Reliability and One-Year Stability of the PIN3 Neighborhood Environmental Audit in Urban and Rural Neighborhoods

Anna K. Porter · Fang Wen · Amy H. Herring ·
Daniel A. Rodríguez · Lynne C. Messer ·
Barbara A. Laraia · Kelly R. Evenson

Published online: 3 April 2018
© The New York Academy of Medicine 2018

Abstract Reliable and stable environmental audit instruments are needed to successfully identify the physical and social attributes that may influence physical activity. This study described the reliability and stability of the PIN3 environmental audit instrument in both urban and rural neighborhoods. Four randomly sampled road segments in and around a one-quarter mile buffer of participants' residences from the Pregnancy, Infection, and Nutrition (PIN3) study were rated twice, approximately 2 weeks apart. One year later, 253 of the year 1 sampled roads were re-audited. The instrument included 43 measures that resulted in 73 item scores for calculation of percent overall agreement, kappa statistics, and log-linear models. For same-day reliability, 81% of items had moderate to outstanding

kappa statistics (kappas ≥ 0.4). Two-week reliability was slightly lower, with 77% of items having moderate to outstanding agreement using kappa statistics. One-year stability had 68% of items showing moderate to outstanding agreement using kappa statistics. The reliability of the audit measures was largely consistent when comparing urban to rural locations, with only 8% of items exhibiting significant differences ($\alpha < 0.05$) by urbanicity. The PIN3 instrument is a reliable and stable audit tool for studies assessing neighborhood attributes in urban and rural environments.

Keywords Built environment · Physical activity · Active transport · Measurement · Walkability

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11524-018-0243-9>) contains supplementary material, which is available to authorized users.

A. K. Porter (✉) · F. Wen
Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina – Chapel Hill, 123 W Franklin St, Building C, Suite 410, Chapel Hill, NC 27516, USA
e-mail: akporter@email.unc.edu

A. H. Herring
Department of Statistical Science, Duke University, Durham, NC, USA

D. A. Rodríguez
Department of City and Regional Planning, Institute for Transportation Studies, University of California, Berkeley, Berkeley, CA, USA

L. C. Messer
Oregon Health & Science University - Portland State University School of Public Health, Portland, OR, USA

B. A. Laraia
Division of Community Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, CA, USA

K. R. Evenson
Department of Epidemiology, Gillings School of Global Public Health, Center for Health Promotion and Disease Prevention, University of North Carolina – Chapel Hill, Chapel Hill, NC, USA

Introduction

Ecological models posit that there are multiple determinants of health behaviors, including intrapersonal, interpersonal, organization, community, and public policy [1]. In the past four decades, there has been concerted research devoted to measuring physical attributes of neighborhoods in an effort to identify conditions associated with health behaviors, specifically physical activity. From a community standpoint, the physical environment in which people reside may help or hinder physical activity, as the behavior occurs many times in space and place [2]. Indeed, many aspects of the physical environment are designed with physical activity in mind, such as in the case of bicycle lanes and public parks [3].

In the quest to identify the most salient factors of the physical environment that promote physical activity, a number of different assessment methods and measures have been developed. One such method is environmental auditing [4]. Environmental audits are tools designed to assess the physical and social spaces within a community, such as parks, trails, or the neighborhood. In the case of a neighborhood physical environment audit, auditors systematically observe and record characteristics of the environment for a designated space, such as on a street-by-street basis within a certain area around the location of interest. Although audits occur mostly by foot or in a motor vehicle, desktop audits reliant on digital photography (e.g., Google Street View) are increasingly common [5–9]. A review of the literature identified 29 neighborhood physical environment audit tools designed for in-person assessment, published between 2001 and 2014 [10–39].

Of the 29 identified neighborhood physical environment audit tools, 90% ($n = 26$) reported on the reliability of the measurement tool [10, 13–18, 20–24, 26–29, 32–37, 39]. External reliability is an important consideration with assessment tools such as neighborhood physical environment audits, to determine if the measures are consistent between (inter-) and within (intra-) raters, as well as from one time to another (i.e., test-retest). Among studies reporting external reliability, only one exclusively assessed reliability compared to a gold standard [17]. All other studies reported on inter-rater

reliability [10–16, 18–39], and a subset of these ($n = 5$) also assessed intra-rater reliability [14, 19, 23, 24, 36]. Of the studies that reported on inter-rater reliability, 48% ($n = 12$) assessed simultaneous audits [13, 14, 18, 19, 21, 22, 26–29, 32], one assessed audits from an approximate one-week interval [12], one assessed audits from an approximate two-week interval [15], and one assessed audits from an approximate two-month interval [10]. For the remaining 8 studies, the time frame for assessing reliability was unclear [16, 20, 23, 24, 34–37].

In addition to reliability, stability of the measures over time is important to consider when determining at what point and how often the neighborhood physical environment should be assessed for research studies. It is expected that neighborhood environments will change over time, but an understanding of the time frame for that change within different aspects of the environment, both social and physical/structural, is warranted. As far as the authors are aware, to date, no neighborhood audit instrument has been assessed for stability of the measures over a long time frame (e.g., 1 year). Further, contextual and methodological factors may influence reliability and stability metrics. For example, urbanicity may influence the reliability of a measure. Of the 29 neighborhood physical environment audit tools examined, only two audits examined both urban and rural locations [13, 29], and of these, only the Wisconsin Assessment of Social and Built Environment (WASABE) [13], examined if there were differences in urban and rural environments.

The Pregnancy, Infection, and Nutrition (PIN3) study developed the PIN3 neighborhood audit instrument to assess road segments near participants' home addresses, in both urban and rural areas. Since its development, the PIN3 neighborhood audit instrument has been used either entirely or in part to assess the neighborhood environment for other research projects [12, 15, 40, 41]. Given the need for more detailed information on the reliability and stability of neighborhood physical environment audit instruments, the primary aim of this study was to assess the inter-rater reliability over same-day and two-week time frames, as well as the one-year stability of the PIN3 neighborhood audit instrument items. In addition, this study assessed the PIN3 audit tool as to whether inter-rater reliability differed between urban and rural locations.

Methods

Study Sample

The PIN3 Study was designed to identify risk factors associated with preterm birth. Participants of the PIN3 study resided primarily in central North Carolina, specifically in the counties of Alamance, Chatham, Durham, and Orange. These counties were the focus of the neighborhood audit. More information on the study can be found at <http://www.cpc.unc.edu/pin>. This study was reviewed and approved by the Institutional Review Board at the University of North Carolina at Chapel Hill (UNC).

PIN3 Neighborhood Audit Instrument and Protocol

The development of the environmental audit instrument and the data collection procedures have been previously described [39, 42]. Briefly, the PIN3 neighborhood audit instrument included 43 measures individually rated to reflect 5 domains: residential land use; nonresidential land use; public, residential and nonresidential space/aesthetics; mobility (walking and bicycling) amenities; and transit and road characteristics. These measures reflect aspects of the physical and social neighborhood environment that are hypothesized to influence physical activity behaviors.

To select the street segments for environmental auditing, PIN3 study participants' home locations were identified with the global positioning system (GPS) or address geocoding [42]. A customized road network was created from the county-level road centerline datasets for the study area; this customized road network was determined to be more accurate and complete as compared to four publicly and commercially available road datasets [42]. Using this customized road network, all roads within one-quarter mile of the PIN3 participants' home locations, identified as the participant's home neighborhood, were selected for auditing [39]. A road segment was defined by the length between two intersections, between an intersection and a cul-de-sac, or an intersection and a dead-end road.

Each road segment was further identified as either urban or rural. Urbanicity was defined using the United States (US) Census Bureau's 2000 urban and rural classification, where urban is defined using the following criteria: (1) core census block groups had a population density of at least 1000 people per square mile, and (2)

the surrounding census block groups had a population density of at least 500 people per square mile; under certain conditions, the US Census Bureau defined less densely settled areas as urban [43]. In the PIN3 study, a segment was identified as urban if the midpoint fell within an urban block group, otherwise it was identified as rural [39]. Over the course of two summers in 2005 (year 1) and 2006 (year 2), 10,770 road segments were audited by trained raters that independently rated in teams of two simultaneously from within a motor vehicle.

Inter-Rater Reliability and Stability Road Sampling Process

In year 1, a 5% random sample of the audited road segments were selected for reliability testing ($n = 481$), with oversampling in the rural areas. Each road segment was rated twice by teams of two, with the second rating of a road segment occurring approximately 2 weeks after the first rating. The teams differed from time 1 to time 2, such that no person ever rated the same road segment twice.

In year 2, a sub-sample of the year 1 sampled roads ($n = 253$) were re-audited by two teams, each comprised of two raters. Each day, team 1 would rate a portion of the selected road segments in the morning, and then rate another portion of the selected road segments in the afternoon. Team 2 would rate the same segments on the same day, but at alternate times. This was repeated day by day until all road segments in the sub-sample were rated.

Statistical Analysis

The original 43 measures were subdivided into 73 item scores for calculation of proportion of overall agreement (PO) and kappa statistics; see Supplemental Table 1 for a complete list of items assessed. For each of the items in the audit instrument, agreements between the (1) two-time ratings from year 1 (two-week agreement), (2) first time ratings from year 1 and year 2 (one-year agreement), and (3) two-time ratings from year 2 (same-day agreement) were evaluated and calculated using PO and kappa. PO was calculated to indicate the proportion of cases for which raters at time 1 and time 2 agreed. Kappa statistics were calculated to verify whether agreement exceeded chance levels. Kappa values less than 0.2 were considered poor agreement, 0.2 to < 0.4 fair, 0.4 to < 0.6 moderate, 0.6 to < 0.8 substantial, and 0.8 to

1.0 outstanding [44]. Simple kappa statistics were calculated for ratings of binary or nominal items, and weighted kappa statistics were calculated for ratings of ordinal items.

To test agreement by urbanicity, log-linear models were used to explore whether any significant exact agreement and/or linear-by-linear association existed [45–47]. The exact agreement was applied on both ordinal and nominal ratings to measure whether ratings agree exactly, while the linear-by-linear association was used on ordinal ratings to measure ratings that did not agree exactly but tended to be high together or low together. The Pearson scaling adjustment was applied to all log-linear models due to overdispersion. Log-likelihood-ratio tests were performed on log-linear models to test whether the urbanicity of road segments significantly affected two-week exact agreement and linear-by-linear association. For all analyses, an alpha level of 0.05 was used to indicate statistical significance. Statistical analyses were carried out using SAS Version 9.3 (SAS Institute Inc. Cary, NC.).

Results

Among the road segments that were selected and rated during the first auditing in year 1, 481 were revisited and rated a second time by a different two-person team in an approximately two-week range (median 16 days, inter-quartile range 10–20 days). Of these, 16 segments were excluded because of changes in road the network between time points so that segments were not structurally comparable, and one was excluded because it was rated by the same rater at each time point. This left 464 segments available for analysis on year 1. An additional 211 road segments were further excluded because they were not rated in year 2, leaving 253 road segments available for year 1 to year 2 comparisons.

Kappa values, color coded based on level of agreement, for same-day reliability, two-week reliability, and one-year stability are reported in Table 1. Additionally, ratings at time 1, inter-rater reliability (IRR) size, PO, and Kappa (95% CI) values for same-day reliability, two-week reliability, and one-year stability are reported in Supplemental Table 1. For same-day reliability (year 2, $n = 253$), PO ranged from 0.60 (children/youth visible and physically active) to 1.00 (housing authority/US Department of Housing and Urban Development projects, religious structures, bus facilities, highest speed

limit sign on segment, flashing warning signs, pavement marking/crosswalks, share the road bicycle sign, speed bumps, and curb extensions). Considering the kappa statistics, 37% ($n = 26$) had outstanding agreement, 36% ($n = 25$) had substantial agreement, 11% ($n = 8$) had moderate agreement, 10% ($n = 7$) had fair agreement, and 6% ($n = 4$) had poor agreement.

For two-week reliability (year 1, $n = 464$), PO ranged from 0.53 (overall condition of vacant/underdeveloped land) to 1.00 (housing authority/HUD projects, religious structures, bus facilities, flashing warning signs, share the road bicycle signs, curb extensions). Considering the kappa statistics, 30% ($n = 21$) had outstanding agreement, 17% ($n = 12$) had substantial agreement, 32% ($n = 23$) had moderate agreement, 7% ($n = 5$) had fair agreement, and 14% ($n = 10$) had poor agreement.

Results of the two-week agreement comparison (year 1, $n = 464$) of urban and rural areas is reported in Supplemental Table 2. Out of the 61 tested items, 8% ($n = 5$) showed significant differences in strength of agreement when comparing urban and rural locations. Perceived walkability of streets, presence of visible security warning signs, children/youth visible, and presence of a median/traffic island showed significantly higher agreement in rural areas as compared to urban areas. Overall condition of resident-kept grounds showed significantly lower agreement in rural areas as compared to urban areas.

For one-year stability (year 1 compared to year 2, $n = 253$), PO ranged from 0.56 (walkable street) to 1.00 (highest speed limit sign on segment, flashing warning signs, share the road bicycle sign, and curb extensions). Considering the kappa statistics, 26% ($n = 18$) had outstanding agreement, 17% ($n = 12$) had substantial agreement, 29% ($n = 20$) had moderate agreement, 16% ($n = 11$) had fair agreement, and 13% ($n = 9$) had poor agreement.

Discussion

This study examined the one-day and two-week reliability and one-year stability of the PIN3 neighborhood audit instrument. For a majority of items, overall percent agreement was outstanding across the time frames examined. Reliability was the strongest with repeated testing within 1 day, with 81% having moderate to outstanding kappa statistics. Two-week reliability was slightly lower, with 77% having moderate to

Table 1 Test-retest reliability and stability of PIN3 audit ratings, presented by level of agreement

Q # ^a	Same-day Reliability	Two-week Reliability	One-year Stability
	Kappa	Kappa	Kappa
Residential Land Use[§]			
1 Walkable street (subjective assessment)			
2 Residential units			
3 Types of residential housing			
Single family-detached			
Multi-family/apartment/duplex			
Mobile homes/trailer home			
Housing authority/HUD ^b projects			
New construction/renovation			
4 Overall condition of most residential units ^{§#}			
5 Overall condition of residential-kept grounds [§]			
6 Traditional lawn or landscaped front yards? [§]			
7 Porches [§]			
8 Some form of decoration [§]			
9 Border (fences/shrubs) [§]			
10 Visible security warning signs [§]			
Any burned/boarded up/abandoned residential units [§]			
11			
Nonresidential Land Use			
12 Commercial land use			
13 Industrial land use	-		
14 Agricultural land			
15 Religious structures			-
16 Overall condition of most buildings			
Any burned/boarded up/ abandoned nonresidential units			
17			
18 Home-based businesses			
Public, Residential, and Nonresidential Space / Aesthetics			
19 Vacant/underdeveloped land			
Overall condition of vacant/underdeveloped land			
20			
21 General condition of public spaces ^{&}			
22 Adults visible			
22 Children/youth visible			
23 Are adults being physically active			
23 Are children/youth being physically active			
22&23 Adults visible and physically active			
22&23 Children/youth visible and physically active			
Any public or neighborhood park or playground [§]			
24			
25 Overall condition of park and/or playground			
26 Visible dogs			
27 Amount of litter			
28 Type of litter			
Nonalcoholic cans/bottles/paper			
Alcoholic cans/bottles			
Large items			
Other litter			
29 Graffiti			
Walking and Bicycling Amenities			
30 Presence of sidewalk			

Table 1 (continued)

31	Sidewalk buffer			
32	Sidewalk condition			
30&32	Sidewalk condition			
33	Footpath along road			
34	Any trails that you can see in this segment			
35	Trees shading walking area			
36	Road oriented public lighting			
36	Pedestrian oriented public lighting			
Transit and Road Characteristics				
37	Bus facilities			
38	Number of lanes			
39	Road paved			
40	Highest speed limit sign on segment			
41	Shoulder or bike lane			
42	On-street parking			
43	Traffic control devices, crossing aids, and signs in segment			
_1	Traffic light(s)			
_2	Flashing warning sign(s)	-	-	-
_3	Stop sign(s)			
_4	Pavement marking/ crosswalk(s)			
_5	Yield to pedestrian paddles/ signal/crossing street sign(s)			
_6	Share the road bicycle sign	-		
_7	Other pedestrian or bike friendly traffic signs			
_8	Bicycle parking facilities			
_9	Speed bumps			
_10	Median / traffic island			
_11	Curb extension(s)	-		
_12	Neighborhood entrance signs			
_13	Neighborhood crime watch			
_14	No trespassing(s)			
_15	Beware of dog / invisible fence			
_16	Billboard			-
43 [†]	Control devices oriented for cars			
43 [€]	Signs for cars about bike/ pedestrian			

^a Complete categorization of these items is available in Supplemental Table 1. Numbering system is from Evenson et al. (2009) additional file #1[39]

^b HUD: The U.S. Department of Housing and Urban Development

#Segments could not be determined due to woods or other reasons, so they were set to missing

&Segments rated as "not applicable (private road)" were set to missing

§Non residential segments were left as missing

†Derived as a count for presence of traffic lights, stop signs, speed bumps, median/traffic islands, and curb extensions (#43_1, 43_3, 43_9, 43_10, and 43_11)

€Presence of either a flashing warning sign, "share the road" bicycle sign or other pedestrian or bike friendly traffic signs (#43_2, 43_6, and 43_7)

-Kappa not reported due to extremely small cell sizes

Outstanding agreement (kappa 0.8 to 1.0)

Substantial agreement (kappa 0.6 to <0.8)

Moderate agreement (kappa 0.4 to <0.6)

Fair agreement (kappa 0.2 to <0.4)

Poor agreement (kappa <0.2)

outstanding kappa statistics. This is comparable to the reliability of other audit tools, such as the Microscale Audit of Pedestrian Streetscapes (MAPS), which reported 76% of items having moderate to outstanding kappa statistics, i.e., greater than or equal to 0.40 [12]. One-year stability had more variable findings, with 68% showing moderate to outstanding kappa statistics. Finally, when urban and rural locations were compared, the reliability of the audit measures were largely consistent, with only 8% of items exhibiting significant differences by urbanicity. Certain aspects of the environment are expected to change over time, and these data provide important information to researchers to improve understanding of how neighborhood environmental audit tools are best utilized.

When comparing specific items, the level of agreement that was exhibited for same-day reliability was similar or slightly lower when examined for two-week reliability. Items assessing structural elements, such as presence/types of residential housing, lighting, presence of sidewalks, and traffic control devices showed fairly consistent substantial to outstanding reliability over the two time frames. Items that showed poor reliability at both time frames were primarily related to the social environmental features that are transient, such as visible people or dogs. The lack of reliability for the social environment found here is consistent with the findings from reliability testing of other audits [10, 15]. The few items that had a substantial shift in reliability between time points were related to types of litter, which showed moderate reliability within 1 day, but generally fair to poor over the two-week interval. It is reasonable to expect that the presence of litter would change in a two-week period, indicating that they may not be a reliable or stable estimate of neighborhood environments over time. Similarly, the lower reliability estimates for the social environment may indicate a lack of stability in the environment over a two-week time frame, rather than a lack of reliability between raters.

For one-year stability, individual item kappa scores were fairly consistent with two-week reliability. One notable difference was for the item “type of residential housing: new construction/renovation”, which had a kappa of 0.60 for two-week reliability, but was 0.06 at 1 year. This is understandable, as a construction site for residential housing would not likely last a year. Other types of residential housing had substantial to outstanding agreement over the one-year time frame, along with other structural aspects of the sidewalks, land use, and

many of the road characteristics. Visible adults or dogs showed poor agreement at both two-week and one-year audits, and types of litter showed low agreement over all time-frames—indicating again the changeable nature of certain elements of an environment over time.

In this study, it was shown that the PIN3 instrument had consistent levels reliability in both urban and rural locations for a large majority of items assessed. Interestingly, four of the five items that were different by urbanicity showed greater reliability in rural areas. This could be due to a lower variability in the rural areas which allows for raters to more easily identify specific neighborhood features. Further, of the five items that showed differences when examined by urbanicity, one was presence of visible children, which was problematic for the other reliability metrics assessed in this study. As previously stated, two other measurement tools have been used in both urban and rural locations: WASABE [13] and the Checklist [10]. However, neither examined the reliability of the estimates by urbanicity.

Overall, these findings indicate that the PIN3 instrument is a reliable audit tool for studies assessing neighborhood locations in both urban and rural environments and over time. The use of this tool in research practice would in large part be dictated by the study question. Projects that assess the built environment of a neighborhood can use the PIN3 neighborhood environmental audit instrument with the knowledge that the built environment factors identified in one auditing session will in large part be stable up to a year after the auditing took place. Conversely, studies that measure social aspects of the environment should do so with caution—the presence of individuals on the street, as well as the behaviors they engage in, are largely variable even within 1 day. From this perspective, it would be important to consider the day of week and time of day that audits are conducted, the weather, and repeating audits within 1 day to fully capture the social aspects of a neighborhood environment.

Strengths and Limitations

This study is the first to examine the stability of neighborhood physical environmental audit items over a one-year time frame. A strength of this study is the breadth of reliability and stability estimates calculated, which gives greater understanding of how these measures might be best used in practice. In addition, the comparison of these measures by urbanicity indicates the value of the

PIN3 audit for assessing both urban and rural neighborhood environments.

There are a few limitations of this study. With one-year stability, it is possible that some of the road segments were audited by the same person in both years, but it is unlikely that they would recall the roads from 1 year to the next. For some segments, the raters had difficulty understanding where to begin and end; this may have adversely impacted reliability. In addition, ratings could vary due to weather conditions at the time of the audit. For certain items, the diversity of the neighborhood environments was limited so that an option within an item was not noted (e.g., segment had visible billboards); in these instances, kappa was calculated as zero.

Conclusions

This study was the first to examine the reliability and stability of a neighborhood physical environmental audit tool in both urban and rural areas. The PIN3 neighborhood audit tool showed high reliability and stability for many of the items measured, and was largely consistent in both across neighborhood urbanicity. Items related to structural aspects of a neighborhood were largely reliable and stable over the time frames studied, whereas social aspects were less so. Other neighborhood audit tools should consider examining in more detail the reliability and the stability of their measures, to provide better understanding of how neighborhood environments should be assessed. In practice, projects aiming to assess the neighborhood environment should consider which characteristics to assess, whether it be the social or built environment; this consideration will dictate the number of audits that may be required.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding Funding for this study was provided by the National Institutes of Health (NIH)/National Cancer Institute (#CA109804-01). AKP was supported by a National Research Service Award T32 post-doctoral research fellowship, funded by the NIH, National Heart, Lung, and Blood Institute (#T32HL007055).

Compliance with ethical standards This study was reviewed and approved by the Institutional Review Board at UNC.

References

1. Sallis JF, Owen N, Fisher EB. Ecological models of health behavior. In: Glanz K, Rimer BK, Viswanath K, editors. *Health behavior and health education: theory, research, and practice*. 4th ed. San Francisco, CA: Jossey-Bass; 2008. p. 465–86.
2. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Annu Rev Public Health*. 2006;27:297–322.
3. Sallis JF. Measuring physical activity environments: a brief history. *Am J Prev Med*. 2009;36(4 Suppl):S86–92.
4. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am J Prev Med*. 2009;36(4 Suppl):S99–123.e112.
5. Bethlehem JR, Mackenbach JD, Ben-Rebah M, Compennolle S, Glonti K, Bárdos H, 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.
6. 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):1224–9.
7. Bader MDM, Mooney SJ, Lee YJ, Sheehan D, Neckerman KM, Rundle AG, et al. Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health Place*. 2015;31:163–72.
8. Mooney SJ, Bader MDM, 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):626–35.
9. Rundle AG, Bader MD, Richards CA, Neckerman KM, Teitler JO. Using Google Street View to audit neighborhood environments. *Am J Prev Med*. 2010;40(1):94–100.
10. Brownson RC, Hoehner CM, Brennan LK, Cook RA, Elliott MB, McMullen KM. Reliability of 2 instruments for auditing the environment for physical activity. *J Phys Act Health*. 2004;1(3):191–208.
11. Nickelson J, Wang AR, Mitchell QP, Hendricks K, Paschal A. Inventory of the physical environment domains and subdomains measured by neighborhood audit tools: a systematic literature review. *J Environ Psychol*. 2013;36:179–89.
12. Millstein RA, Cain KL, Sallis JF, Conway TL, Geremia C, Frank LD, et al. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). *BMC Public Health*. 2013;13:403.
13. Malecki KC, Engelman CD, Peppard PE, Nieto FJ, Grabow ML, Bernardinello M, et al. The Wisconsin Assessment of the Social and Built Environment (WASABE): a multi-dimensional objective audit instrument for examining neighborhood effects on health. *BMC Public Health*. 2014;14.
14. Su M, Du YK, Liu QM, et al. Objective assessment of urban built environment related to physical activity—development, reliability and validity of the China Urban Built Environment Scan Tool (CUBEST). *BMC Public Health*. 2014;14:109.

15. Scanlin K, Haardoefer R, Kegler MC, Glanz K. Development of a pedestrian audit tool to assess rural neighborhood walkability. *J Phys Act Health*. 2014;11(6):1085–96.
16. Leung CW, Gregorich SE, Laraia BA, Kushi LH, Yen IH. Measuring the neighborhood environment: associations with young girls' energy intake and expenditure in a cross-sectional study. *Int J Behav Nutr Phys Act*. 2010;7:52.
17. Parsons JA, Singh G, Scott AN, Nisenbaum R, Balasubramaniam P, Jabbar A, et al. Standardized observation of neighbourhood disorder: does it work in Canada? *Int J Health Geogr*. 2010;9:6.
18. Schaefer-McDaniel N. Neighborhood stressors, perceived neighborhood quality, and child mental health in New York City. *Health Place*. 2009;15(1):148–55.
19. Michael YL, Keast EM, Chaudhury H, Day K, Mahmood A, Sarte AF. Revising the senior walking environmental assessment tool. *Prev Med*. 2009;48(3):247–9.
20. Franzini L, Elliott MN, Cuccaro P, Schuster M, Gilliland MJ, Grunbaum JA, et al. Influences of physical and social neighborhood environments on children's physical activity and obesity. *Am J Public Health*. 2009;99(2):271–8.
21. Furr-Holden CD, Smart MJ, Pokorni JL, et al. The NifETy method for environmental assessment of neighborhood-level indicators of violence, alcohol, and other drug exposure. *Prev Sci*. 2008;9(4):245–55.
22. Zhu X, Lee C. Walkability and safety around elementary schools economic and ethnic disparities. *Am J Prev Med*. 2008;34(4):282–90.
23. King D. Neighborhood and individual factors in activity in older adults: results from the neighborhood and senior health study. *J Aging Phys Act*. 2008;16(2):144–70.
24. Kamphuis CB, Giskes K, Kavanagh AM, et al. Area variation in recreational cycling in Melbourne: a compositional or contextual effect? *J Epidemiol Community Health*. 2008;62(10):890–8.
25. Foltete J-C, Piombini A. Urban layout, landscape features and pedestrian usage. *Landsc Urban Plan*. 2007;81:225–34.
26. Zenk SN, Schulz AJ, Mentz G, House JS, Gravlee CC, Miranda PY, et al. Inter-rater and test-retest reliability: methods and results for the neighborhood observational checklist. *Health Place*. 2007;13(2):452–65.
27. Suminski RR, Heinrich KM, Poston WS, Hyder M, Pyle S. Characteristics of urban sidewalks/streets and objectively measured physical activity. *J Urban Health*. 2008;85(2):178–90.
28. Clifton KJ, Livi Smith AD, Rodriguez D. The development and testing of an audit for the pedestrian environment. *Landsc Urban Plan*. 2007;80(1–2):95–110.
29. Hoehner CM, Ivy A, Ramirez LK, Handy S, Brownson RC. Active neighborhood checklist: a user-friendly and reliable tool for assessing activity friendliness. *Am J Health Promot*. 2007;21(6):534–7.
30. Wright PA, Kloos B. Housing environment and mental health outcomes: A levels of analysis perspective. *J Environ Psychol*. 2007;27(1):79–89.
31. McMillan TE. The relative influence of urban form on a child's travel mode to school. *Transp Res A Policy Pract*. 2007;41(1):69–79.
32. Boarnet MG, Day K, Alfonzo M, Forsyth A, Oakes M. The Irvine-Minnesota inventory to measure built environments: reliability tests. *Am J Prev Med*. 2006;30(2):153–9.
33. Andresen EM, Malmstrom TK, Miller DK, Wolinsky FD. Reliability and validity of observer ratings of neighborhoods. *J Aging Health*. 2006;18(1):28–36.
34. Laraia BA, Messer L, Kaufman JS, Dole N, Caughy M, O'Campo P, et al. Direct observation of neighborhood attributes in an urban area of the US south: characterizing the social context of pregnancy. *Int J Health Geogr*. 2006;5:11.
35. Craig CL, Brownson RC, Cragg SE, Dunn AL. Exploring the effect of the environment on physical activity: a study examining walking to work. *Am J Prev Med*. 2002;23(2 Suppl):36–43.
36. Pikora TJ, Bull FC, Jamrozik K, Knuiman M, Giles-Corti B, Donovan RJ. Developing a reliable audit instrument to measure the physical environment for physical activity. *Am J Prev Med*. 2002;23(3):187–94.
37. Caughy MO, O'Campo PJ, Patterson J. A brief observational measure for urban neighborhoods. *Health Place*. 2001;7(3):225–36.
38. Loukaitou-Sideris A, Liggett R, Iseki H, Thurlow W. Measuring the effects of built environment on bus stop crime. *Environ Plan B: Plan Design*. 2001;28(2):255–80.
39. Evenson KR, Sotres-Alvarez D, Herring AH, Messer L, Laraia BA, Rodríguez DA. Assessing urban and rural neighborhood characteristics using audit and GIS data: derivation and reliability of constructs. *Int J Behav Nutr Phys Act*. 2009;6:44.
40. Knapp EA, Nau C, Brandau S, DeWalle J, Hirsch AG, Bailey-Davis L, et al. Community Audit of Social, Civil, and Activity Domains in Diverse Environments (CASCADDE). *Am J Prev Med*. 2017;52(4):530–40.
41. Brown AF, Morris DM, Kahn KL, et al. The healthy community neighborhood initiative: rationale and design. *Ethn Dis*. 2016;26(1):123–32.
42. Frizzelle BG, Evenson KR, Rodriguez DA, Laraia BA. The importance of accurate road data for spatial applications in public health: customizing a road network. *Int J Health Geogr*. 2009;8:24.
43. Census 2000 urban and rural classifications. <https://www.census.gov/geo/reference/ua/urban-rural-2000.html>. Accessed Sept 26, 2017.
44. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–74.
45. Tanner MA, Young MA. Modeling agreement among raters. *J Am Stat Assoc*. 1985;80(389):175–80.
46. Tanner MA, Young MA. Modeling ordinal scale disagreement. *Psychol Bull*. 1985;98(2):408–15.
47. Agresti A. *An introduction to categorical data analysis*. New York, NY: Wiley; 1996.