Published in final edited form as:

Am J Prev Med. 2020 January ; 58(1): 152–160. doi:10.1016/j.amepre.2019.08.032.

Drop-and-Spin Virtual Neighborhood Auditing: Assessing Built Environment for Linkage to Health Studies

Jesse J. Plascak, PhD^{1,2}, Andrew G. Rundle, PhD³, Riddhi A. Babel, MPH¹, Adana A.M. Llanos, PhD, MPH^{1,2}, Celine M. LaBelle⁴, Antoinette M. Stroup, PhD^{1,2,5}, Stephen J. Mooney, PhD⁶

¹Department of Biostatistics and Epidemiology, School of Public Health, Rutgers, The State University of New Jersey, Piscataway, New Jersey

²Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey

³Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York

⁴Edward J. Bloustein School of Planning and Public Policy, Rutgers, The State University of New Jersey, New Brunswick, New Jersey

⁵New Jersey State Cancer Registry, New Jersey Department of Health, Trenton, New Jersey

⁶Department of Epidemiology, University of Washington, Seattle, Washington

Abstract

Introduction: Various built environment factors might influence certain health behaviors and outcomes. Reliable, resource-efficient methods that are feasible for assessing built environment characteristics across large geographies are needed for larger, more-robust studies. This paper reports the item response prevalence, reliability, and rating time of a new virtual neighborhood audit protocol, drop-and-spin auditing, developed for assessment of walkability and physical disorder characteristics across large geographic areas.

Methods: Drop-and-spin auditing, a method where a Google Street View (GSV) scene was rated by spinning 360° around a point location, was developed using a modified version of the virtual audit tool CANVAS. Approximately 8,000 locations within Essex County, New Jersey were assessed by 11 trained auditors. Thirty-two built environment items per a location within GSV were audited using a standardized protocol. Test–retest and inter-rater κ statistics were from a 5% subsample of locations. Data were collected in 2017–2018 and analyzed in 2018.

Results: Roughly 70% of GSV scenes had sidewalks. Among those, two thirds were in good condition. At least five obvious items of garbage or litter were present in 41% of GSV scenes.

Address correspondence to: Jesse J. Plascak, PhD, Department of Biostatistics and Epidemiology, School of Public Health, Rutgers, The State University of New Jersey, 683 Hoes Lane West, Room 209, Piscataway NJ 08854. jesse.plascak@rutgers.edu.

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

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Maximum test–retest reliability indicated substantial agreement ($\kappa = 0.61$) for all items. Inter-rater reliability of each item, generally, was lower than test–retest reliability. The median time to rate each item was 7.3 seconds.

Conclusions: Compared with segment-based protocols, drop-and-spin virtual neighborhood auditing is quicker and similarly reliable for assessing built environment characteristics. Assessment of large geographies may be more feasible with the use of drop-and-spin virtual auditing.

INTRODUCTION

Identifying environmental determinants of health is a priority for population health research and practice.¹⁻⁸ Modifiable built environment factors—for example, vacant lots, deteriorated infrastructure, presence of sidewalks—are associated with health behaviors/outcomes.^{1-3,9-11} Development of methods to assess built environment characteristics that result in reliable, accurate, and resource-efficient measurements is critical for population health research.

Virtual systematic social observation (vSSO), whereby trained auditors code online imagery from mapping programs such as Google Street View (GSV), is a promising method for characterizing built environments across broad geographic areas.^{9,12-14} A recent review summarizes 54 studies that used GSV in health-related research.¹⁴ Nearly all reviewed virtual audit tools yielded reliable and valid measures of walkability/pedestrian safety, physical disorder/aesthetics, or access to health-related resources (e.g., food, physical activity).¹⁴ A majority of reviewed studies audited segments within predefined regions surrounding the residence of participants of an extant health study (e.g., buffers).^{15,16} Every reviewed study reported reliability results based on audits of street segments.

Measures commonly assessed through vSSO, including walkability and physical disorder/aesthetics, have been found to spatially autocorrelate at distances up to approximately 1,000 meters.^{13,17-20} By definition, this spatial autocorrelation indicates that vSSO responses recorded within 1,000 meters of one another collect redundant information. Thus, traversing a segment might be an inefficient sampling scheme for conducting a neighborhood audit because spatially autocorrelated characteristics at the segment's beginning can predict what will be found at the end (the average 2010 U.S. block face is approximately 942 meters).^{21,22} The use of street segments as the sampling unit within vSSO studies may be a methodologic convention retained from in-person audit protocols—protocols from which virtual versions were adapted.

Drop-and-spin vSSO, a method where a rater/auditor is dropped into a GSV scene at a predetermined location and then collects data by spinning the GSV perspective 360° around that location, might be a quicker way to assess large areas compared with segment-based audits. Moreover, drop-and-spin vSSO protocols avoid challenges with date mixing: abrupt changes in image date and scene sometimes experienced when “walking” a street segment in GSV that is caused by merged imagery collected from different GSV cars that traversed the same street segment on different dates.²³ Auditors of segment-based vSSO protocols must be trained additionally to handle mixed-date scenes (e.g., treat segment as missing, ignore characteristics in certain scenes/dates), which complicate protocols and results.

This study investigated whether drop-and-spin vSSO reliably and efficiently collects neighborhood data across large geographic areas. This study reports the sampling strategy, resource utilization, item-specific prevalence, and reliability of a three-construct audit protocol applied across nearly 8,000 locations in Essex County, New Jersey using the previously validated CANVAS neighborhood audit system.²⁴

METHODS

Study Sample

The study region, Essex County, New Jersey, contains Newark—the most populous municipality in the state—plus numerous suburbs in the western part of the county (Figure 1). Geographic information system techniques were used to randomly generate candidate audit locations (the point on a street where an auditor can “drop”) across non-highway roads as determined by the U.S. Census’ 2016 TIGER road file (sample location details within Appendix Figure 1).²⁵ The 8,262 candidate locations within the main Essex County data set were, on average, 142 meters apart from the next nearest candidate location (SD=18 meters; range, 76–277 meters).^{13,17–19} Candidate audit locations of the reliability study were a 5% random subsample ($n=414$) of the 8,262 candidate locations of the main data set. No two auditors assessed the same location within the Essex County data set. Each of the 11 auditors assessed the same reliability audit locations.

Measures

Audited items were based on previously validated neighborhood audit tools,^{26–28} as well as three newly developed items (Table 1 and Appendix Table 1). Twenty items were related to constructs of walkability or pedestrian safety and 12 items were related to the construct physical disorder/aesthetics. Three new items were related to physical disorder/aesthetics (Appendix Table 1). A summary including each item name, full wording, brief description, categorization of responses, and overall frequency within the complete Essex County data set is in Table 1 and Appendix Table 1.

The 11 auditors were paid research assistants. Auditors were introduced to the study’s overall goal as well as construct definitions during a 1-hour orientation session. The 4-hour training session on a following day explained each of the 32 items, introduced CANVAS, and involved hands-on training based on 45 non–Essex County training locations throughout New Jersey. A protocol facilitated training and was a reference during auditing (Appendix Figure 2). CANVAS’s real-time inter-rater reliability calculations enabled training that was dynamic, tailored, and iterative as audit items with low agreement were focused on as training progressed. Auditing of the 8,262 candidate study locations proceeded after responses to items rated in the 45 test locations were as consistent as possible.

Items were divided by auditor teams to limit the number of different items rated per auditor. Reasons for this were twofold: first, to increase specialization, and potentially agreement, of items; and second, to reduce rating burnout by facilitating quicker movement through GSV locations. Walkability or pedestrian safety items ($n=20$) were split into sidewalk- ($n=10$) and

intersection-related ($n=10$) and physical disorder/aesthetics items ($n=12$) were rated as a group.

Virtual audits took place at the Rutgers School of Public Health, on dedicated workstations containing two monitors, one of which was 24" with 1,920 X 1,080 resolution for viewing GSV scenes. Assessment of the 8,262 candidate locations occurred in November 2017–June 2018. Reliability audits were conducted in March–June 2018.

Statistical Analysis

Frequencies of non-reliability sample item responses were calculated to provide overall estimates of item prevalence. Cohen's κ statistic was used to summarize test–retest reliability agreement.^{29,30} The test–retest reliability data sets were constructed by merging auditor responses from the main and reliability data sets. Although each auditor assessed nearly identical locations in the reliability subsample, the number of locations assessed within the main data set of approximately 8,000 varied by auditor based on their auditing rate. Moreover, responses of *cannot tell*, *not applicable*, or *skipped*—considered as missing values—varied across items. Thus, the number of test–retest pairs varied by auditor and item.

Average, pairwise Cohen's κ values summarized inter-rater reliability agreement. Inter-rater reliability data sets were created by merging auditor responses to reliability data sets within item groupings (i.e., merging sidewalk-related items separately from intersection-related items). Although the identical 414 reliability candidate locations were submitted to CANVAS for each auditor, variation in missing values, GSV image updates, and the protocol for returning audit locations submitted to CANVAS (appendix to Mooney et al.¹³) resulted in inter-rater reliability data locations that were less than the 414 submitted within all items. Resulting κ values were interpreted according to agreement ranges suggested by Landis and Koch³⁰ (<0, poor; 0.0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.0, almost perfect). Item-specific ranges of test–retest κ values were compared to inter-rater κ values to identify items with potential for reliable measurement (e.g., high maximum test–retest κ values), or increased inter-rater reliability with additional group-based training (e.g., large discrepancies between inter-rater and test–retest κ values). Item-specific median rating time was calculated from the audit location data set median rating time per location, a measure automatically collected by CANVAS (i.e., location data set median rating time / n items = item-specific median rating time). Data were managed and analyzed using SAS, version 9.4, R, version 3.5.1 (with the *IRR* package), and ArcGIS, version 10.3 in 2018.

RESULTS

Approximately 8,000 scenes were audited of the 8,262 candidate locations submitted to GSV (intersection-related: 7,988, sidewalk-related: 8,016, physical disorder-related: 7,989). Audited locations were fewer than candidate locations and varied by item groupings for various reasons: (1) GSV scene updates over time and various times at which candidate locations were uploaded to CANVAS; (2) inaccurate road files used in generating candidate locations; and (3) the CANVAS process for identifying an available GSV scene near a

candidate location incorporated a stochastic component that sometimes led to failure to re-identify the identical GSV scene at the identical location (appendix to Mooney and colleagues¹³). The median time to audit each item was 7.3 seconds. Median audit times varied by construct group: intersection-related, 6.4 seconds; sidewalk-related, 6.5 seconds; and physical disorder-related, 8.6 seconds. GSV image dates ranged from August 2007 to May 2018 with 37% of all images dated from August 2013 to October 2013. Time duration between image date and rating date ranged 1 month to 10.7 years, with a median of 4.3 years.

Less than one fifth of assessed intersections had a pedestrian crossing warning sign (18.0%) or a pedestrian signal (15.2%) (Table 1). Nearly two thirds of intersections intended for pedestrian crossing had some type of pedestrian crossing marks. Sidewalks were present in 68.4% of GSV scenes. Approximately two thirds of these sidewalks were in good condition (e.g., even and without cracks). Presence of obstructions on the sidewalk were recorded in <10% of all sidewalks for each of the four obstruction types (cars, garbage cans, signs, other obstruction). At least five items of garbage, litter, or broken glass were observed in 41% of all GSV scenes. Prevalence of abandoned cars or burned out buildings was very low: 0.2% and 0.3%, respectively. Nearly 86% of buildings and 88% of yards were considered to be in either very well- or moderately well-kept condition. Evidence of graffiti was observed in 11% of GSV scenes. Presence of team sports equipment in public spaces was rare (2.4%). Half of scenes had some form of exterior yard decoration and about 25% of scenes were found to contain outdoor seating in private spaces (Appendix Table 1).

Tables 2 and 3 summarize the researchers' reliability findings. The maximum test-retest reliability values indicated at least substantial agreement ($\kappa = 0.61$) for nearly all items (Table 2). Moreover, the lowest, item-specific test-retest κ indicated at least moderate agreement ($\kappa = 0.41$) for all items except "presence of a highway" (0.31), "sidewalk to curb distance" (0.39), and "yard condition" (0.31). There was, however, variability in test-retest reliability among construct groupings. Generally, test-retest reliability was highest among the intersection-related walkability and pedestrian safety items. Time duration between rating identical GSV scenes ranged from 0 to 194 days with a mean of 45 days (SD=35 days). Rater-specific average time durations between ratings of identical GSV scenes were calculated to explore the relationship between test-retest κ values and duration of time between ratings. A Wilcoxon rank-sum test comparing time duration between rating identical GSV scenes by minimum and maximum κ values indicated little difference between rater-specific time duration by κ values (normal approximation Z-value=0.82; results not shown).

Inter-rater reliability analysis indicated at least moderate levels of agreement for all items with the exception of "presence of highway" (0.35), "sidewalk to curb distance" (0.09), "pole or sign obstruction" (0.19), and "team sports equipment" (0.29) (Table 3 and Appendix Table 3). Similar to test-retest reliability, inter-rater κ values varied by construct groupings. Inter-rater κ values ranged from 0.35 to 0.88 for intersection-related walkability and pedestrian safety items, 0.09 to 0.95 for sidewalk-related walkability and pedestrian safety items, and 0.29 to 0.72 for physical disorder and aesthetics items. Moreover, the frequency of items with inter-rater $\kappa = 0.61$ was 7/10 among intersection-related walkability

and pedestrian safety items, 2/10 among sidewalk-related walkability and pedestrian safety items, and 2/12 among physical disorder and aesthetics items. Items within the intersection-related walkability and pedestrian safety construct seemed to be more reliable than those in the other construct groupings.

Comparison of inter-rater κ values (Table 3 and Appendix Table 3) and minimum test–retest κ values (Table 2 and Appendix Table 2) yielded numerous instances of at least substantial agreement within raters (test–retest agreement) ($\kappa_{\min} = 0.61$) but no greater than moderate agreement between raters ($\kappa < 0.41$). Items meeting these criteria included (test–retest $\kappa_{\min}$, inter-rater κ): “team sports equipment” (0.74, 0.29), “building conditions” (0.64, 0.40), “curb cuts” (0.85, 0.50), “pole or sign obstruction” (0.66, 0.19), and “car obstruction” (0.66, 0.41).

DISCUSSION

This study reports the prevalence of indicators of intersection-, sidewalk- and disorder-related conditions in Essex County, New Jersey using a new vSSO approach. Drop-and-spin virtual neighborhood auditing was found to be efficient and reliable, suggesting that vSSO audits can efficiently estimate walkability or physical disorder constructs across large geographic regions (larger than county/metropolitan area). The high test–retest and inter-rater reliability values observed in this study indicated that the 20 items measuring walkability and pedestrian safety and 12 measuring physical disorder and aesthetics can be reliably replicated. Comparisons of the magnitudes of these two measures of reliability values indicated several items—“team sports equipment,” “building conditions,” “curb cuts,” “pole or sign obstruction,” and “car obstruction”—for which inter-rater concordance of item responses might be improved with additional auditor training. These situations indicated item ratings that can be replicated reliably by the same rater but not between different raters, potentially due to each rater consistently following different decision rules for choosing a response to each of the above audit items. The median time per item to rate a typical 360° GSV scene using this drop-and-spin method was just 7.3 seconds.

Comparisons to other reliability studies of neighborhood audits using GSV are limited owing to between-study variability in audited items and differences in reliability measures or evaluation process. Few studies have reported process information necessary to calculate a summary measure of rating time.^{9,16,24,31-39} Previously reported rating times per item range widely from an average of 7.5 seconds³² to 135 seconds,³¹ with the majority of studies reporting 15 seconds per item.^{16,24,34,36-39} Assuming these reporting times are representative of all GSV studies, the observed 7.3 seconds per item measured via the drop-and-spin protocol is among the fastest virtual neighborhood audit methods. Although the 7.7-second difference in rating time per item in the majority of studies versus drop-and-spin seems short (15 seconds – 7.3 seconds = 7.7 seconds), differences over an entire study comparing methods can be large as total rating time is a function of rated items, rated locations, and rating time per item. For example, if segment-based auditing was used in this study for the identical rating time (i.e., approximately 519 hours), <3,900 locations would have been rated.

It is likely that per-item rating times were shorter in this study because the drop-and-spin technique bypassed the need to “walk” the GSV street segment. This saves mouse clicks and assesses a smaller area, allowing speedier audit times. Moreover, because only one GSV scene is used for the full audit, ambiguities are avoided that arise when an assessed item is visible in one scene but not a differently dated scene within the same segment.²³

Both the observed test–retest and inter-rater reliability are similar to those found in other GSV neighborhood audit studies. Broadly, this study supports previous findings of lower reliability among items that are more subjective (e.g., condition judgment) or difficult to ascertain remotely within static images such as those item responses that are more dependent on large object size, color variation, or camera angle, including building/yard/sidewalk condition, presence of garbage or litter, and presence of curb cuts.^{12,13,32,34,35,40} This study also confirmed that items assessing the presence of large, consistently placed, unambiguous objects have higher reliability, including presence of sidewalks, traffic signals, and pedestrian signs or signals.^{16,24,35,41} The three items developed for this neighborhood audit —“team sports equipment,” “yard decorations,” and “fences”—yielded mixed findings. Both test–retest and inter-rater concordance were at least moderate for both “yard decorations” and “fences,” but inter-rater κ of “team sports equipment” was only 0.29. As noted above, however, the substantial-to-perfect test–retest agreement of “team sports equipment” suggests that inter-rater reliability can be improved with additional group-based training with the goal of each auditor using the same decision rules.

Limitations

This study has several important limitations, most of which relate to reliance on GSV as a source of data for assessment of neighborhood characteristics. As proprietary data, Google Inc. does not readily make available their GSV image collection protocol, including: exact image date/time (currently only month and year); timing of image acquisition, processing, and posting online; protocol of image updates and processing (e.g., when, where, priority locations, objects or items selected for censoring); protocol for drivers while gathering images (e.g., route taken, time of day restrictions, weather restrictions, seasonal restrictions); and, most importantly, the permanence of GSV image availability, both historical and future image posting. Several of these factors can influence the occurrence and prevalence of neighborhood characteristics commonly audited.¹⁴ Inherent to any efficiency-related strengths of this drop-and-spin method is the notion of item response spatial autocorrelation. Although segment-based studies have found significant spatial autocorrelation,^{13,17-19} drop-and-spin studies should also quantify the distances at which audit items spatially correlate. This is important to all vSSO techniques, as a previous study demonstrated that recent images are more common in urban areas whereas older images are over-represented in rural areas, potentially leading to exposure misclassification bias if temporal trends in built environment characteristics vary by urban–rural area.²³ Similarly, drop-and-spin audits should investigate whether resulting audit constructs are as valid as those found in segment-based audits.^{19,42} Lastly and as noted in previous studies, virtual neighborhood audits are inferior to in-person audits for assessing non-visual cues (e.g., sounds and smells), characteristics that might rely more on visual perspective (e.g., object size, color variation,

angle), and if image dates vary substantially across the study area or are incongruent with other data sets (i.e., epidemiologic data sets).^{23,32,33,40,42}

CONCLUSIONS

Efficient and reliable methods for assessing specific neighborhood characteristics continue to remain a priority as evidence emerges of the role of built environment factors in population health and the increased availability of built environment data through online platforms such as GSV. The drop-and-spin approach to conducting virtual neighborhood audits has numerous advantages compared with the traditional, segment-based auditing approach. Drop-and-spin auditing of these 32 items related to walkability/pedestrian safety and physical disorder/aesthetics appears to be a reliable and resource-efficient means to conducting virtual neighborhood audits across large geographic areas. Drop-and-spin reliability replication studies are needed prior to recommending widespread implementation of this neighborhood audit method. Future studies should also seek to formally test the spatial characteristics of audit results, including not only profiling the autocorrelation structure but also assessing predictive accuracy of interpolation models to better inform the choice of drop-and-spin versus segment-based auditing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This study was supported by funds from the Cancer Institute of New Jersey Cancer Prevention and Control pilot award (P30CA072720-19 to JJP) and National Cancer Institute (K07CA222158-01 to JJP). This study was also partly supported by the Columbia Population Research Center (P2CHD058486 to AGR), the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (1R01HD087460-01 to AGR), and the National Library of Medicine (1K99LM012868 to SJM).

REFERENCES

1. Bancroft C, Joshi S, Rundle A, et al. Association of proximity and density of parks and objectively measured physical activity in the United States: a systematic review. *Soc Sci Med*. 2015;138:22–30. 10.1016/j.socscimed.2015.05.034. [PubMed: 26043433]
2. Gomez SL, Shariff-Marco S, DeRouen M, et al. The impact of neighborhood social and built environment factors across the cancer continuum: current research, methodological considerations, and future directions. *Cancer*. 2015;121(14):2314–2330. 10.1002/cncr.29345. [PubMed: 25847484]
3. Branas CC, South E, Kondo MC, et al. Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. *Proc Natl Acad Sci USA*. 2018;115(12):2946–2951. 10.1073/pnas.1718503115. [PubMed: 29483246]
4. Osypuk TL, Joshi P, Geronimo K, Acevedo-Garcia D. Do social and economic policies influence health? A review. *Curr Epidemiol Rep*. 2014;1(3):149–164. 10.1007/s40471-014-0013-5. [PubMed: 25984439]
5. National Academies of Sciences, Engineering, and Medicine. A Framework for Educating Health Professionals to Address the Social Determinants of Health. Washington, DC: The National Academies Press; 2016 10.17226/21923.
6. National Academies of Sciences, Engineering, and Medicine. Bringing public health into urban revitalization: Workshop in brief. Washington, DC: The National Academies Press; 2016 10.17226/21847.

7. Robert Wood Johnson Foundation. Built environment and health: How our communities are planned, designed and built can have a major influence on our health. www.rwjf.org/en/our-focus-areas/topics/built-environment-and-health.html. Published 2018 Accessed August 19, 2019.
8. HHS. Physical activity: Built environment approaches combining transportation system interventions with land use and environmental design Healthy People 2020. Washington, DC: HHS; 2016.
9. Odgers CL, Caspi A, Bates CJ, Sampson RJ, Moffitt 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. 10.1111/j.1469-7610.2012.02565.x. [PubMed: 22676812]
10. Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*. 2012;125(5):729–737. 10.1161/circulationaha.110.969022. [PubMed: 22311885]
11. Mayne S, Jose A, Mo A, et al. Neighborhood disorder and obesity-related outcomes among women in Chicago. *Int J Environ Res Public Health*. 2018;15(7):1395 10.3390/ijerph15071395.
12. 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. 10.1016/j.healthplace.2010.08.007. [PubMed: 20797897]
13. Mooney SJ, Bader MD, Lovasi GS, Neckerman KM, Teitler JO, Rundle AG. Validity of an ecometric neighborhood physical disorder measure constructed by virtual street audit. *Am J Epidemiol*. 2014;180(6):626–635. 10.1093/aje/kwu180. [PubMed: 25122584]
14. Rzotkiewicz A, Pearson AL, Dougherty BV, Shortridge A, Wilson N. Systematic review of the use of Google Street View in health research: major themes, strengths, weaknesses and possibilities for future research. *Health Place*. 2018;52:240–246. 10.1016/j.healthplace.2018.07.001. [PubMed: 30015181]
15. Mooney SJ, Bader MD, Lovasi GS, Neckerman KM, Rundle AG, Teitler JO. Using universal kriging to improve neighborhood physical disorder measurement. *Sociol Methods Res*. In press. Online 5 6, 2018 10.1177/0049124118769103.
16. Wilson JS, Kelly CM, Schootman M, et al. Assessing the built environment using omnidirectional imagery. *Am J Prev Med*. 2012;42(2):193–199. 10.1016/j.amepre.2011.09.029. [PubMed: 22261217]
17. Kruger DJ, Reischl TM, Gee GC. Neighborhood social conditions mediate the association between physical deterioration and mental health. *Am J Community Psychol*. 2007;40(3–4):261–271. 10.1007/s10464-007-9139-7. [PubMed: 17924185]
18. Hunter MCR, Brown DG. Spatial contagion: gardening along the street in residential neighborhoods. *Landsc Urban Plan*. 2012;105(4):407–416. 10.1016/j.landurbplan.2012.01.013.
19. Duncan DT, Aldstadt J, Whalen J, Melly SJ. Validation of walk scores and transit scores for estimating neighborhood walkability and transit availability: a small-area analysis. *GeoJournal*. 2013;78(2):407–416. 10.1007/s10708-011-9444-4.
20. Plascak JJ, Llanos AA, Chavali LB, et al. Sidewalk conditions in northern New Jersey: using Google Street View imagery and ordinary kriging to assess infrastructure for walking. *Prev Chronic Dis*. 2019;16:E60 10.5888/pcd16.180480. [PubMed: 31095921]
21. Griffith DA. Effective geographic sample size in the presence of spatial autocorrelation. *Ann Assoc Am Geogr*. 2005;95(4):740–760. 10.1111/j.1467-8306.2005.00484.x.
22. Donaldson K How big is your neighborhood? U.S. Census Bureau. SEHSD Working Paper #FY2013-064; 2013.
23. Curtis JW, Curtis A, Mapes J, Szell AB, Cinderich A. Using Google Street View for systematic observation of the built environment: analysis of spatio-temporal instability of imagery dates. *Int J Health Geogr*. 2013;12:53 10.1186/1476-072x-12-53. [PubMed: 24298903]
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. 10.1016/j.healthplace.2014.10.012. [PubMed: 25545769]

25. U.S. Census Bureau and Department of Commerce, Economics and Statistics Administration 2016 topologically integrated geographic encoding and referencing (TIGER) system. Washington, DC: U.S. Census Bureau; 2018 <ftp://ftp2.census.gov/geo/pvs/tiger2010st/>.
26. Earls FJ, Brooks-Gunn J, Raudenbush SW, Sampson RJ. Project on human development in Chicago neighborhoods: community survey, 1994–1995 10.3886/icpsr25762.
27. Clifton KJ, Smith ADL, Rodriguez D. The development and testing of an audit for the pedestrian environment. *Landsc Urban Plan.* 2007;80(1–2):95–110. 10.1016/j.landurbplan.2006.06.008.
28. Day K, Boarnet M, Alfonzo M, Forsyth A. The Irvine–Minnesota inventory to measure built environments: development. *Am J Prev Med.* 2006;30(2):144–152. 10.1016/j.amepre.2005.09.017. [PubMed: 16459213]
29. Cohen J A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20(1):37–46. 10.1177/001316446002000104.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174. 10.2307/2529310. [PubMed: 843571]
31. Badland HM, Opit S, Witten K, Kearns RA, Mavoa S. Can virtual streetscape audits reliably replace physical streetscape audits? *J Urban Health.* 2010;87(6):1007–1016. 10.1007/s11524-010-9505-x. [PubMed: 21104331]
32. Bethlehem JR, Mackenbach JD, Ben-Rebah M, et al. The SPOTLIGHT virtual audit tool: a valid and reliable tool to assess obesogenic characteristics of the built environment. *Int J Health Geogr.* 2014;13:52 10.1186/1476-072x-13-52. [PubMed: 25515179]
33. Griew P, Hillsdon M, Foster C, Coombes E, Jones A, Wilkinson P. 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 10.1186/1479-5868-10-103. [PubMed: 23972205]
34. Gullón P, Badland HM, Alfayate S, et al. Assessing walking and cycling environments in the streets of Madrid: comparing on-field and virtual audits. *J Urban Health.* 2015;92(5):923–939. 10.1007/s11524-015-9982-z. [PubMed: 26349472]
35. Kepper MM, Sothorn MS, Theall KP, et al. A reliable, feasible method to observe neighborhoods at high spatial resolution. *Am J Prev Med.* 2017;52(1 suppl 1):S20–S30. 10.1016/j.amepre.2016.06.010. [PubMed: 27989289]
36. Mooney SJ, DiMaggio CJ, Lovasi GS, et al. Use of Google Street View to assess environmental contributions to pedestrian injury. *Am J Public Health.* 2016;106(3):462–469. 10.2105/ajph.2015.302978. [PubMed: 26794155]
37. Mygind L, Bentsen P, Badland H, Edwards N, Hooper P, Villanueva K. Public open space desktop auditing tool—establishing appropriateness for use in Australian regional and urban settings. *Urban For Urban Green.* 2016;20:65–70. 10.1016/j.ufug.2016.08.001.
38. Pliakas T, Hawkesworth S, Silverwood RJ, et al. Optimising measurement of health-related characteristics of the built environment: comparing data collected by foot-based street audits, virtual street audits and routine secondary data sources. *Health Place.* 2017;43:75–84. 10.1016/j.healthplace.2016.10.001. [PubMed: 27902960]
39. Vanwolleghem G, Ghekiere A, Cardon G, et al. Using an audit tool (MAPS global) to assess the characteristics of the physical environment related to walking for transport in youth: reliability of Belgian data. *Int J Health Geogr.* 2016;15:41 10.1186/s12942-016-0069-1. [PubMed: 27846881]
40. 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.* 2012;45(suppl 1):S108–S112. 10.1007/s12160-012-9419-9.
41. Nesoff ED, Milam AJ, Pollack KM, et al. Novel methods for environmental assessment of pedestrian injury: creation and validation of the inventory for pedestrian safety infrastructure. *J Urban Health.* 2018;95(2):208–221. 10.1007/s11524-017-0226-2. [PubMed: 29442222]
42. Mooney SJ, Bader MD, Lovasi GS, et al. Street audits to measure neighborhood disorder: virtual or in-person? *Am J Epidemiol.* 2017;186(3):265–273. 10.1093/aje/kwx004. [PubMed: 28899028]

Figure 1.
Population density by census tract, New Jersey, U.S. Census 2010 Decennial Census.

Table 1.

Item Details and Response Frequency of Neighborhood Characteristics Audited in Essex County, New Jersey, 2017–2018

Item name	Full question	Item summary	Categorization	% 'Yes' / '1'
Walkability/pedestrian safety: intersection-related				
NJ.CANVAS.14	Can you clearly see an intersection?	Clear intersection	Yes (1) vs no (2)	60.1
PEDS.23.6	Is there a pedestrian crossing warning sign?	Pedestrian crossing sign	Yes (1) vs no (2)	18.0
PEDS.23.2	Is there a pedestrian signal?	Pedestrian signal	Yes (1) vs no (2)	15.2
Minn.Irvine.2	Consider the places that are intended for pedestrians to cross the street. Are these places marked for pedestrian crossing?	Pedestrian crossing marks	All/some (1) vs none (2)	64.0
Minn.Irvine.4	How is the road marked at crosswalks?	Type of pedestrian crosswalk marks	Painted solid lines (1), painted zebra lines (2), other (3)	68.5
Minn.Irvine.9	What kind of traffic signal can you see?	Traffic signal type	Traffic signal (1), stop sign (2), yield sign (3), multiple (4), none (5)	19.9
Minn.Irvine.19	Is the street a one-way or a two-way street?	One-way street	One-way (1) vs two-way (2)	10.7
Minn.Irvine.21	How many lanes are there for cars?	Number of lanes	Few (2) (1) vs many (2+) (2)	87.7
Minn.Irvine.92	Can you see a highway (elevated or below ground) here?	Presence of highway	This is a highway (1), present (2), not present (3)	5.4
Minn.Irvine.92.1	Is the highway a barrier to walking?	Highway barrier	Can be overcome/can be somewhat overcome (1) vs cannot be overcome/no destinations on far side (2)	80.1
Walkability/pedestrian safety: sidewalk-related				
PEDS.4.1	What type of sidewalk or path (paved or unpaved) is there?	Sidewalk presence	Sidewalk (1) vs none (2)	68.4
PEDS.12	Is the sidewalk complete/continuous?	Complete sidewalk	Complete (1) vs incomplete (2)	90.4
PEDS.6	In what condition is the sidewalk or pedestrian path?	Sidewalk condition	Good (1) vs fair/poor (2)	64.7
PEDS.10	How wide is the sidewalk?	Sidewalk width	4 ft. (1) vs <4 ft. (2)	84.1
PEDS.5	How far is the sidewalk or path from the curb?	Sidewalk from curb distance	5 ft. (1) vs <5 ft. (2)	98.7
PEDS.7.2	Do parked cars obstruct the path?	Car obstruction	Yes (1) vs no (2)	6.2
PEDS.7.4	Do garbage cans obstruct the path?	Garbage can obstruction	Yes (1) vs no (2)	5.5
PEDS.7.1	Do poles or signs obstruct the path?	Pole or sign obstruction	Yes (1) vs no (2)	4.8
PEDS.7.5	Does anything else obstruct the path?	Other obstruction	Yes (1) vs no (2)	8.1
Minn.Irvine.8	Do you see any curb cuts?	Curb cuts	Yes (1) vs no (2)	52.1
Physical disorder/aesthetics				
PHDCN.1	Is there garbage, litter, or broken glass in the street or on the sidewalks (5 items)?	Presence of garbage, litter, or glass	Yes (1) vs no (2)	41.2
PHDCN.4	Are there abandoned cars?	Abandoned cars	Yes (1) vs no (2)	0.2
PHDCN.5	How would you rate the condition of most of the buildings?	Building condition	Very well/moderately well kept (1) vs fair/poor condition (2)	86.3
NJ.CANVAS.15	How would you rate the condition of most of the yards?	Yard condition	Very well/moderately well kept (1) vs fair/poor condition (2)	87.8

Item name	Full question	Item summary	Categorization	% 'Yes' / '1'
Minn.Irvine.141	Are there dumpsters visible?	Dumpsters	1 (1) vs none (2)	8.9
PHDCN.3	Is there graffiti, or evidence of graffiti that has been painted over, on buildings, signs, or walls?	Presence of graffiti	Yes (1) vs no (2)	11.3
PHDCN.6	Do you see burned out buildings?	Burned buildings	Yes (1) vs no (2)	0.3
PHDCN.7	Do you see boarded up or abandoned buildings?	Abandoned buildings/boarded up	Yes (1) vs no (2)	6.2
NJ.CANVAS.13	Do you see any outdoor seating (e.g., benches, porch swings, restaurantseating)?	Outdoor seating	Yes (1) vs no (2)	27.1

Table 2.

Test–Retest Reliability of Neighborhood Characteristics Audited in Essex County, New Jersey, 2017–2018

Item summary	κ (Min)	κ (Max)
Walkability/pedestrian safety: intersection-related (3 raters)		
Clear intersection	0.745	0.854
Pedestrian crossing sign	0.435	0.818
Pedestrian signal	0.778	0.878
Pedestrian crossing marks	0.700	0.841
Type of pedestrian crosswalk marks	0.823	1.000
Traffic signal type	0.867	0.908
One-way street	0.756	0.890
Number of lanes	0.755	0.962
Presence of highway	0.311	0.797
Highway barrier	NA	NA
Walkability/pedestrian safety: sidewalk-related (2 raters)		
Sidewalk presence	0.984	0.986
Complete sidewalk	0.522	0.565
Sidewalk condition	0.552	0.699
Sidewalk width	0.596	0.680
Sidewalk from curb distance	0.392	NA
Car obstruction	0.658	0.852
Garbage can obstruction	0.524	0.825
Other obstruction	0.600	0.784
Pole or sign obstruction	0.655	0.789
Curb cuts	0.852	0.882
Physical disorder/aesthetics (3 raters)		
Presence of garbage, litter, or glass	0.516	0.807
Abandoned cars	NA	NA
Building condition	0.635	0.731
Yard condition	0.305	0.700
Dumpsters	0.662	0.671
Presence of graffiti	0.783	0.918
Burned buildings	NA	NA
Abandoned buildings/boarded up	0.790	0.800
Outdoor seating	0.752	0.853

κ , Cohen's Kappa; NA, too few observations to calculate statistic.

Table 3.

Inter-rater Reliability of Neighborhood Characteristics Audited in Essex County, New Jersey, 2017–2018

Item summary	κ
Walkability/pedestrian safety: intersection-related (raters=4)	
Clear intersection	0.645
Pedestrian crossing sign	0.686
Pedestrian signal	0.789
Pedestrian crossing marks	0.577
Type of pedestrian crosswalk marks	0.881
Traffic signal type	0.788
One-way street	0.741
Number of lanes	0.856
Presence of highway	0.346
Highway barrier	NA
Walkability/pedestrian safety: sidewalk-related (raters=3)	
Sidewalk presence	0.949
Complete sidewalk	0.594
Sidewalk condition	0.523
Sidewalk width	0.632
Sidewalk from curb distance	0.094
Car obstruction	0.407
Garbage can obstruction	0.423
Other obstruction	0.544
Pole or sign obstruction	0.188
Curb cuts	0.503
Physical disorder/aesthetics (raters=4)	
Presence of garbage, litter, or glass	0.472
Abandoned cars	NA
Building condition	0.401
Yard condition	0.434
Dumpsters	0.593
Presence of graffiti	0.637
Burned buildings	NA
Abandoned buildings/boarded up	0.722
Outdoor seating	0.574

κ , average pairwise Cohen's Kappa; NA, too few observations to calculate statistic.