.7	47	19.7	18	14.0	
55-64	119	32.4	77	32.4	42	32.6	
65-74	80	21.8	44	18.5	36	27.9	
≥ 75	59	16.0	37	15.5	22	17.1	
Education							16.3 (2) p<.001
≤ High school	142	38.7	75	31.5	67	51.9	
1-3 years college	107	29.2	73	30.7	34	26.4	
≥ College graduate	118	32.2	90	21.7	28	21.7	
Household income							19.1 (4) p=.001
≥ \$75,000	80	21.8	63	26.5	17	13.2	
\$50-74,999	69	18.8	46	19.3	23	17.8	
\$35-49,999	64	17.4	46	19.3	18	14.0	
\$25-34,999	36	9.8	20	8.4	16	12.4	
≤ 24,999	91	24.8	45	18.9	46	35.6	
Missing	27	7.4	18	7.6	9	7.0	
Marital status							.516 (1) p=.473
Married or living together	256	69.8	163	68.5	93	72.1	
Unmarried	111	30.3	75	31.5	36	27.9	
Employment							1.07 (3) p=.783
Employed	140	38.2	95	39.9	45	34.9	
Retired	111	30.3	71	29.8	40	31.0	

	Total n=367		Urban n=238		Rural n=129		Chi-square test for urban/rural difference
	N	%	N	%	N	%	
Unable to work or unemployed	56	15.3	34	14.3	22	17.1	.314 (1) p=.646
Student/homemaker	60	16.4	38	16.0	22	17.1	
Insurance							
No	22	6.0	13	5.5	9	7.0	.009 (1) p=1.0
Yes	345	94.0	225	94.5	120	93.0	
Has usual source of care							1.73(2) p=.422
No	6	1.6	4	1.7	2	1.6	
Yes	361	98.4	234	98.3	127	98.4	3.61(1) p=.057
Smoking status							
Never smoked	233	63.8	155	65.4	78	60.9	
Former smoker	81	22.2	53	22.4	28	21.9	1.96 (1) p=.162
Current smoker	51	14.0	29	12.2	22	17.2	
Physical activity							10.67(4) p=.030
No	111	30.2	64	26.9	47	36.4	
Yes	256	69.8	174	73.1	82	63.6	10.69 (1) p=.001
Home owner							
No	43	11.7	32	13.5	11	8.5	70.2(1) p<.001
Yes	324	88.3	206	86.6	118	91.5	
Years at residence							10.69 (1) p=.001
<5	78	21.3	59	24.8	19	14.7	
5-9.99	70	19.1	42	17.7	28	21.7	
10-19.99	78	21.3	52	21.9	26	20.2	
20-29.99	63	17.2	44	18.5	19	14.7	
≥ 30	78	21.3	41	17.2	37	28.7	70.2(1) p<.001
Self-rated health							
Excellent/very good/good	305	83.1	209	87.8	96	74.4	
Fair/poor	62	16.9	29	12.2	33	25.6	70.2(1) p<.001
Census tract percent living in poverty							
<9.14% (median)	183	49.9	157	66.0	26	20.2	
≥ 9.14%	184	50.1	81	34.0	103	79.8	

Table 2

Scales and items measuring self-reported neighborhood conditions.

SCALES
Social Control (n=5 items, scale/item range: 1–5)
CN1. If a group of neighborhood children were skipping school and hanging out on a street corner, how likely is it that your neighborhoods would do something about it?
CN2. If some children were spray-painting graffiti on a local building, how likely is it that your neighbors would do something about it?
CN3. If a child was showing disrespect to an adult, how likely is it that people in your neighborhood would scold that child?
CN4. If there was a fight in front of your house and someone was being beaten or threatened, how likely is it that would neighbors would break it up?
CN5. Suppose that because of budget cuts the fire station closest to your home was going to be closed down by the city. How likely is it that neighborhood residents would organize to try to do something to keep the fire station open?
Social Cohesion (n=5 items, scale/item range: 1–4)
CH1. People around here are willing to help their neighbors.
CH2. This is close-knit neighborhood.
CH3. People in this neighborhood can be trusted.
CH4. People in this neighborhood generally don't get along with each other.
CH5. People in this neighborhood do not share the same values.
Physical Disorder (n=6 items, scale/item range: 1–4)
PD1. There is a lot of graffiti in my neighborhood.
PD2. My neighborhood is noisy.
PD3. Vandalism is common in my neighborhood.
PD4. There are lots of abandoned buildings in my neighborhood.
PD5. My neighborhood is clean.
PD6. People in my neighborhood take good care of their houses and apartments.
Social Disorder (n=9 items, scale/item range: 1–4)
SD1. There are too many people hanging around on the streets near my home.
SD2. There is too much drug use in my neighborhood.
SD3. There is too much alcohol use in my neighborhood.
SD4. I'm always having trouble with my neighbors.
SD5. There is a lot of crime in my neighborhood.
SD6. In my neighborhood, people watch out for each other.
SD7. The police protection in my neighborhood is adequate.
SD8. My neighborhood is safe.
SD9. I can trust most people in my neighborhood.
Fear (n=3 items, item range: 0 or 1, scale range: 0–3)
NF1. During the past week, how many days did you fear being robbed, attacked, or physically injured? (none[0] vs. any [1])
NF2. During the past week, how many days were you worried that your home would be broken into? (none[0] vs. any [1])
NF3. During the past week, how many days did you feel afraid to leave the house? (none[0] vs. any [1])
ITEMS
Multidimensional Neighboring (n=4 items, item range: 1–5 [MN1,2] or 1–3 [MN3])
MN1. I feel strongly attached to this residence.
MN2. I feel at home in this neighborhood.

SCALES

MN3. I would have better contacts with friends, family, etc. if I lived in another neighborhood.

Table 3

Spearman correlation coefficients for scales and items measuring self-reported neighborhood conditions at Time 1.

Scales and items	Social Control	Social Cohesion	Physical Disorder	Social Disorder	Fear	MN1	MN2	MN3	MN4
Social Control	-	-	-	-	-	-	-	-	-
Social Cohesion	.57*	-	-	-	-	-	-	-	-
Physical Disorder	.30*	.36*	-	-	-	-	-	-	-
Social Disorder	.34*	.55*	.72*	-	-	-	-	-	-
Fear	.10	.21*	.28*	.31*	-	-	-	-	-
MN1	.28*	.39*	.21*	.30*	.05	-	-	-	-
MN2	.25*	.57*	.36*	.41*	.14*	.63*	-	-	-
MN3	.26*	.43*	.32*	.43*	.13*	.31*	.38*	-	-
MN4	.22*	.33*	.05	.16*	.12*	.14*	.22*	.12*	-

* p≤.05

Table 4

Descriptive, test-retest, and internal consistency reliability statistics for scales and items measuring self-reported neighborhood conditions by urban/rural and high/low neighborhood poverty strata (n=367). P values of <.05 indicate significant differences in reliability across strata. Higher mean scores for all scales reflect greater neighborhood disadvantage.

SCALES									
Descriptives by Strata					Reliability by Strata			Test-Retest	
					Internal Consistency				
Strata	Scale	Range	Time 1	Time 2	Mean (SD)	Time 1	Time 2	Cronbach's α	ICC (95% CI)
Social Control									
Total	1-5.0	1-5.0	1.79 (.70)	1.67 (.68)		.71	.75		.71 (.65-.76)
Urban	1-5.0	1-4.6	1.82 (.74)	1.71 (.72)		.73	.77		.79 (.73-.83)
Rural	1-3.4	1-5.0	1.72 (.60)	1.58 (.58)		.64	.69		.45 (.29-.59)
						Urban vs. rural: p=.467			Urban vs. rural: p=.005
Low Poverty	1-4.4	1-4.6	1.72 (.63)	1.66 (.65)		.65	.75		.76 (.69-.81)
High Poverty	1-5.0	1-5.0	1.85 (.76)	1.68 (.71)		.75	.76		.66 (.57-.75)
						Low vs. high poverty: p=.526			Low vs. high poverty: p=.202
Social Cohesion									
Total	1-3.8	1-3.6	1.62 (.53)	1.55 (.50)		.85	.86		.76 (.72-.80)
Urban	1-3.8	1-3.6	1.65 (.55)	1.58 (.52)		.85	.87		.77 (.71-.82)
Rural	1-3.0	1-2.6	1.56 (.49)	1.51 (.45)		.86	.85		.75 (.66-.82)
						Urban vs. rural: p=.464			Urban vs. rural: p=.430
Low Poverty	1-3.6	1-3.6	1.58 (.51)	1.52 (.51)		.87	.88		.77 (.70-.82)
High Poverty	1-3.8	1-3.6	1.66 (.55)	1.58 (.49)		.84	.84		.76 (.68-.82)
						Low vs. high poverty: p=.484			Low vs. high poverty: p=.425
Physical Disorder									
Total	1-3.8	1-3.5	1.38 (.44)	1.31 (.42)		.80	.83		.73 (.67-.77)
Urban	1-3.8	1-3.5	1.38 (.47)	1.30 (.45)		.82	.86		.78 (.72-.82)
Rural	1-2.5	1-2.5	1.39 (.39)	1.33 (.36)		.74	.76		.59 (.46-.69)
						Urban vs. rural: p=.499			Urban vs. rural: p=.012
Low Poverty	1-2.5	1-2.5	1.27 (.33)	1.23 (.33)		.71	.77		.62 (.52-.70)
High Poverty	1-3.8	1-3.5	1.50 (.50)	1.40 (.48)		.82	.85		.75 (.68-.81)
						Low vs. high poverty: p=.451			Low vs. high poverty: p=.046

SCALES											
Descriptives by Strata					Reliability by Strata						
					Internal Consistency			Test-Retest			
Strata	Scale	Range	Time 1	Time 2	Mean (SD)	Time 1	Time 2	Cronbach's α	Time 1	Time 2	ICC (95% CI)
Social Disorder											
Total	1-3.2	1-3.9			1.48 (.46)		1.40 (.44)	.88	.89		.75 (.70-.79)
Urban	1-3.2	1-3.9			1.47 (.49)		1.39 (.47)	.90	.91		.78 (.72-.82)
Rural	1-2.8	1-2.4			1.49 (.41)		1.43 (.38)	.85	.82		.68 (.58-.77)
								Urban vs. rural: p=.511			Urban vs. rural: p=.108
Low Poverty	1-2.9	1-3.9			1.37 (.42)		1.34 (.42)	.90	.90		.74 (.67-.80)
High Poverty	1-3.2	1-3.2			1.58 (.47)		1.47 (.46)	.86	.87		.74 (.67-.80)
								Low vs. high poverty: p=.497			Low vs. high poverty: p=.477
Fear											
Total	0-3	0-3			0.10 (.41)		0.13 (.49)	.64	.74		.70 (.64-.74)
Urban	0-3	0-3			0.10 (.43)		0.14 (.50)	.71	.73		.72 (.66-.78)
Rural	0-2	0-3			0.09 (.36)		0.12 (.48)	.48	.77		.63 (.52-.73)
								Urban vs. rural: p=.512			Urban vs. rural: p=.331
Low Poverty	0-3	0-3			0.09 (.36)		0.15 (.52)	.47	.73		.62 (.52-.70)
High Poverty	0-3	0-3			0.10(.45)		0.11 (.47)	.76	.76		.77 (.70-.82)
								Low vs. high poverty: p=.539			Low vs. high poverty: p=.182

ITEMS									
Descriptives by Urban/Rural Strata					Test-Retest Reliability by Strata				
					Internal Consistency			Test-Retest	
Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Cronbach's α	ICC (95% CI)
MNL. I feel strongly attached to this residence.									
Strongly agree	45.9	49.6	40.8	48.3	55.4	51.9	.58	.59 (.52-.65)	
Agree	38.9	33.5	43.3	34.0	30.8	32.6	.62	.64 (.57-.72)	
Neither	10.3	12.8	10.9	13.5	9.2	11.6	.52	.48 (.35-.61)	
					Urban vs. rural: p=.029			Urban vs. rural: p=.029	
Disagree	2.5	2.2	1.7	1.7	3.9	3.1	.62	.60 (.50-.70)	
Strongly disagree	2.5	1.9	3.4	2.5	0.8	0.8	.54	.58 (.49-.68)	
					Low vs. high poverty: p=.784			Low vs. high poverty: p=.784	

ITEMS	Descriptives by Urban/Rural Strata						Test-Retest Reliability by Strata			
	%Total		% Urban		% Rural		Strata	% Agreement	PABAK	Weighted kappa (95% CI)
	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2				
MN2. I feel at home in this neighborhood										
Strongly agree	59.2	62.1	55.9	34.3	65.4	63.6	Total	73.6	.67	.58 (.50–.67)
Agree	34.8	33.2	37.0	33.2	30.8	33.3	Urban	73.9	.67	.62 (.52–.71)
Neither	3.3	1.9	3.8	2.1	2.3	1.6	Rural	73.6	.67	.51 (.36–.66)
Disagree	0.8	1.1	1.3	1.3	0	0.8	Low Poverty	72.1	.65	Urban vs. rural: p=.258
Strongly disagree	1.9	1.6	2.1	2.1	1.5	0.8	High Poverty	75.5	.69	.54 (.41–.67)
										.62 (.51–.73)
										Low vs. high poverty: p=.351
MN3. I would have better contacts with friends, family, etc. if I lived in another neighborhood										
Strongly agree	3.5	6.0	3.4	6.7	3.9	4.7	Total	53.0	.41	.41 (.34–.49)
Agree	10.9	13.4	10.1	10.9	12.3	17.8	Urban	52.2	.40	.38 (.28–.47)
Neither	11.1	5.7	10.5	5.9	12.3	5.4	Rural	52.7	.41	.40 (.27–.53)
Disagree	32.9	25.3	33.6	26.1	31.5	25.3	Low Poverty	49.2	.36	Urban vs. rural: p=.749
Strongly disagree	41.6	49.6	42.4	50.4	40.0	49.6	High Poverty	57.1	.46	.34 (.24–.45)
										.47 (.37–.58)
										Low vs. high poverty: p=.091
MN4. How many of your closest neighbors do you typically stop and chat with when you run into them?										
>6	31.3	32.4	29.2	37.5	37.5	37.6	Total	66.9	.50	.59 (.53–.66)
4–6	37.8	37.0	39.2	39.5	36.6	44.8	Urban	70.7	.56	.65 (.58–.73)
≤ 3	31.0	30.7	31.7	23.0	25.9	17.6	Rural	60.0	.40	.47 (.35–.59)
							Low Poverty	71.8	.58	Urban vs. rural: p=.013
							High Poverty	62.9	.44	.67 (.58–.75)
										.53 (.43–.63)
										Low vs. high poverty: p=.033

Table 5

Odds ratio (OR) and 95% confidence interval (CI) estimates from logistic regressions of the association of continuous measures of worse self-reported neighborhood characteristics and fair/poor self-rated health, adjusting for education and income. Models include a naïve analysis (traditional logistic regression using mean neighborhood conditions measured at Time 1 and 2) and a measurement error-corrected analysis (regression calibration using calibrated mean neighborhood characteristics at Time 1 and 2) (n=367).

	Naïve Analysis		Regression Calibration	
	Mean Neighborhood Conditions		Calibrated Mean Neighborhood Conditions	
	OR	95% CI	OR	95% CI
Social Control				
Total sample	0.99	0.62–1.60	1.05	0.60–1.83
Urban	1.04	0.57–1.89	1.05	0.53–2.07
Rural	1.36	0.54–3.42	1.81	0.45–7.21
Social Cohesion				
Total sample	2.32	1.28–4.22	2.60	1.31–5.15
Urban	2.25	1.07–4.77	2.41	1.02–5.67
Rural	3.67	1.24–10.81	4.98	1.43–17.37
Physical Disorder				
Total sample	1.47	0.75–2.88	1.59	0.69–3.63
Urban	1.61	0.71–3.64	1.73	0.67–4.44
Rural	1.53	0.43–5.46	1.82	0.28–11.91
Social Disorder				
Total sample	2.21	1.12–4.39	2.67	1.23–5.79
Urban	1.87	0.81–4.36	2.61	1.05–6.45
Rural	4.87	0.99–23.96	5.93	1.57–22.09
Neighborhood Fear				
Total sample	1.17	0.66–2.09	1.22	0.59–2.52
Urban	1.07	0.51–2.23	1.08	0.44–2.64
Rural	1.78	0.63–5.03	2.11	0.51–8.78
MN1. I feel strongly attached to this residence.				
Total sample	1.19	0.86–1.65	1.23	0.83–1.83
Urban	1.08	0.69–1.67	1.09	0.66–1.80
Rural	1.49	0.85–2.61	1.69	0.80–3.58
MN2. I feel at home in this neighborhood.				
Total sample	1.37	0.95–1.98	1.46	0.93–2.28
Urban	1.31	0.82–2.08	1.36	0.80–2.33
Rural	1.71	0.89–3.28	2.01	0.69–5.85
MN3. I would have better contacts with friends, family, etc. if I lived in another neighborhood.				
Total sample	1.59	1.20–2.10	2.48	1.07–5.78
Urban	1.56	1.08–2.26	1.94	1.09–3.45
Rural	1.78	1.14–2.79	2.48	1.07–5.78

	Naïve Analysis		Regression Calibration	
	Mean Neighborhood Conditions		Calibrated Mean Neighborhood Conditions	
	OR	95% CI	OR	95% CI
MN4. How many of your closest neighbors do you typically stop and chat with when you run into them?				
Total sample	1.16	0.75–1.78	1.29	0.77–2.18
Urban	1.06	0.59–1.92	1.17	0.60–2.30
Rural	1.37	0.70–2.69	1.75	0.69–4.47