Street Audits to Measure Neighborhood Disorder: Virtual or In-Person?

Stephen J. Mooney*, Michael D. M. Bader, Gina S. Lovasi, Julien O. Teitler, Karestan C. Koenen, Allison E. Aiello, Sandro Galea, Emily Goldmann, Daniel M. Sheehan, and Andrew G. Rundle

* Correspondence to Dr. Stephen J. Mooney, Harborview Injury Prevention & Research Center, University of Washington, 401 Broadway, 4th Floor, Seattle, WA 98122 (e-mail: sjm2186@u.washington.edu).

Neighborhood conditions may influence a broad range of health indicators, including obesity, injury, and psychopathology. In particular, neighborhood physical disorder—a measure of urban deterioration—is thought to encourage crime and high-risk behaviors, leading to poor mental and physical health. In studies to assess neighborhood physical disorder, investigators typically rely on time-consuming and expensive in-person systematic neighborhood audits. We compared 2 audit-based measures of neighborhood physical disorder in the city of Detroit, Michigan: One used Google Street View imagery from 2009 and the other used an in-person survey conducted in 2008. Each measure used spatial interpolation to estimate disorder at unobserved locations. In total, the virtual audit required approximately 3% of the time required by the in-person audit. However, the final physical disorder measures were significantly positively correlated at census block centroids ($r = 0.52$), identified the same regions as highly disordered, and displayed comparable leave-one-out cross-validation accuracy. The measures resulted in very similar convergent validity characteristics (correlation coefficients within 0.03 of each other). The virtual audit–based physical disorder measure could substitute for the in-person one with little to no loss of precision. Virtual audits appear to be a viable and much less expensive alternative to in-person audits for assessing neighborhood conditions.

data collection; Detroit, Michigan; epidemiologic methods; Google Street View; social environment; spatial analysis; urban health

Abbreviations: CANVAS, Computer-Assisted Neighborhood Visual Assessment System; DNHS, Detroit Neighborhood Health Study.

Neighborhood conditions may influence a broad range of public health indicators (1), including obesity (2, 3), injury (4, 5), and psychopathology (6–8). In particular, neighborhood physical disorder—visual indications of neighborhood deterioration sometimes referred to as the "broken windows" phenomenon—may encourage crime (9, 10) and high-risk behaviors (11), resulting in poorer health outcomes (12); however, this theory is controversial (13, 14). Neighborhood physical disorder is modifiable (e.g., through community development or blight removal programs) and disproportionately affects disadvantaged communities, so interventions that remove physical disorder may have broad effects on community health (14). However, measuring physical disorder is challenging.

Neighborhood physical disorder is often measured using systematic neighborhood audits, wherein trained observers record social or physical characteristics of street segments according to explicit rules (15). However, the large-scale use of systematic audits has been limited by the substantial logistic and economic costs of deploying a research team to observe a representative subset of a city's streets and neighborhoods (16). Therefore, an alternate approach to characterizing neighborhood physical disorder that maintains validity of the construct while providing logistical and financial advantages is needed. To this end, rather than visiting neighborhoods in person, researchers have begun auditing streets using captured street imagery, such as that offered on Google Street View (Google, Inc., Mountain View, California) (17). These "virtual audits" typically cost less and incur a smaller logistical burden than in-person audits would (18). Many street audit items can be assessed from remote imagery with reliability

Figure 1. A screen capture of Google Street View imagery of Detroit, Michigan, 2015. Prior studies have confirmed that Google Street View imagery is typically detailed enough to reliably assess common indicators of physical disorder. For example, in this image, building maintenance is good, and no abandoned buildings or graffiti are present; however, there is a vacant lot. Image © Google, Inc., 2015.

that is comparable to that of in-person assessment (19–23) (Figure 1).

However, comparing reliability of items alone provides an incomplete picture of how in-person and virtual audits differ as methods for assessing context. Sample plans for in-person audits typically prefer geographically clustered streets to minimize time and logistic travel costs, whereas virtual audits incur no benefits from clustered designs. On the other hand, virtual audit scales typically exclude items that cannot be measured using remote imagery alone, such as noise and smells (24). Therefore, the selection of an audit strategy should not depend on item-level reliability alone. Rather, researchers will want to weigh the validity of the final measure against the costs and benefits of each approach. Yet, to the best of our knowledge, there has been no previous study in which researchers have compared the development and validation of a measure computed from a virtual audit to those from one computed from in-person neighborhood audit.

In this analysis, we directly compared 2 measures of neighborhood physical disorder across Detroit, Michigan, a city with substantial variability in disorder. One measure used in-person street audits conducted in 2008 as part of the Detroit Neighborhood Health Study (DNHS) (25). The second used virtual audits was conducted with the Computer Assisted Neighborhood Visual Assessment System (CANVAS) (24) using Google Street View imagery dating primarily from 2009. Here, we describe and explore the costs and benefits of each approach, using cross-validation to estimate error in each method.

METHODS

Street audit samples

Samples for each study were taken independently and not with the intent of comparing methods. The DNHS research team first sampled block groups from each of Detroit's 54 municipally defined neighborhoods in proportion to the total number of people living in the neighborhood according to the 2000 US Census (26). From each neighborhood, the team chose at least 1 but no more than 4 block groups. Second, the team divided all remaining block groups in the city of Detroit into quartiles based on population density and randomly sampled 10 block groups from each quartile. This procedure

resulted in the selection of a total of 138 block groups for evaluation. Three block groups made up entirely of a park, cemetery, or factory were excluded. The DNHS in-person audit team then assessed both sides of every public street found in the remaining block groups as detailed below, resulting in a total of 4,138 assessed street segments clustered in 135 block groups.

The CANVAS virtual audit sample has been described in detail previously (27). Briefly, investigators used ArcGIS (Esri, Redlands, California) to identify coordinates of points representing a 2-km grid across the Detroit city limits, with a 1-km grid oversample in neighborhoods in the highest quartile of population density (7,770 residents/km^2) for the Detroit metropolitan area as estimated by the 2006–2010 American Community Survey (28) and a 0.5-km grid oversample in neighborhoods in which subjects in the Fragile Families and Childhood Well Being Study (29) resided. Street View imagery was not available at every point selected for the CANVAS sample because not every sample point fell within a street. The CANVAS team designed an algorithm (described in the Web Appendix of Mooney et al. (27)) to randomly select a street segment within 50 meters if possible. From each selected segment, the algorithm selected one side at random to audit. Ultimately, 15 (2.9%) sample points fell in parks or large industrial areas in which no public streets were present, resulting in 502 unique block faces to be virtually audited. Figure 2 shows the audited segments from the DNHS and CANVAS samples. The CANVAS point sample included Highland Park and Hamtramck, 2 communities fully contained within Detroit's boundaries, whereas the DNHS sampling plans excluded them because they are administratively independent of the city of Detroit. More details regarding each sample are available in Web Appendix 1 (available at https://academic.oup.com/aje).

Audit items and physical disorder measures

The DNHS survey's 19-item audit instrument was originally developed from the Inner-City Mental Health Study Predicting HIV/AIDS and Other Drug Transitions (30), a study of urban conditions, drug use, and mental health in New York, New York. The DNHS team altered the instrument slightly to account for differences between Detroit's built environment and New York's. For example, an item assessing the presence of apartment buildings with doormen was irrelevant for Detroit, where most housing was designed for 1 or 2 families. DNHS audit training materials are presented in Web Appendix 2.

Two DNHS auditors walked each selected street segment together during June and July of 2008 and recorded the presence or absence of each item on each street segment. For each block group, the proportion of streets falling in the block group for which the audit item could be observed from the street segment (even if the disorder indicator was physically located in an adjacent block group across a boundary street) was taken as the audit measure for that block group. An exploratory factor analysis on the original 19-item audit instrument, in which block groups were treated as the unit of analysis, indicated that 3 audit items assessing building quality formed a cohesive factor: 1) presence of buildings with broken windows, boarded-up windows, or boarded-up doors; 2) presence of buildings with outside damage that can only be corrected by major repairs, such as damaged siding, shingles, boards, brick,
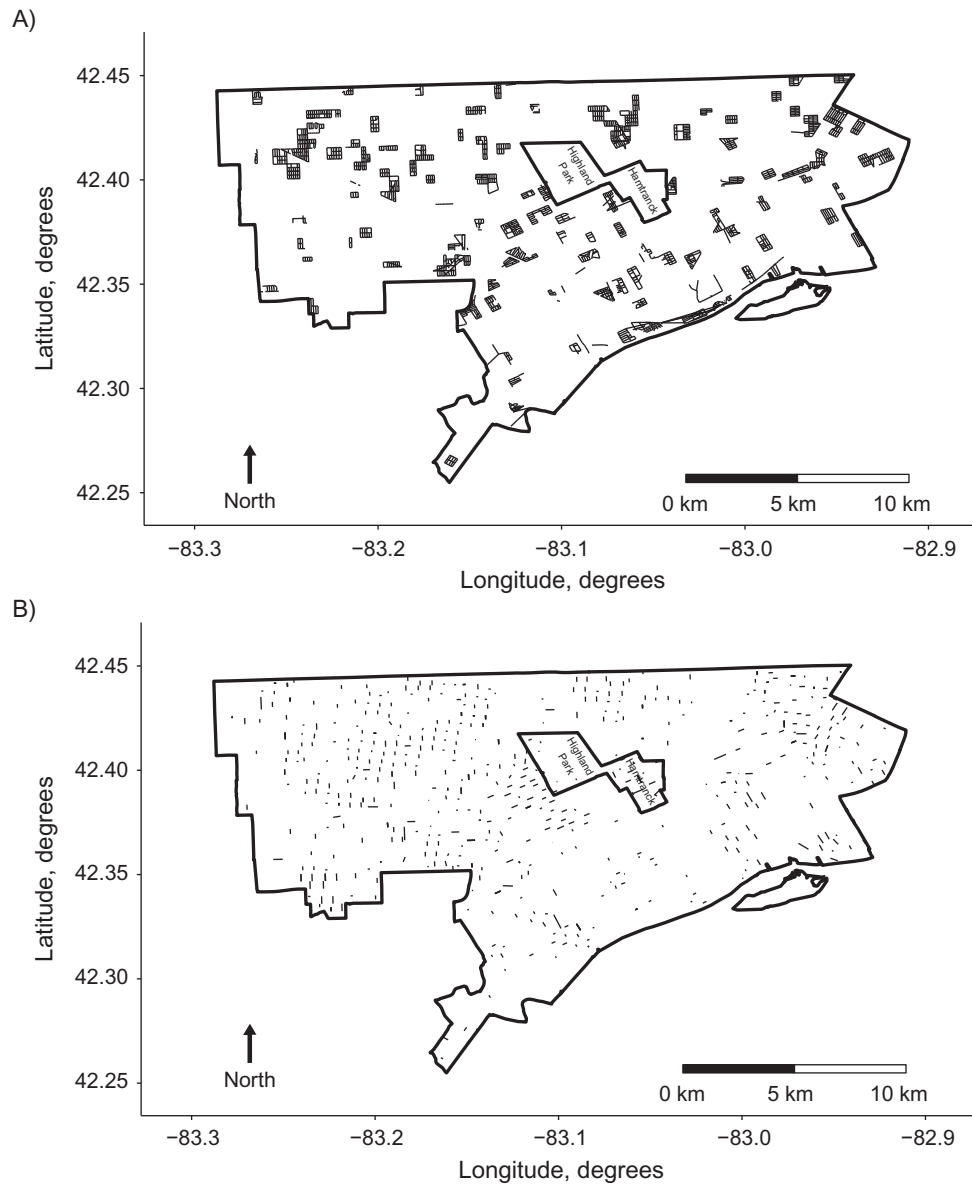
Figure 2. Blocks audited in the 2 studies described here. A) The blocks audited in person by the Detroit Neighborhood Health Study, 2008. B) The block faces audited using the Computer-Assisted Neighborhood Visual Assessment System, 2009.

concrete, and stucco; and 3) presence of entirely vacant buildings (25). The component score computed from the factor analysis was taken as the measure of physical disorder for each block group for research purposes (e.g., see Keyes et al. (25)).

The CANVAS audit measures, including inter-rater reliability statistics, have been described in more detail previously (27). Briefly, 9 virtual audit items to assess indicators of neighborhood physical disorder were developed from 3 validated in-person audit inventories (31–33). These 9 items included 1) litter, 2) empty alcohol bottles, 3) graffiti, 4) burned-out buildings, 5) abandoned buildings, 6) abandoned cars, 7) poor building maintenance, 8) vacant lots, and 9) bars on windows.

Seven auditors performed the CANVAS virtual audits over the period from August 2012 to May 2013. Image capture dates

for Google Street View imagery on the audited block faces ranged from August 2007 to October 2011, with 98% of the audited imagery dating from June to September 2009. From the 9 audit items, the CANVAS team fit a 2-parameter item response theory model (34). From this model, for each audited block face, the CANVAS team estimated a latent ability score using the observed response pattern that represented the physical disorder level for that block face. This technique is similar to what others have done previously to measure disorder (27, 31).

Spatial interpolation

Each team used ordinary kriging to interpolate the level of physical disorder at unobserved locations (35). Because

ordinary kriging relies on spatial covariance for estimation, each team tested that autocorrelation was present for both measures (for DNHS, Moran's $I = 0.14$, 95% confidence interval: 0.07, 0.21; for CANVAS, Moran's $I = 0.20$, 95% confidence interval: 0.17, 0.23). For DNHS observations, the centroid of the measured block group was used as the location of measurement. For CANVAS observations, the midpoint of a line between the ends of the street segment acted as the measurement location. Nearly all streets in Detroit are straight; thus, this location typically fell in the midpoint of the street segment. Because the DNHS analysis aggregated measures to the block group before kriging, whereas the CANVAS analysis estimated the measure at the block-face level, the CANVAS spatial model was based on more observations than was the DNHS model (500 vs. 135). The decision to aggregate or not is a direct consequence of sampling plan chosen in each audit; a direct comparison of measures either as aggregated or not would thus not be appropriate.

Statistical analysis

First, as a rough assessment of the validity of the neighborhood audit processes, we identified streets audited by both surveys as those for which both the start points were within 20 m of each other and the end points were within 20 m of each other. Because the audits used different training materials to calibrate auditors and different protocols for assessment (DNHS assessed both sides of the street in a single audit whereas CANVAS assessed only one block face), we expected low agreement in general, even for roughly similar items. Moreover, because a year elapsed between audits, we expected weaker agreement for items for which physical presence would be less stable, such as street cleanliness and abandoned cars.

Next, because accuracy of physical disorder estimates rather than item-level reliability was our core focus, we used 2 methods to assess the similarities between the measures' estimates of the spatial distribution of physical disorder. First, we created a scatterplot of estimated physical disorder level at the centroids of block groups selected for the DNHS audit using the CANVAS measures. Second, using ordinary kriging to estimate in a 100-m grid over all of Detroit, we computed separate raster surfaces of estimated physical disorder using the DNHS data and the CANVAS data and compared maps of the 2 measures.

Next, to assess predictive accuracy of the kriging models from the DNHS and CANVAS, we used "leave-one-out cross-validation" (36); that is, we computed the root mean square of the differences between the measured value at each sampled point and the estimate kriged for that point using all other measured values. From the resulting error distribution, we computed the root mean square error and median absolute deviation. To ensure comparability of empirical estimates across models, we transformed all measures to z scores before kriging. As a direct comparison of measure accuracy, we also computed mean squared error and median absolute deviation using the CANVAS measure to predict DNHS scores. For all mean squared error and median absolute deviation results, lower numbers indicate less error in the spatial model. Cross-validation thus provides an empirical estimate of error in the

spatial model; the distribution of the differences between what we actually observed and what we would have estimated at any given observed location had we not observed that location also estimates the error for unsampled points in the city.

Next, to test convergent validity, we estimated correlations between the physical disorder measure at each census tract centroid and 3 census-block level measures from the 2010 US Census (37) and 2009–2013 American Community Survey 5-year estimates (28) that were available at the census block–group level and thought to be correlated with disorder: housing vacancy (strongly positive expected), adult unemployment (moderately positive expected), and median home value (strongly negative expected).

Finally, we conducted a brief simulation to determine the amount of error induced by using one measure instead of the other. First, for each of the 879 block groups in Detroit, we simulated an outcome Y_i as a function of the DNHS disorder estimate for that block group and a standard normal error term, as follows:

$$Y_i = 0.5(\text{Disorder}_{\text{DNHS}}i) + N(0, 1).$$

Then we fitted a model using the simulated Y_i as the observed outcome variable and the CANVAS disorder measure as the observed disorder level:

$$Y_i = \beta(\text{Disorder}_{\text{CANVAS}}i) + \varepsilon.$$

The difference between 0.5 and the estimate for β then represents the extent to which the true association with the DNHS disorder measure would be obscured by the use of CANVAS instead. We repeated this simulation 1,000 times to explore the distribution of resulting β estimates.

Although the reliability and validity of the ordinary kriged measures described above formed the core of our investigation, we also explored a hybrid approach in which we combined the measures (Web Appendix 3). Hybrid approaches might be of interest for researchers looking to augment measures taken in targeted in-person audits (e.g., by interviewers visiting subject homes) with virtual audits intended to improve estimation accuracy over a wider area. For this investigation, we used co-kriging, a geospatial technique in which spatial estimation of a primary measure is augmented through cross-correlation with a secondary spatially located measure (38). We computed 2 co-kriged estimates: one in which the DNHS measure augmented the CANVAS measure and one in which the CANVAS measure augmented the DNHS measure.

Finally, although the comparison of the CANVAS measure as constructed to the DNHS measure as constructed addresses the goal of comparing 2 different final measures using best practices for each approach, comparing a CANVAS measure using the DNHS techniques to the DNHS measure removes measure construction as a source of heterogeneity. This analysis and its conclusions are described in detail in Web Appendix 4.

All analyses used 64-bit R for Windows, version 3.1.0 (R Foundation for Statistical Computing, Vienna, Austria). We used the "ltm" package to construct item response theory models and the "geoR" and "gstat" packages for kriging.

RESULTS

Efficiency, logistics, and related concerns

Including food and bathroom breaks, DNHS audits required 3,350 hours of auditor time in the field (Janie Slayden, University of Michigan, email communication, 2015), or slightly more than 24 minutes per pair per street segment. Because 16 members of the DNHS team were based in Ann Arbor, Michigan, which is roughly a 45-minute drive from Detroit, approximately 600 additional person-hours of time were spent in transit between Ann Arbor and Detroit. Including the reliability subsample, CANVAS audits required approximately 105 person-hours of auditor time, or a little less than 13 minutes per block face audited.

The DNHS training procedure involved 2 individuals training 16 auditors in one 8-hour day, after which subsequent untrained auditors were paired with trained auditors during audits and were trained on the job. CANVAS training required approximately 30 person-hours total for the 7 auditors, including time for street auditors to learn the CANVAS system and to apply audit items appropriately. CANVAS audits further required 5 dual-monitor computers at 5 cubicles in an academic office building, whereas DNHS audits further required the use of 2 vans to transport the audit teams to the locations to be audited.

Item-level reliability

Fifty-two street segments were audited in both surveys, though the CANVAS team audited only one block face of the segment. For the 4 audit items for which text and training materials were most comparable, κ coefficients ranged from −0.06 (for abandoned cars) to 0.43 (for boarded-up or abandoned buildings). Between-study pairwise κ coefficients for the 4 directly comparable items are shown in Table 1.

Composite physical disorder measures

Leave-one-out cross-validation indicated a moderate amount of in-sample prediction error in both the DNHS and CANVAS measures. Prediction error estimates from the CANVAS audit were somewhat smaller, indicating that the predictions were more internally reliable. Error estimates that were computed by

Table 1. κ Statistics for Comparable Items Assessed on Matching Street Segments From an In-Person Audit (Detroit Neighborhood Health Study, 2008) and a Virtual Audit (Computer Assisted Neighborhood Visual Assessment System, 2009), Detroit, Michigan

Item Assessed	κ	Audited Streets Where Indicator of Disorder Was Present, %	
		DNHS	CANVAS
Boarded up or abandoned buildings	0.43	57	35
Vacant lots	0.31	27	33
Clean street	0.08	69	92
Abandoned cars	−0.06	14	4

Abbreviations: CANVAS, Computer-Assisted Neighborhood Visual Assessment System; DNHS, Detroit Neighborhood Health Study.

Table 2. Leave-One-Out Cross-Validation Results From an In-Person Audit (Detroit Neighborhood Health Study, 2008) and a Virtual Audit (Computer Assisted Neighborhood Visual Assessment System, 2009), Detroit, Michigan

| Model	Number of Cross-Validated Locations	z Score for Root Mean Square Error	z Score for Median Absolute Deviation
CANVAS	502	0.72	0.49
DNHS	135	0.96	0.68
CANVAS predictions treating DNHS observations as truth	135	0.89	0.68
Co-kriging DNHS with CANVAS	135	0.86	0.69
Co-kriging CANVAS with DNHS	501	0.71	0.48

Abbreviations: CANVAS, Computer-Assisted Neighborhood Visual Assessment System; DNHS, Detroit Neighborhood Health Study.

subtracting CANVAS's estimated physical disorder level at each block group centroid in the DNHS audit sample from the actual DNHS physical disorder observation for that block group were slightly lower than the cross-validation error estimates from DNHS (Table 2), suggesting the measures assessed closely related underlying constructs and that the spatial coverage in CANVAS had increased accuracy compared with the more focused coverage in the DNHS. Co-kriging both measures together did not produce substantially less cross-validation error (Table 2 and Web Appendix 3).

Kriged physical disorder estimates at all 2010 census block–group centroids in Detroit (*n* = 879) from both measures were
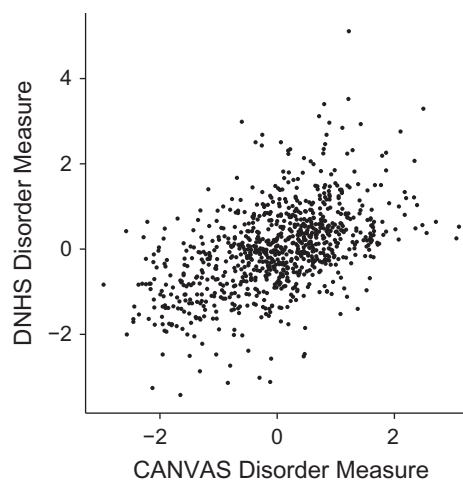
Figure 3. Scatterplot of disorder estimates (computed as a *z* score) at census block group centroids for the Detroit Neighborhood Health Study (DNHS) and the Computer-Assisted Neighborhood Visual Assessment System (CANVAS). The DNHS in-person neighborhood audit was conducted in the summer of 2008, and the CANVAS virtual audit used Google Street View imagery dating from the summer of 2009. The Spearman correlation coefficient for the measures was 0.52.
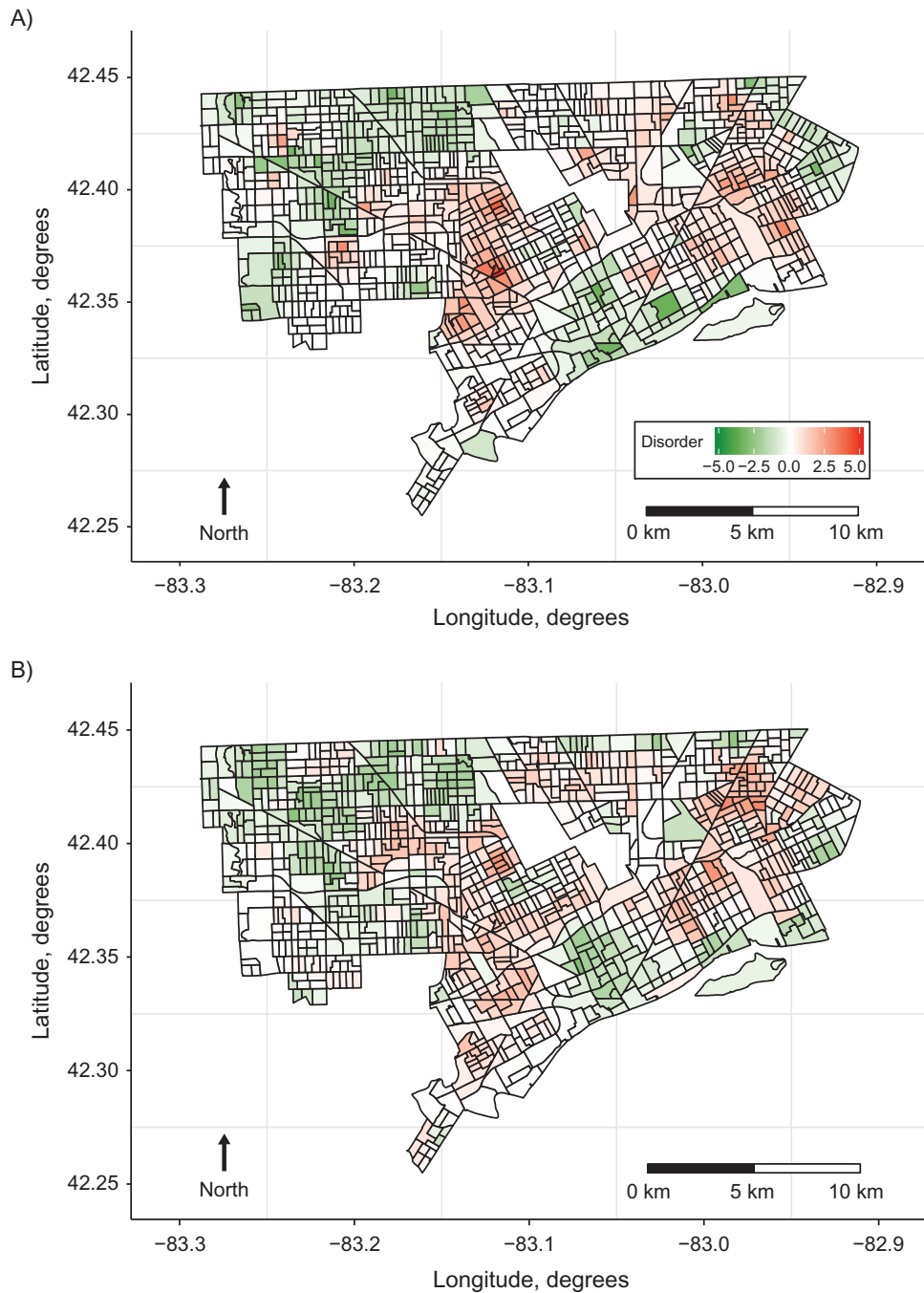
Figure 4. Estimates of disorder and block group boundaries in Detroit from the Detroit Neighborhood Health Study's 2008 in-person audit (A) and a 2009 virtual audit using the Computer-Assisted Neighborhood Visual Assessment (B).

positively but imperfectly correlated (Spearman's $\rho = 0.52$, Figure 3). Although maps of the 2 audits together revealed portions of the city in which the measures diverged, we could not discern systematic traits of the neighborhoods in which the measures diverged. Figure 4 displays 100-m grid kriged raster surfaces estimating physical disorder across the city using each survey's measures: Panel A shows DNHS estimates and panel B shows CANVAS estimates.

Convergent validity

Both measures varied as expected (and with nearly identical strength) with census measures hypothesized to correlate with neighborhood physical disorder (Table 3). For example, housing vacancy was correlated with the DNHS measure ($\rho = 0.47$) and with the CANVAS measure ($\rho = 0.44$). Finally, over 1,000 simulation runs, β coefficient estimates

Table 3. Correlation Coefficients Relating Each Disorder Measure as Computed at the Block Group Centroid to Characteristics From the US Census for Each of 876 Block Groups in Detroit, Michigan[a]

Census Characteristic	Expected Correlation	DNHS[b]	CANVAS[c]
Housing vacancy[d], %	Positive	0.47	0.44
Unemployment[e], %	Positive	0.17	0.18
Households below poverty line[e], %	Positive	0.29	0.27
Owner occupancy[d], %	Negative	−0.19	−0.14
Median home value[e], dollars	Negative	−0.28	−0.31

Abbreviations: CANVAS, Computer-Assisted Neighborhood Visual Assessment System; DNHS, Detroit Neighborhood Health Study.
[a] All coefficients significantly different from zero, with $P < 0.001$.
[b] In-person audit, 2008.
[c] Virtual audit, 2009.
[d] From the 2010 US Census estimates.
[e] From the 2009–2013 American Community Survey 5-year estimates.

ranged from 0.08 to 0.59 (mean, 0.34), indicating that 68% of the strength of association of the true measure would be recovered by the use of the substitute measure in an otherwise error-free study.

DISCUSSION

We compared spatially interpolated estimates from in-person and virtual audit measures of physical disorder in Detroit, Michigan, a city with substantial spatially patterned physical disorder. Although reliability of individual audit items ranged from poor to mediocre, the spatial distribution of physical disorder in Detroit computed using the 2 techniques was similar. Cross-validation results suggest that the benefits of a more evenly distributed spatial sample afforded by virtual audits outweigh the precision losses due to auditing fewer segments and fewer items within the scale. Ultimately, because of its smaller sample size, its use of one auditor per segment, and its lower per-segment audit time, the virtual audit required approximately only 3% of the total person-time investment that the in-person audit did. Overall, these findings support the use of virtual audits for estimating neighborhood physical disorder.

Where comparable items were assessed on streets audited by both groups, item-level reliability between the audit techniques was low. This likely represents several artifacts of measurement: 1) Auditors were trained differently, such that items may reflect different calibration levels (e.g., the CANVAS measure for litter was calibrated to detect even small amounts of litter, such as a discarded paper bag), 2) DNHS audited both sides of each street, whereas CANVAS audited only one, and 3) a year passed between audits. Detroit's foreclosure crisis began in 2006; demolition following abandonment may explain why vacant lot prevalence was similar between audits despite CANVAS's audit being limited to one side of the street. However, despite weak item-level reliability,

the estimates of disorder at block group centroids were strongly correlated, suggesting that the social processes underlying the development of neighborhood disorder were consistent over 2 years and strongly spatially patterned.

Although both audits aimed to assess physical disorder, the sample design and items used in the 2 audits were different. The DNHS sample was spatially clustered, a decision driven in part by the need to transport audit teams into the field. The CANVAS team audited far fewer segments overall and incorporated more audit items into the final measure. Although sampling density and item selection decisions are affected both by audit logistics and by resources available to commit to the audit, optimal spatial sampling for physical disorder audits is an area for future research.

Audit cost and precision of the final estimate are not the only criteria relevant to choosing a neighborhood audit technique. Whereas a virtual audit may be perceived as remote and technocratic, an in-person audit may help auditors forge connections with community members, develop field data collection skills, and generate hypotheses of future scientific importance. Furthermore, subject to logistic constraints, in-person audits may be able to control season, time of day, and day of week for audits, which may be important in some contexts (e.g., prevalence of litter may be affected by a neighborhood's garbage pickup day). Conversely, remote auditing removes researcher concern regarding physical safety of auditors, allows for real-time quality control and reliability metrics, and minimizes study disruptions due to weather or related logistical details. Furthermore, given Google's archive of past Street View images, future researchers may be able to use archived imagery to study neighborhood change retrospectively. Although these endpoints were not the primary focus of this analysis, researchers should not overlook them when choosing an audit strategy.

This study has several notable strengths. To our knowledge, this is the first study in which the key trade-off researchers must make in selecting an audit technique (i.e., cost, coverage, and validity) has been directly addressed. Our conclusions are further strengthened by the similar timeframes of audits and the use of identical spatial interpolation techniques. However, our findings should be considered in light of the following limitations. First, the 1-year gap between the in-person audit and the capture of most audited imagery means that direct comparisons between the items are limited. We note, however, that any differences that accrued between 2008 and 2009 would result in less similar measures; thus, the similarity we observed may underestimate the similarity we would have observed had we compared the in-person audit to an audit of imagery captured the same year. In addition, both audits captured the characteristics of a street segment at only one time point, which precluded capture of dynamic and shorter term changes in the environment such as neighborhood clean-up initiatives. Second, whereas DNHS audited both sides of the street, CANVAS audited only one, making direct comparison of item-level results on identical units impossible. Third, there is no gold-standard measure of disorder (14); therefore, we were able only to compare the measures with each other and cannot provide evidence that either measure more fully captures the construct of physical disorder. Fourth, we observe that resident perceptions of disorder may differ from research

team observations because of neighborhood stigma and related subject-specific factors ([13](#), [39](#)) and that for some health behaviors and outcomes, subject perceptions may be more relevant than independent observations ([40](#)). Nonetheless, because independent observations of physical disorder avoid the risk of recall bias intrinsic to perceived measures ([41](#), [42](#)), they are commonly used ([14](#)).

In conclusion, the similarities we found between a measure of neighborhood physical disorder generated from in person audits and a measure developed from Google Street View imagery suggest that virtual audits are a viable and much less expensive alternative to in-person audits for physical disorder. By massively reducing the cost of measurement, virtual audits may revolutionize research determining how disorder affects health, thus setting the stage for effective public health interventions.

REFERENCES

1. Kawachi I, Berkman LF. *Neighborhoods and Health*. New York, NY: Oxford University Press; 2003.

2. Gordon-Larsen P, Nelson MC, Page P, et al. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*. 2006;117(2): 417–424.

3. Lovasi GS, Hutson MA, Guerra M, et al. Built environments and obesity in disadvantaged populations. *Epidemiol Rev*. 2009;31:7–20.

4. Wei E, Hipwell A, Pardini D, et al. Block observations of neighbourhood physical disorder are associated with neighbourhood crime, firearm injuries and deaths, and teen births. *J Epidemiol Community Health*. 2005;59(10):904–908.

5. Cerdá M, Ransome Y, Keyes KM, et al. Revisiting the role of the urban environment in substance use: the case of analgesic overdose fatalities. *Am J Public Health*. 2013;103(12): 2252–2260.

6. Evans GW. The built environment and mental health. *J Urban Health*. 2003;80(4):536–555.

7. Casciano R, Massey DS. Neighborhood disorder and anxiety symptoms: new evidence from a quasi-experimental study. *Health Place*. 2012;18(2):180–190.

8. Leventhal T, Brooks-Gunn J. Moving to opportunity: an experimental study of neighborhood effects on mental health. *Am J Public Health*. 2003;93(9):1576–1582.

9. Wilson JQ, Kelling GL. Broken windows. *The Atlantic*. March, 1982:29–38.

10. Skogan W. Fear of crime and neighborhood change. *Crime Justice*. 1986;8:203–229.

11. Cohen D, Spear S, Scribner R, et al. "Broken windows" and the risk of gonorrhea. *Am J Public Health*. 2000;90(2):230–236.

12. Ross CE, Mirowsky J. Neighborhood disadvantage, disorder, and health. *J Health Soc Behav*. 2001;42(3):258–276.

13. Sampson RJ, Raudenbush SW. Seeing disorder: neighborhood stigma and the social construction of "broken windows". *Soc Psychol Q*. 2004;67(4):319–342.

14. Skogan W. Disorder and decline: the state of research. *J Res Crime Delinq*. 2015;52(4):464–485.

15. Reiss AJ. Systematic observation of natural social phenomena. *Sociol Methodol*. 1971;3:3–33.

16. Sampson RJ, Raudenbush SW. Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. *AJS*. 1999;105(3):603–651.

17. Anguelov D, Dulong C, Filip D, et al. Google street view: capturing the world at street level. *Computer*. 2010;43(6):32–38.

18. Charreire H, Mackenbach J, Ouasti M, et al. Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: a systematic review (the SPOTLIGHT project). *Health Place*. 2014;25:1–9.

19. Badland HM, Opit S, Witten K, et al. Can virtual streetscape audits reliably replace physical streetscape audits? *J Urban Health*. 2010;87(6):1007–1016.

20. Rundle AG, Bader MD, Richards CA, et al. Using Google Street View to audit neighborhood environments. *Am J Prev Med*. 2011;40(1):94–100.

21. Wilson JS, Kelly CM, Schootman M, et al. Assessing the built environment using omnidirectional imagery. *Am J Prev Med*. 2012;42(2):193–199.

22. Kelly CM, Wilson JS, Baker EA, et al. Using Google Street View to audit the built environment: inter-rater reliability results. *Ann Behav Med*. 2013;45(suppl 1):S108–S112.

23. Odgers CL, Caspi A, Bates CJ, et al. Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *J Child Psychol Psychiatry*. 2012;53(10):1009–1017.

24. Bader MD, Mooney SJ, Lee YJ, et al. Development and deployment of the Computer Assisted Neighborhood Visual

Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health Place*. 2015;31:163–172.

25. Keyes KM, McLaughlin KA, Koenen KC, et al. Child maltreatment increases sensitivity to adverse social contexts: neighborhood physical disorder and incident binge drinking in Detroit. *Drug Alcohol Depend*. 2012;122(1-2):77–85.

26. Bureau of the Census, US Department of Commerce. *Census 2000 gatewat*. https://www.census.gov/main/www/cen2000.html. Accessed June 8, 2017.

27. Mooney SJ, Bader MD, Lovasi GS, et al. Validity of an ecometric neighborhood physical disorder measure constructed by virtual street audit. *Am J Epidemiol*. 2014; 180(6):626–635.

28. Bureau of the Census, US Department of Commerce. *American Fact Finder*. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml. Accessed June 8, 2017.

29. Reichman NE, Teitler JO, Garfinkel I, et al. Fragile families: sample and design. *Child Youth Serv Rev*. 2001;23(4-5):303–326.

30. Frye V, Blaney S, Cerdá M, et al. Neighborhood characteristics and sexual intimate partner violence against women among low-income, drug-involved New York City residents results from the IMPACT studies. *Violence Against Women*. 2014; 20(7):799–824.

31. 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):1–41.

32. Clifton KJ, Smith ADL, Rodriguez D. The development and testing of an audit for the pedestrian environment. *Landscape Urban Plan*. 2007;80(1-2):95–110.

33. Day K, Boarnet M, Alfonzo M, et al. The Irvine-Minnesota inventory to measure built environments: development. *Am J Prev Med*. 2006;30(2):144–152.

34. Rizopoulos D. ltm: an R package for latent variable modeling and item response theory analyses. *J Stat Softw*. 2006;17(5): 1–25.

35. Bader MDM, Ailshire JA. Creating measures of theoretically relevant neighborhood attributes at multiple spatial scales. *Sociol Methodol*. 2014;44(1):322–368.

36. Davis BM. Uses and abuses of cross-validation in geostatistics. *Math Geol*. 1987;19(3):241–248.

37. Bureau of the Census, US Department of Commerce. *2010 Census summary file 1*. https://www.census.gov/prod/cen2010/doc/sf1.pdf. Published September 2012. Accessed June 8, 2017.

38. Bivand RS, Pebesma EJ, Gomez-Rubio V, et al. *Applied Spatial Data Analysis With R*. New York, NY: Springer; 2008.

39. Franzini L, Caughy MOB, Nettles SM, et al. Perceptions of disorder: contributions of neighborhood characteristics to subjective perceptions of disorder. *J Environ Psychol*. 2008; 28(1):83–93.

40. Blacksher E, Lovasi GS. Place-focused physical activity research, human agency, and social justice in public health: taking agency seriously in studies of the built environment. *Health Place*. 2012;18(2):172–179.

41. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol*. 1990;43(1):87–91.

42. Avolio BJ, Yammarino FJ, Bass BM. Identifying common methods variance with data collected from a single source: an unresolved sticky issue. *J Manage*. 1991;17(3):571–587.