HHS Public Access
Author manuscript
Am J Prev Med. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:
Am J Prev Med. 2017 January ; 52(1 Suppl 1): S20–S30. doi:10.1016/j.amepre.2016.06.010.

A Reliable, Feasible Method to Observe Neighborhoods at High Spatial Resolution

Maura M. Kepper, MPH[1], **Melinda S. Sothern, PhD**[1], **Katherine P. Theall, PhD**[2], **Lauren A. Griffiths, MPH**[1], **Richard Scribner, MD, MPH**[3], **Tung-Sung Tseng, PhD**[1], **Paul Schaettle, BA**[2], **Jessica M. Cwik, MSW**[4], **Erica Felker-Kantor, MSPH, MA**[2], and **Stephanie T. Broyles, PhD**[5]

[1]Louisiana State University Health Sciences Center, School of Public Health, Behavioral and Community Health Sciences Department, New Orleans, Louisiana

[2]Tulane University School of Public Health and Tropical Medicine, Global Community Health and Behavioral Sciences, New Orleans, Louisiana

[3]Louisiana State University Health Sciences Center, School of Public Health, Department of Epidemiology, New Orleans, Louisiana

[4]Tulane University School of Public Health and Tropical Medicine, Department of Epidemiology, New Orleans, Louisiana

[5]Pennington Biomedical Research Center, Contextual Risk Factors, Baton Rouge, Louisiana

Abstract

Introduction—Systematic social observation (SSO) methods traditionally measure neighborhoods at street level and have been performed reliably using virtual applications to increase feasibility. Research indicates that collection at even higher spatial resolution may better elucidate the health impact of neighborhood factors, but whether virtual applications can reliably capture social determinants of health at the smallest geographic resolution (parcel level) remains uncertain. This paper presents a novel, parcel-level SSO methodology and assesses whether this new method can be collected reliably using Google Street View and is feasible.

Methods—Multiple raters (N=5) observed 42 neighborhoods. In 2016, inter-rater reliability (observed agreement and kappa coefficient) was compared for four SSO methods: (1) street-level in person; (2) street-level virtual; (3) parcel-level in person; and (4) parcel-level virtual. Intra-rater reliability (observed agreement and kappa coefficient) was calculated to determine whether parcel-level methods produce results comparable to traditional street-level observation.

Results—Substantial levels of inter-rater agreement were documented across all four methods; all methods had >70% of items with at least substantial agreement. Only physical decay showed

---

Address correspondence to: Maura M. Kepper, MPH, Louisiana State University Health Sciences Center, School of Public Health, 2020 Gravier Street, Lions Building Office 225, New Orleans LA 70112. mmohle@lsuhsc.edu.

No financial disclosures were reported by the authors of this paper.

higher levels of agreement (83% of items with >75% agreement) for direct versus virtual rating source. Intra-rater agreement comparing street- versus parcel-level methods resulted in observed agreement >75% for all but one item (90%).

Conclusions—Results support the use of Google Street View as a reliable, feasible tool for performing SSO at the smallest geographic resolution. Validation of a new parcel-level method collected virtually may improve the assessment of social determinants contributing to disparities in health behaviors and outcomes.

Introduction

Neighborhood social determinants of health have been shown to impact the biological processes and behavioral risk factors during childhood and may promote chronic diseases later in life, such as obesity, diabetes, and cardiovascular disease.[1, 2] Health outcomes may be influenced by safety concerns (e.g., crime, victimization, poorly lit streets); the social environment (e.g., perceived collective efficacy, social capital, trust among neighbors); the built environment (e.g., access to parks/playgrounds, sidewalks, walkability); or disorder (e.g., graffiti, litter, dilapidated homes).[3–6] The challenges of measuring social and physical factors and determining the geographic resolution that captures the environment in which an individual lives have limited public health research efforts and produced inconsistent results.[7] The purpose of this work was to determine if virtual applications can used to collect neighborhood data at high spatial resolution to improve sensitivity to individual health outcomes and maximize data flexibility for specific research questions.

Systematic social observation (SSO) measures social determinants by overcoming limitations of common environmental measures, such as the individual's perceptions of their environment and archival sources like Census data.[8] However, the majority of SSO work has focused on: a street segment (section of a street between adjacent intersections or intersection and dead end), a block face (one side of a street segment); or an entire block (four street segments).[12] A neighbourhood described at the smallest geographic unit possible (parcel level) may produce data suitable for microlevel studies looking at individual health outcomes. Studies have identified three important considerations that support the need for SSO performed at the parcel level[9–11]:

1. Spatial resolution is a key factor contributing to differences in environmental data and its relationship to outcomes.
2. There are differences in costs and benefits of higher- versus lower-resolution data.
3. Higher spatial resolution that increases variation may be necessary for microlevel research studies.

Performing SSO at higher spatial resolution (i.e., parcel level) allows researchers to differentiate the health impact of detailed neighborhood factors that are under the control of the individual (i.e., their house and yard condition) while still capturing more-distal factors (i.e., land use or street condition).

Technological advancements have enabled virtual data collection and increased feasibility and reliability of observing neighborhoods for street-level SSO.[8, 13–17] For example, Odgers and colleagues[8] demonstrated that a neighborhood characterized as the most disadvantaged using Census data also displayed the highest levels of SSO-rated disorder, decay, and dangerousness by raters taking a virtual walk down the street. This paper presents a new SSO methodology for collecting data on neighborhood social determinants of health at the parcel level using virtual applications and assesses whether this new method using Google Street View (GSV) is more feasible (less cost and time) and yet as reliable as direct observation.

Methods

Study Population

Data were collected in Southeast Louisiana as part of two cross-sectional-longitudinal research projects. Residential addresses of children (N=170) participating in either the Molecular and Social Determinants of Health in Developing Youth study (NIMHHD 5U54MD008176-02) or the Neighborhood Stress and Physiology Among Children study (NIEHS R01ES020447) were collected to characterize neighborhood environments.

Each participant's street segment of residence was eligible for observation. Probability sampling methods were used to randomly select 54 participants, each contributing one street segment (N=54), across three levels of poverty and three levels of land use. Poverty categories were defined as Census tracts with:

1. ≥20% of individuals below the poverty threshold (high poverty);
2. 10%–19.9% of individuals below the poverty threshold (middle poverty); or
3. <10% of individuals below the poverty threshold (low poverty), based on the 2005–2009 American Community Survey.[18]

Land use categories were defined using GSV evaluation:

1. all parcels classified as residential land use (residential);
2. one or more parcel classified as commercial/business, industrial building, or institution (commercial); or
3. one or more parcel classified as a vacant lot/open space, recreational facility, park, or playground (recreational).

The authors attempted to select six street segments per poverty-by-land use cell; however, inadequate numbers within each cell limited effort to ensure equal distribution (Appendix 1). The final street segment sample (n=42) reflected mixed poverty level and land use and allowed instrument reliability to be determined across urban and rural neighborhood types (Table 1).

Survey Development

A parcel- and street-level instrument was developed using online software (QuestionPro, version 14.2.3) to measure aspects of the neighborhood physical and social environment using both direct and virtual observation. The survey derived questions from and was comparable in length to the Project on Human Development in Chicago Neighborhoods (PHDCN) study and SSO Inventory: Tally of Observations in Urban Regions.[8, 12] These street-level surveys were adapted for parcel-level data collection to ensure the same outcome measures could be derived for comparison. The resulting parcel-level survey adequately addressed the feasibility limitations (e.g., travel time, staff costs) by employing GSV and eliminated subjectivity and difficulty by requiring responses about a single parcel rather than the entire street segment. Previous research on the reliability of using GSV to perform neighborhood audits reported lower inter-rater reliability for conditions requiring detailed observations at the street level (i.e., presence of garbage), which suggested these observations may be less obvious in a GSV image, street-level responses are more subjective, and detailed data are more difficult to collect.[8] A content validity analysis of the parcel-level instrument involved extensive review of the literature by the survey developer on the concepts and items within each domain. An interdisciplinary group of subject matter experts reviewed the survey content to reveal linguistic errors, advise on relevance, and sort items. The final survey was comprised of questions that measured physical disorder, physical decay, safety, street safety, and land use.

Raters were recruited from Louisiana State University Health Sciences Center School of Public Health and Tulane University School of Public Health and Tropical Medicine. All raters were master's- or doctoral-level graduate students or Masters of Public Health. Raters were trained in a 4-hour classroom-based session regarding the use of data collection software, the procedure for performing observations, and definitions of parcel- and street-level attributes. Raters were considered proficient when they had general agreement of virtual observations among themselves and the supervisor. An online procedure manual was created to provide raters access to directions for rating and definitions for response options that included example images to increase inter-rater reliability. The procedure manual was used during training, and raters could access the manual online or offline during virtual or direct observations.

Data Collection

Each street segment was assessed four ways by two independent raters:

1. direct SSO at the street segment level;
2. direct SSO at the parcel level;
3. virtual SSO at the street segment level; and
4. virtual SSO at the parcel level.

The order in which the four ratings were performed on each street segment was randomized, and the two independent raters never performed the same sequence of the four ratings on a single street segment. Randomization of rating order was necessary to reduce rater's (or experimenter's) bias, as knowledge gained once the rater performed one method of observation may have biased the time taken and responses given to complete subsequent methods.

All parcels were surveyed on street segments with ≤20 parcels. If >20 parcels, random selection was used to select 20 parcels for observation. Parcels were defined as a distinct, continuous portion, or tract of land that may or may not include a building and may vary in size. To perform direct observation, raters traveled to the neighborhood. To perform virtual observation, raters used GSV's rotational and zoom capabilities to take a virtual walk down the participant's street. Feasibility was measured as the difference between the minutes required to perform a virtual rating versus a direct rating including travel time.

Observations were performed over 13 weeks in August to November 2015. Five raters completed a total of 188 parcel-level ratings (or 2,437 individual parcels) and 192 street-level ratings. All ratings (n=8) were completed for 42 (78%) street segments. Twelve street segments were excluded for:

1. safety concerns when performing direct observation (n=2);
2. GSV not available on the street segment or unable to identify the participant's residence (n=4);
3. structures were not visible for observation using GSV owing to objects (i.e., trees) obstructing view (n=2); or
4. missing data (n=4).

The GSV images were dated 2011–2015, with an average of 23 months (range, 2–56 months) elapsed between the GSV image and the in-person SSO data collection.

Statistical Analysis

Summary measures for each street segment were created. Parcel-level data were aggregated to street level. Signs of physical disorder included the presence of garbage/litter on the street, in residential yards, and on commercial/businesses/industrial properties (rated 0–4: heavy, moderate, light, none; dichotomized for analysis); graffiti or graffiti that had been painted over on buildings or signs (rated 0–1); and residential porches that were cluttered with personal items (rated 0–1). Signs of physical decay included the condition (rated 1–4: poor/badly deteriorated, fair, well-kept/good, not present; dichotomized as good/well kept versus fair and poor/badly deteriorated for analysis) of residential units and yards, commercial/businesses, industrial units, vacant lots, sidewalks, and streets. Signs of street safety included the presence of speed limit signs (rated 0–1), lighting (rated 1–4: >75%, 50%–74%, 25%–49%, <25%; dichotomized <50% for analysis), and bike lanes (rated 0–1). Safety was assessed based on raters' perceptions of whether the neighborhood was a "safe place to live" or whether the rater would feel "safe walking at night" (rated 1–5; definitely safe to definitely unsafe; dichotomized as safe versus unsafe for analysis). Land use was categorized as residential, commercial/business, industrial, vacant lot/open space, recreational facility, or other (dichotomized for analysis). For parcel-level data, the number of parcels in each land use category was divided by the total parcels on the street segment to determine the percentage land use (>0%=present). Descriptive information for each SSO measure is reported in Table 2.

The authors examined the inter-rater reliability (observed agreement and simple kappa coefficient [SAS FREQ procedure, AGREE option]) for both street-level and street-level characteristics aggregated from parcel-level data observed directly (in-person) and virtually (GSV). Intra-rater reliability (observed agreement and simple kappa coefficient) was calculated to compare street-level to street-level outcomes aggregated from parcel-level data collected by the same rater using both direct and virtual sources to determine whether parcel-level methods produced results comparable to more traditional street-level SSO. Substantial agreement was defined as >75%. All analyses were conducted using SAS, version 9.3 in 2016.

Results

The prevalence for each item by rater, method (street and parcel level), and source (direct and virtual) is presented in Table 2. Observed agreement and kappa values (κ) for inter-rater reliability for all four methods (direct street and parcel level and virtual street and parcel level) are presented in Table 3. The domain with the highest levels of agreement was land use; all items exceeded 75% agreement and inter-rater agreement was substantial (all κ>0.61) for all items except "vacant lots observed directly at the street level." Across all four methods, all items within physical decay (except measurement of vacant lot deterioration) showed agreement >75%; however, only six items (25%) had substantial kappa values. A summary of items with agreement >75% (Table 4) illustrates three main findings. First, levels of observed inter-rater agreement were comparable across all four methods; inter-rater agreement was similar (71%–75%) for all methods. Second, physical decay showed higher levels of agreement (83% of items with >75% agreement) when collected directly, although inter-rater agreement for items from other domains was similar across direct and virtual data collection methods. Third, inter-rater agreement was comparable for all domains.

Intra-rater agreement comparing street- versus parcel-level methods exceeded 75% for all but one item (assessing vacant lot condition) (Table 3). Overall, 100% of physical disorder, safety, street safety, and land use items exceeded 75% agreement for both direct and virtual data collection, with 73% of items having kappa values for intra-rater reliability indicate substantial agreement (κ>0.61).

Feasibility measured as travel time averaged 18.9 minutes for direct ratings. The time to perform street-level ratings virtually (7.8 minutes) was 18.5 minutes faster than direct ratings including travel time (26.4 minutes) (Appendix Table 2). The time to perform parcel-level ratings virtually (16.1 minutes) was 13.5 minutes faster than direct ratings including travel time (34.9 minutes) (Appendix Table 2). Parcel-level ratings took an average of 11.1 minutes longer than street-level ratings.

Discussion

A parcel-level SSO method using a virtual data source (GSV) produced reliable results when compared to three previously validated methods: direct street-level observation, virtual street-level observation, and direct parcel-level observation. Substantial levels of inter-rater agreement were documented across all four methods; all methods had >70% of items with at least substantial agreement. Furthermore, intra-rater reliability comparing street- and parcel-level measures resulted in at least 90% agreement for both direct and virtual data collection. Thus, street- and parcel-level observation methods produced similar outcomes. Observations collected virtually did not differ from direct observation, yet were more feasible.

Use of SSO is valuable for assessing social determinants at higher spatial resolutions, which may enhance understanding of mechanisms by which the physical and social environment influence health behaviors and outcomes. This study adds to the literature by determining whether SSO performed at different spatial resolutions (street versus parcel) can be collected reliably using different sources (direct versus virtual). This new SSO method that assesses the environment at the parcel level within a virtual context appears reliable for assessing markers of disorder, decay, safety, and land use. However, the current study was performed in Southeast Louisiana and may not be generalizable to studies performing SSO in other regions. Similar to previous SSO studies, limitations that may explain lower agreement include difficulty identifying detailed signs of physical disorder within a virtual context, time lapse between the street imagery and direct observation that may affect more-fluid characteristics (e.g., garbage or litter), lack of coverage by GSV for all streets (especially small or newly developed streets) and natural barriers (i.e., trees) may obstruct the view of more-detailed data, and subjective judgment on quality for certain variables.[8, 13–15, 23] Some individuals (raters) may be more likely to perceive disorder or adverse conditions (i.e., safety) than others, and raters' perceptions may differ if other individuals are nearby.[24] Thus, factors such as gender, ethnicity, and knowledge of the local area/previous exposure to neighborhood disorder may influence observations. Despite some variations, levels of agreement were similar between virtually and directly collected data, which may be due to a smaller interval (on average 23 months) between direct and virtual data collection than reported in previous studies.[14]

Of particular interest was whether the use of GSV increased the feasibility of performing high–spatial resolution SSO. Regardless of resolution, observational data require upfront costs for GIS expertise, software for data collection/entry, and study staff. The primary additional cost of obtaining observational data at high resolution is time.[9] Parcel-level ratings took an average of 11.1 minutes longer than street-level ratings. However, virtual observation increased feasibility compared with both direct parcel- and street-level rating methods. Therefore, although the time to complete parcel-level ratings was longer than street-level ratings overall, parcel-level ratings performed virtually averaged 5 minutes faster than the traditional direct street-level ratings (Appendix Table 2). Furthermore, parcel-level rating time was comparable to other instruments designed for observational data at lower resolutions (street level), which ranged from 10.6 to 20 minutes per segment.[9] In this study, street-level observations averaged only 7.6 minutes to complete. Although the order in which ratings were performed was randomized, cases where street-level ratings followed previous ratings may have experienced greater bias than parcel-level observations, as street-level measures were fewer, less detailed, and easier to remember than parcel-level measures. Thus, rating time likely decreased for subsequent street-level ratings more than subsequent parcel-level ratings, which could have exaggerated the difference between average rating times for street- and parcel-level ratings. Nevertheless, the use of GSV reduced the time necessary to perform ratings compared with direct observation. Ultimately, this study demonstrated that GSV reduced barriers to performing observations at high spatial resolutions; time demands were comparable to traditional SSO methods.

This method was developed to collect data on social determinants of health at the smallest geographic unit possible (parcel level). Parcel-level data were aggregated to street-level outcomes to compare outcomes across methods. However, aggregating parcel-level data to a lower resolution (street level) results in the loss of some detail of the parcel-level data, which may limit ability to detect relationships with behavioral outcomes. For example, parcel-level data produced the number and percentage of residential units on the street that were in good, fair, or poor condition. For comparison purposes, these data were aggregated to the street-level survey, which asks: *In general, how would you rate the condition of most of the residential units in the block face?* Similarly, Leonard et al.[9] reported that detail was lost when parcel-level data were averaged to higher levels of aggregation, which produced less variability among observations, and may not have identified relationships between neighborhood environment and outcomes measures. However, parcel-level data collected on a single street segment may provide a myopic view that is not indicative of the entire neighborhood environment (e.g., access to healthy food or physical activity environments that may not be on an individual's street). Future research should benefit from obtaining small area geo-referenced data to allow exploration of associations to health behaviors (i.e., physical activity) and outcomes (i.e., obesity, inflammatory, and metabolic markers) that may differ based on measurement resolution (street- versus parcel-level) and different aggregations derived from parcel-level data. Parcel-level data provide flexibility to measure social determinants of health in relation to a specific research question, which is necessary because the relevant geographic scale is likely to differ by the environmental variable of interest (e.g., walkability), behaviors of interest (e.g., walking versus biking), population (e.g., age group, those with or without access to automobiles), and calls for researchers to report results using multiple geographic scales.[10]

Conclusions

This study validated a new parcel-level method that collects data virtually. Parcel-level virtual SSO methods may be particularly useful to understand the complex interactions of how the environment "gets under the skin" to predispose to metabolic dysfunction associated with obesity, diabetes, cardiovascular disease, and cancer.[2, 25–27] Observational methods that reliably assess microlevel neighborhood factors must keep pace with the rapid development of this research area in order to isolate mechanisms through which neighborhood social determinants influence behavior and health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the Schools of Public Health and Medicine, Department of Pediatrics, Louisiana State University Health Sciences Center (LSUHSC)–New Orleans; the LSUHSC School of Public Health Jim Finks Endowed Chair in Health Promotion Research Fund; Louisiana Cancer Research Center; Nutrition Obesity Research Center at Pennington Biomedical Research Center; National Institute of Diabetes and Digestive and Kidney Diseases (Clinical Nutrition Research Unit) 1P30 DK072476; R01 HD49046; National Institute on Minority Health and Disparities Mid-South Transdisciplinary Collaborative Centers grant 5U54MDO08176-02; and National Institute of Environmental Health Sciences R01ES020447. STB is supported in part by 1 U54 GM104940 from the National Institute of General Medical Sciences of NIH, which funds the Louisiana Clinical and Translational Science Center.

References

1. Zenk SN, Schulz AJ, Mentz G, et al. Inter-rater and test-retest reliability: methods and results for the neighborhood observation checklist. Health Place. 2007; 13(2):452–465. http://dx.doi.org/10.1016/j.healthplace.2006.05.003. [PubMed: 16809060]
2. Miller GE, Chen E, Parker KJ. Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. Psychol Bull. 2011; 137(6):959–997. http://dx.doi.org/10.1037/a0024768. [PubMed: 21787044]
3. Kimbro RT, Brooks-Gunn J, McLanahan S. Young children in urban areas: Links among neighborhood characteristics, weight status, outdoor play, and television watching. Soc Sci Med. 2011; 72(5):668–676. http://dx.doi.org/10.1016/j.socscimed.2010.12.015. [PubMed: 21324574]
4. Cradock A, Kawachi I, Colditz G, et al. Neighborhood social cohesion and youth participation in physical activity in Chicago. Soc Sci Med. 2009; 68(3):427–435. http://dx.doi.org/10.1016/j.socscimed.2008.10.028. [PubMed: 19036490]
5. Foster S, Giles-Corti B. The built environment, neighborhood crime and constrained physical activity: An exploration of inconsistent findings. Prev Med. 2008; 47(3):241–251. http://dx.doi.org/10.1016/j.ypmed.2008.03.017. [PubMed: 18499242]
6. McNeill L, Kreuter M, Subramanian S. Social environment and physical activity: A review of concepts and evidence. Soc Sci Med. 2006; 63(4):1011–1022. http://dx.doi.org/10.1016/j.socscimed.2006.03.012. [PubMed: 16650513]
7. Diez Roux AV, Mair C. Neighborhoods and Health. Ann NY Acad Sci. 2010; 1186:125–245. http://dx.doi.org/10.1111/j.1749-6632.2009.05333.x. [PubMed: 20201871]
8. Odgers CL, Caspi A, Bates CJ, Sampson RJ, Moffit TE. 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. http://dx.doi.org/10.1111/j.1469-7610.2012.02565.x. [PubMed: 22676812]
9. Leonard TCM, Caughy MO, Mays JK, Murdoch JC. Systematic neighborhood observations at high spatial resolution: methodology and assessment of potential benefits. Plos One. 2011; 6(6):e20225. http://dx.doi.org/10.1371/journal.pone.0020225. [PubMed: 21673983]
10. Brownson RC, Hoehner CM, Day K, Forseth A, Sallis JF. Measuring the built environment for physical activity: state of the science. Am J Prev Med. 2009; 36(Suppl 4):S99–S123. http://dx.doi.org/10.1016/j.amepre.2009.01.005. [PubMed: 19285216]
11. Schaefer-McDaniel N, Caughy MO, O'Campo PH, Geary W. Examining methogological details of neighborhood observations and the relationship to health: A literature review. Soc Sci Med. 2010; 70(2):277–292. http://dx.doi.org/10.1016/j.socscimed.2009.10.018. [PubMed: 19883966]
12. Earls, FJ.; Raudenbush, SW.; Reiss, AJ.; Sampson, RJ. Project on Human Development in Chicago Neighborhoods (PHDCN): Systematic Social Observation. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research; 2005. http://doi.org/10.3886/ICPSR13578.v1
13. Badland H, Opit S, Witten K, Kearns R, Mavoa S. Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? J Urb Health. 2010; 87(6):1007–1016. http://dx.doi.org/10.1007/s11524-010-9505-x.
14. Clarke P, Ailshire J, Melendez R, Bader M, Morenoff J. Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit instrument. Health Place. 2010; 16(6):1224–1229. http://dx.doi.org/10.1016/j.healthplace.2010.08.007. [PubMed: 20797897]

15. Rundle AG, Bader MD, Richards CA, Neckerman KM, Teitler JO. Using Google Street View to Audit Neighborhood Environments. Am J Prev Med. 2011; 40(1):94–100. http://dx.doi.org/10.1016/j.amepre.2010.09.034. [PubMed: 21146773]
16. Charreire H, Mackenbach JD, Ouasti M, et al. Using remote sensing to define environmental characteristics related to physical activity and dietary behaviors: a systematic review (the spotlight project). Health Place. 2014; 25:1–9. http://dx.doi.org/10.1016/j.healthplace.2013.09.017. [PubMed: 24211730]
17. Kelly CM, Wilson JS, Baker EA, Miller DK, Schootman M. Using Google Street View to Audit the built environment: inter-rater reliability results. Ann Behav Med. 2013; 45(Suppl 1):S108–S112. http://dx.doi.org/10.1007/s12160-012-9419-9. [PubMed: 23054943]
18. U.S. Census Bureau American Community Survey. 2013. www.census.gov/acs/www/
19. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assessment, Research and Evaluation. 2004; 9(4):1–19.
20. Hartmann DP. Considerations in the choice of interobserver reliability estimates. J Appl Behav Anal. 1977; 10(1):103–116. http://dx.doi.org/10.1901/jaba.1977.10-103. [PubMed: 16795538]
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1):159–174. http://dx.doi.org/10.2307/2529310. [PubMed: 843571]
22. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990; 43(6):543–549. http://dx.doi.org/10.1016/0895-4356(90)90158-L. [PubMed: 2348207]
23. Griew P, Hillsdon M, Foster C, et al. Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. Int J Behav Nutr Phys Act. 2013; 10:103–110. http://dx.doi.org/10.1186/1479-5868-10-103. [PubMed: 23972205]
24. Sampson RJ, Raudenbush RW. Seeing disorder: Neighborhood stigma and the social construction of “broken windows”. Soc Psychol Q. 2004; 67(4):319–342. http://dx.doi.org/10.1177/019027250406700401.
25. Godfrey KM, Lillycrop KA, Burge GC, Gluckman PD, Hanson MA. Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. Pediatr Res. 2007; 61(5 Pt 2):5R–10R. http://dx.doi.org/10.1203/pdr.0b013e318045bedb.
26. Kuzawa CW, Sweet E. Epigenetics and the embodiment of race: developmental origins of U.S. racial disparities in cardiovascular health. Am J Hum Biol. 2009; 21(1):2–15. http://dx.doi.org/10.1002/ajhb.20822. [PubMed: 18925573]
27. Essex ML, Boyce WT, Hertzman C, et al. Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. Child Dev. 2013; 84(1):58–75. http://dx.doi.org/10.1111/j.1467-8624.2011.01641.x. [PubMed: 21883162]

Table 1

Descriptive Statistics for Selected Street Segments (n=42)

Descriptive	% or mean(range)
Land use[a]	
Residential[b]	33.3
Commercial/Business[c]	26.2
Recreational/Open Space[d]	40.8
Poverty[e]	
Low (<10%)	35.7
Medium (10%–19.99%)	26.2
High (≥20%)	38.1
% Black[f]	50.3 (0.0, 100)
% under the age 18[f]	20.9 (4.8, 32.1)
% female-headed household[f]	45.8 (0.0, 99.3)
% unemployed[f]	9.6 (2.0, 23.2)
% receiving public assistance[f]	12.3 (0.0, 54.5)

[a]Land use categories were defined using Google Street View to observe street segments.

[b]Residential land use only.

[c]Commercial/business land use had at least one parcel on the street segment as one of the following: commercial/institutional/industrial property.

[d]Recreational/open space land use had at least one parcel on the street segment as one of the following: vacant lots/open space/parks/recreational facilities.

[e]Poverty is defined as the % of the population in the census tract classified as below the official poverty threshold according to the 2005–2008 American Community Survey.[18]

[f]Characteristic of the census tract in which the street segment is located (2010 U.S. Census).

Table 2

Prevalence (%) of Street Segments Within Domains by Rater, Method (Street- and Parcel-Level) and Source (Direct and Virtual)

	Direct (n=168 ratings)				Virtual (n=168 ratings)			
	Street-level (n=84 ratings)		Parcel-level (n=84 ratings)		Street-level (n=84 ratings)		Parcel-level (n=84 ratings)	
Items	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)
Physical disorder								
Moderate/heavy garbage or litter	57.1	59.5	78.6	71.4	50.0	54.8	64.3	73.8
Graffiti present	16.7	9.5	14.3	12.2	14.3	4.8	9.5	16.7
Cluttered residential porches present	21.4	19.1	21.4	46.3	11.9	14.3	19.1	28.6
Physical decay								
Fair/poor residential condition[a]	22.5	24.4	23.8	34.2	12.5	22.0	16.7	23.8
Fair/poor yard condition[b]	22.5	24.4	9.8	15.0	12.5	22.0	5.0	15.0
Fair/poor businesses[c]	45.5	40.0	40.0	54.6	27.3	43.8	44.4	66.7
Fair/poor vacant lots[d]	27.8	38.1	16.7	33.3	25.0	38.9	11.1	38.9
Fair/poor street[e]	100.0	97.6	97.5	97.6	95.1	97.6	97.5	97.5
Fair/poor sidewalks[f]	86.5	94.7	83.8	92.1	91.9	94.9	91.9	97.4
Street safety								
Speed limit signs present	23.8	26.2	24.4	23.1	28.6	26.2	69.1	76.2
Bike lane present	2.4	0.0	2.4	2.4	2.4	2.4	2.4	0.0
>50% of street lit	50.0	50.0	51.2	59.5	40.5	52.4	64.3	52.4

	Direct (n=168 ratings)				Virtual (n=168 ratings)			
	Street-level (n=84 ratings)		Parcel-level (n=84 ratings)		Street-level (n=84 ratings)		Parcel-level (n=84 ratings)	
Items	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)	Rater 1 (n=42 ratings)	Rater 2 (n=42 ratings)
Safety								
Rated as unsafe place to live	24.3	36.6	25.0	40.5	13.2	26.5	17.5	34.2
Rated as unsafe to walk at night	23.7	38.9	30.0	52.8	17.1	47.4	21.6	48.6
Alarm systems/security signs present	59.5	59.6	71.4	61.0	53.7	54.8	61.9	59.5
Land use[g]								
Residential	100.0	100.0	100.0	100.0	100.0	10.00	100.0	100.0
Commercial, business	26.2	26.2	23.8	26.8	23.8	26.2	21.4	28.6
Industrial, warehouse, manufacturing	2.4	2.4	2.4	2.4	2.4	2.4	4.8	4.8
Institutional	9.5	11.9	9.5	9.8	11.9	11.9	11.9	11.9
Recreational facility	4.8	4.8	4.8	4.9	4.8	7.1	2.4	4.8
Vacant lot, open space	33.3	50.0	42.9	51.2	33.3	38.1	42.9	42.9

[a] Residential condition rated on a 4 point scale (poor/badly deteriorated; fair; good/well-kept; no residential units on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[b] Residential yard condition rated on a 4 point scale (poor/badly deteriorated; fair; good/wellkept; no residential yards on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[c] Commercial/business condition rated on a 4 point scale (poor/badly deteriorated; fair; good/well-kept; no commercial/businesses on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[d] Vacant lot condition was assessed using three question on physical features used to categorized as (poor/badly deteriorated; fair; good/well-kept; no commercial/businesses on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[e] Street condition were rated on a 4 point scale (poor/badly deteriorated; fair; good/well-kept; under construction; not present). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[f]Sidewalk condition were rated on a 5 point scale (poor/badly deteriorated; fair; good/well-kept; under construction; not present). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[g]Percent of street segments containing at least one parcel of specified land use category.

Table 3

Agreement for Each Item Within Domains by Method (Street- and Parcel-Level) and Source (Direct and Virtual)

	Inter-rater agreement[a] Direct observation (n=84 ratings)				Inter-rater agreement[a] Virtual observation (n=84 ratings)				Intra-rater agreement[b] Street- versus parcel-level Methods (n=84 ratings)			
	Street-level (n=42 ratings)		Parcel-level (n=42 ratings)		Street-level (n=42 ratings)		Parcel-level (n=42 ratings)		Direct (n=42 ratings)		Virtual (n=42 ratings)	
Items	% Agree	K	% Agree	K	% Agree	K	% Agree	K	% Agree	K	% Agree	K
Physical disorder												
Moderate/heavy garbage or litter	64	0.27	83	0.56	67	0.29	67	0.23	81	0.58	83	0.66
Graffiti present	88	0.48	83	0.27	93	--	93	0.69	92	0.62	92	0.59
Cluttered residential porches present	69	0.04	71	0.39	83	0.27	76	0.35	80	0.49	87	0.57
Physical decay												
Fair/poor residential condition[c]	83	0.52	76	0.42	80	0.32	83	0.49	79	0.44	81	0.37
Fair/poor residential yard condition[d]	83	0.52	90	0.55	80	0.32	85	0.20	86	0.55	90	0.58
Fair/poor businesses[e]	91	0.81	89	0.8	73	0.48	67	0.34	89	0.78	90	0.80
Fair/poor vacant lots[f]	65	0.37	67	0.22	53	0.15	59	0.05	70	0.36	66	0.27
Fair/poor street[g]	98	1.00	100	1.00	98	0.66	100	1.00	99	0.66	98	0.79
Fair/poor sidewalks[h]	86	0.23	86	0.38	86	−0.07	89	−0.04	95	0.72	96	0.65

Items	Inter-rater agreement[a] Direct observation (n=84 ratings)				Inter-rater agreement[a] Virtual observation (n=84 ratings)				Intra-rater agreement[b] Street- versus parcel-level Methods (n=84 ratings)			
	Street-level (n=42 ratings)		Parcel-level (n=42 ratings)		Street-level (n=42 ratings)		Parcel-level (n=42 ratings)		Direct (n=42 ratings)		Virtual (n=42 ratings)	
	% Agree	K	% Agree	K	% Agree	K	% Agree	K	% Agree	K	% Agree	K
Street safety												
Speed limit Signs present	83	0.56	78	0.51	88	0.70	93	0.82	88	0.73	95	0.88
Bike lane present	100	--	100	--	100	--	100	--	99	--	99	--
>50% of street lit	48	−0.05	56	0.13	64	0.29	55	0.08	90	0.81	86	0.71
Safety												
Rated as unsafe place to live	69	0.26	71	0.32	77	0.33	69	0.23	96	0.90	97	0.91
Rated as unsafe to walk at night	59	0.23	68	0.35	68	0.31	60	0.14	97	0.94	96	0.90
Alarm systems/security signs present	86	0.70	90	0.79	80	0.61	93	0.85	90	0.79	87	0.74
Land use[i]												
Residential	100	1.00	100	1.00	100	1.00	100	1.00	100	1.00	100	1.00
Commercial, business	100	1.00	93	0.81	93	0.81	93	0.81	96	0.91	95	0.87
Industrial, warehouse, manufacturing	100	--	100	--	100	--	95	--	100	---	98	---
Institutional	98	0.88	100	--	95	0.77	98	0.88	99	0.93	99	0.94
Recreational facility	100	--	100	--	98	--	100	--	100	--	98	--
Vacant lot, open space	79	0.57	93	0.85	90	0.79	95	0.90	94	0.88	93	0.85

Notes: Observed agreement >75% is considered substantial.[19, 20]

Kappa values: 0–0.20=poor, 0.21–0.40=slight, 0.40–0.61=moderate, 0.61–0.80=substantial, 0.80–1.000=almost perfect.[21] Kappa coefficients were not reported for items with a base rate <10% as the expected chance agreement is inflated and the Kappa is lowered in these cases.[22]

[a] Inter-rater agreement was measured as observed agreement and a simple kappa coefficient between two raters.

[b] Intra-rater agreement was measured as observed agreement and a simple kappa coefficient between virtual vs. direct ratings performed by the same rater.

[c] Residential condition rated on a 4 point scale (poor/badly deteriorated; fair; good/well-kept; no residential units on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[d] Residential yard condition rated on a 4 point scale (poor/badly deteriorated; fair; good/well-kept; no residential yards on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[e] Commercial/business condition rated on a 4 point scale (poor/badly deteriorated; fair; good/well-kept; no commercial/businesses on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[f] Vacant lot condition was assessed using three question on physical features used to categorized as (poor/badly deteriorated; fair; good/well-kept; no commercial/businesses on the street). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[g] Street condition were rated on a 4 point scale (poor/badly deteriorated; fair; good/well-kept; under construction; not present). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[h] Sidewalk condition were rated on a 5 point scale (poor/badly deteriorated; fair; good/well-kept; under construction; not present). A dichotomous variable was created good/well-kept vs. fair and poor/badly deteriorated. Prevalence reported for fair, poor/badly deteriorated.

[i] Percent of street segments containing at least one parcel of specified land use category.

Table 4

Summary: n(%) of Items Within Each Domain With Substantial Agreement[a]

		Inter-rater agreement[b] Direct observation		Inter-rater agreement[b] Virtual observation (n=84)		Intra-rater agreement[c] Street- versus parcel-level (n=84)	
Domain	**Total # of items[d]**	**Street-level n(%)**	**Parcel-level n(%)**	**Street-level n(%)**	**Parcel-level n(%)**	**Direct observation n(%)**	**Virtual observation n(%)**
Physical disorder	3	1(34)	2(67)	2(67)	2(67)	3(100)	3(100)
Physical decay	6	5 (83)	5(83)	4(67)	4(67)	5(83)	5(83)
Safety	3	1 (34)	1(34)	2(67)	1(34)	3(100)	3(100)
Street safety	3	2 (67)	2(67)	2(67)	2(67)	3(100)	3(100)
Land use	6	6(100)	6(100)	6(100)	6(100)	6(100)	6(100)
Total	21	15(71)	16(76)	16(76)	15(71)	20(95)	20(95)

[a]Substantial agreement is defined as greater than 75 % agreement from item-level agreement in Table 3.[19,20]

[b]Inter-rater agreement was measured as observed agreement and a simple kappa coefficient between two raters.

[c]Intra-rater agreement was measured as observed agreement and a simple kappa coefficient between virtual vs. direct ratings performed by the same rater.

[d]The total number of items indicates the number of survey questions that were measured within each domain. The total number of items is the denominator used to calculate the percent of items within each domain with substantial agreement.