BRIEF REPORT

Using Google Street View to Audit the Built Environment: Inter-rater Reliability Results

**Cheryl M. Kelly, PhD · Jeffrey S. Wilson, PhD ·
Elizabeth A. Baker, PhD, MPH ·
Douglas K. Miller, MD · Mario Schootman, PhD**

Abstract

Background Observational field audits are recommended for public health research to collect data on built environment characteristics. A reliable, standardized alternative to field audits that uses publicly available information could provide the ability to efficiently compare results across different study sites and time.

Purpose This study aimed to assess inter-rater reliability of built environment audits conducted using Google Street View imagery.

Methods In 2011, street segments from St. Louis and Indianapolis were geographically stratified to ensure representation of neighborhoods with different land use and socioeconomic characteristics in both cities. Inter-rater reliability was assessed using observed agreement and the prevalence-adjusted bias-adjusted kappa statistic (PABAK).

Results The mean PABAK for all items was 0.84. Ninety-five percent of the items had substantial (PABAK≥0.60) or nearly perfect (PABAK≥0.80) agreement.

Conclusions Using Google Street View imagery to audit the built environment is a reliable method for assessing characteristics of the built environment.

Keywords Physical activity · Measurement · Imagery

C. M. Kelly (✉)
Beth-el College of Nursing and Health Sciences, University of Colorado,
1420 Austin Bluffs Parkway,
Colorado Springs, CO 80904, USA
e-mail: ckelly6@uccs.edu

J. S. Wilson
Department of Geography, School of Liberal Arts, Indiana University–Purdue University Indianapolis,
425 University Blvd,
Indianapolis, IN 46202, USA

E. A. Baker
School of Public Health, Saint Louis University,
3545 Lafayette, Ste 300,
St. Louis, MO, USA

D. K. Miller
Regenstrief Institute, Inc., and Center for Aging Research, Indiana University,
420 E. 10th Street, Suite 2000,
Indianapolis, IN, USA

M. Schootman
School of Medicine, Washington University,
4444 Forest Park Avenue, Ste 6700,
St. Louis, MO, USA

Background

Advocates of physical activity promotion have recognized that interventions must address not only individual-level factors (e.g., lack of time or motivation) but also interpersonal (e.g., social support), community or environmental (e.g., improving sidewalks), and policy (e.g., land use planning) factors [1–4]. Public health researchers and practitioners recognize that interventions at the environmental or policy level provide opportunities, support, and cues to help people engage in physical activity and have the potential to benefit the population exposed to the environment, as potential complements to more individually focused interventions [4–6].

Observational field audits are one method used in public health research to collect data on built environment characteristics that affect health-related behaviors and outcomes, including physical activity [7]. However, field audits are time and resource intensive because they require auditors to travel to each location that must be observed. This limits practicality of implementing field audits across large or geographically dispersed areas (e.g., local, regional, national, or international study

sites); for example, in support of nationally representative studies.

New technologies using high-resolution omnidirectional imagery provide a visual record of built environment characteristics that may support more efficient and extendable alternatives to field-based methods. Omnidirectional camera systems, like those used for Google Street View (http://maps.google.com/help/maps/streetview/), collect imagery in multiple directions to create panoramic views. Image users can observe characteristics included on built environment audit instruments by virtually "driving" through a community. Google Street View is the most commonly accessible form of omnidirectional imagery, providing coverage in all major US cities, thousands of miles of roads in smaller towns and rural areas, and in international locations including much of Europe and selected areas in Australia, South America, Africa, and eastern Asia (http://gmaps-samples.googlecode.com/svn/trunk/streetview_landing/streetview-map.html).

Recent studies have shown accurate and consistent agreement between observational field audits and image-based interpretation using Google Street View [8–11]. In the largest of these studies, our group demonstrated that the average prevalence-adjusted, bias-adjusted kappa (PABAK) statistic of all items was 0.81 when comparing field audits to audits conducted using Google Street View, indicating substantial to nearly perfect agreement [8]. However, we are aware of no prior studies that have reported inter-rater reliability results for omnidirectional image-based audits. Inter-rater reliability is important to ensure consistency in measurements across different auditors.

The purpose of this study was to assess inter-rater reliability of built environment audits derived from interpretation of Google Street View imagery using The Active Neighborhood Checklist (i.e., the Checklist), an instrument that assesses the presence or absence of features and conditions of the built environment [12]. Assessing inter-rater reliability of image-based audits is important because this method offers potential to more efficiently implement audits across large or geographically dispersed areas. Imagery may be archived, allowing for the potential to assess temporal changes in built environment characteristics in support of longitudinal studies.

Methods

The study was conducted in suburban and urban areas in Indianapolis, Indiana, and St. Louis, Missouri. Street segments (i.e., both sides of a street between two intersections) were selected in both cities using a geographically stratified sampling design to ensure representation of neighborhoods with different land use and socioeconomic characteristics. GIS data were used to classify census block groups separately in both cities (St. Louis and Indianapolis) based on two poverty classes (≥ or < than 20 % population in poverty), two race classes (≥50 % African American or ≥50 % white population) and above or below the median percentage of commercial land use in the block groups. This stratification created eight categories of block groups; we randomly selected 50 street segments within each category (Table 1).

The Checklist includes 89 items across several sections and subsections (Table 2). The Checklist has six main sections assessing presence or absence of land use characteristics; public transportation; street characteristics; quality of the environment for pedestrians; sidewalks and related features; and shoulders and bike lanes. This tool has demonstrated strong inter-rater reliability when using observational field audits [12]. The Checklist is available online at http://activelivingresearch.org/node/12715.

Four undergraduate and graduate student research assistants participated in a 4-h training session (before conducting any audits) following the protocol of Hoehner and colleagues [12] that included conducting practice audits in the field and using Google Street View. Following this training, two auditors independently assessed the same street segments using Google Street View imagery, blinded to results of the other auditor. Auditors did not audit streets in their own city (i.e., St. Louis auditors viewed Indianapolis streets and vice versa).

Inter-rater reliability was assessed using Cohen's kappa (a measure of inter-rater agreement) and PABAK statistics. Importantly, unlike the traditional kappa coefficient, PABAK accounts for systematic differences between data sources and the distribution of each audit item (i.e., when variability is low) [13]. We followed the commonly used adjectival ratings of Landis and Koch to interpret the

Table 1 Sampling of streets (*n*=288)

	Commercial land use area above median		Commercial land use area below median	
	≥20 % poverty	<20 % poverty	≥20 % poverty	<20 % poverty
≥50 % African American	31 (11 %)	36 (13 %)	38 (13 %)	40 (14 %)
≥50 % White	34 (12 %)	39 (14 %)	32 (11 %)	38 (13 %)

Percentages are of the total number of segments (*n*=288)

S110

ann. behav. med. (2013) 45 (Suppl 1):S108–S112

Table 2 Google Street View inter-rater reliability for active neighborhood checklist items (*n*=75)

Audit tool section[b]	Agreement		Number of items in Landis and Koch value range[a]			
	Kappa mean (range)	PABAK mean (range)	<0.39, fair	0.40–0.59, moderate	0.60–0.79, substantial	0.80–1.00, nearly perfect
Land use (42/50)						
Types of land use (3/4)	0.67 (0.60, 0.74)	0.76 (0.66, 0.88)	0	0	2	1
Predominant uses (8/9)	0.40 (0.13, 0.69)	0.85 (0.74, 0.96)	0	0	2	6
Residential uses (7/8)	0.41 (0.00, 0.75)	0.89 (0.83, 0.99)	0	0	0	7
Parking (4/4)	0.32 (0.23, 0.47)	0.60 (0.51, 0.72)	0	2	2	0
Recreational (4/7)	0.40 (0.00, 0.66)	0.97 (0.94, 0.99)	0	0	0	4
Non-residential (16/18)	0.51 (0.00, 1.00)	0.93 (0.69, 1.00)	0	0	1	15
Public transportation (2/2)	0.52 (0.44, 0.59)	0.90 (0.83, 0.97)	0	0	0	2
Street Characteristic (10/10)	0.62 (0.24, 0.82)	0.91 (0.63, 0.99)	0	0	1	9
Quality of environment (6/9)	0.35 (0.19, 0.57)	0.73 (0.46, 0.97)	0	1	3	2
Sidewalk characteristics (8/11)						
Sidewalk present (1/2)	0.89	0.90	0	0	0	1
Sidewalk continuity (2/2)	0.82 (0.79, 0.86)	0.83 (0.79, 0.86)	0	0	1	1
Sidewalk width (2/2)	0.47 (0.18, 0.76)	0.70 (0.59, 0.81)	0	1	0	1
Curb cuts (1/1)	0.38	0.63	0	0	1	0
Buffer (2/2)	0.80 (0.75, 0.84)	0.82 (0.80, 0.84)	0	0	0	2
Alignment/obstruct (2/2)	0.19 (0.00, 0.39)	0.73 (0.62, 0.83)	0	0	1	1
Shoulder characteristics (5/7)						
Bike route or sign (1/1)	0.44	0.97	0	0	0	1
Shoulder present (1/3)	0.55	0.85	0	0	0	1
Shoulder width (1/1)	0.43	0.93	0	0	0	1
Shoulder continuity (1/1)	1.00	1.00	0	0	0	1
Shoulder obstruct (1/1)	1.00	1.00	0	0	0	1
Total (75/89)			0	4	14	57

[a] Landis and Koch value range assessed using prevalence-adjusted bias-adjusted kappa (PABAK)

[b] Numbers in parentheses are counts of items included from each section of the Checklist/total possible items

PABAK inter-rater agreement results: 0.80 to 1.00 (almost perfect agreement), 0.60 to 0.79 (substantial agreement), 0.40 to 0.59 (moderate agreement), 0.20 to 0.39 (fair agreement), and 0.00 to 0.19 (poor agreement) [14]. All Checklist items were dichotomized as present or absent to be consistent with how the Checklist was reported in previous studies [8, 12]. Items that required the auditor to indicate if something was present on one side of the street, both sides of the street, or not present were dichotomized as present (on one or both sides) vs. absent. Ordinal items (e.g., flat, moderate, or steep for slope) were characterized as present (e.g., moderate or steep) vs. absent (e.g., flat). A total of 75 checklist items were included in the analysis. Fourteen items were excluded due to lack of variability across street segments (less 1 % of streets had each of these items). For example, zero audited streets had an outdoor pool.

Results

Based on the availability of Google Street View imagery, we audited 288 of 400 street segments sampled (149 segments in Indianapolis and 139 segments in St. Louis). At the time of this study, 28 % of selected streets were not captured by Google Street View and were excluded. The number of audited segments in each of the eight stratification classes ranged from 31 to 40. The mean PABAK statistics for all items in the Checklist was 0.84 (Table 2). Table 2 summarizes average kappa and PABAK statistics by different sections and subsections of the Checklist (representing 75 items). Ninety-five percent (71) of the items had substantial or nearly perfect agreement. When comparing items in the land use subsections, PABAK values ranged from 0.60 to 0.97, with parking facilities being the least reliable items (ranging from 0.51 to 0.72) and recreational uses the most

reliable (ranging from 0.94 to 0.99). Mean reliability of items assessing access to public transit, street characteristics, and quality of environment ranged from 0.73 to 0.91, with tree shade and presence of litter being the least reliable among the 18 items in these sections (PABAK=0.46 and 0.66, respectively). Mean reliability of sidewalk characteristics ranged from 0.63 to 0.90 with curb cuts (PABAK=0.63), sidewalk width (PABAK=0.70), and alignments/obstructions (PABAK=0.73) being the least reliable subsections and presence of sidewalks the most reliable (PABAK=0.90). Mean reliability for the section shoulder characteristics ranged from 0.85 to 1.00, with presence of a shoulder the least reliable item (PABAK=0.85) and shoulder continuity and obstructions the most reliable (PABAK=1.00).

When inter-rater reliability was assessed across the different environmental stratifications (e.g., race, poverty, and land use), the pattern of results was essentially identical suggesting that Google Street View is reliable regardless of the racial composition, the poverty level, or the land use mix of the neighborhood.

Discussion

Previous studies suggest that omnidirectional imagery, as exemplified by Google Street View, offers a viable alternative to field audits that can improve efficiency and expand the geographic and temporal scope of and reduce resources required for conducting audits [8–11]. This study builds on previous work by demonstrating substantial inter-rater reliability for most items included on the Checklist when audits are conducted using Google Street View imagery. Results suggest Google Street View audits have inter-rater reliability comparable to observational field audits [12].

Characteristics that demonstrated the lowest reliability (e.g., on-street parking, tree shade on street, sidewalk width, and curb cuts) should be assessed with caution. These characteristics were harder to view on the imagery and could be blocked from view by cars parked along the side of a street. Additionally, consistent with field audit inter-rater reliability [12], items pertaining to environmental quality (e.g., amount of litter) are subject to perceptions and experiences of individual auditors and thus had lower inter-rater reliability.

Several limitations should be noted. First, while the imagery may be reliable, it is not available on all streets presently, although spatial coverage continues to increase over time. Second, we were unable to examine the potential effect of certain factors that may impact inter-rater reliability (e.g., training method, image clarity, viewing angles, and procedures used to view the imagery). Because all auditors were trained the same way and used the same image source and computer program, we could not investigate the potential effects of training method, viewing program, and image on

reliability. The auditors did view the images on different screens, which may have resulted in differences in clarity; however, the substantial to nearly perfect reliability results suggest that this factor did not play a significant role. Additionally, while raters were trained using a protocol, there were no quality control measures conducted throughout the auditing process; however, such quality control mechanisms would likely improve agreement even more.

Third, this study only assessed urban and suburban areas and results cannot be generalized to rural environments. Fourth, information regarding when the images were obtained is an issue with this technique. At the time of data collection, time stamps were not available on images, and it is possible that auditors in this study assessed imagery taken at different times. However, all audits were conducted over a short study period (3–4 months) and it is likely the images had not changed. After our study was completed, Google started including month and year of Street View image acquisition (available at the bottom-left of images when viewed through http://maps.google.com, but not in Google Earth). Date stamps allow researchers and practitioners to better match environmental conditions with temporally concurrent behavior and outcome measures. As research on built environment and physical activity evolves and potentially identifies other important characteristics of the built environment or improved measurement scales, researchers may be able to assess longitudinal changes in communities if multi-temporal and archived imagery are made available.

References

1. Brownson RC, Ballew P, Dieffenderfer B. Evidence-based interventions to promote physical activity: What contributes to dissemination by state health departments. *Am J Prev Med*. 2007; 33(1): S66-S73. quiz S74-68.

2. Brownson RC, Hagood L, Lovegreen SL, et al. A multilevel ecological approach to promoting walking in rural communities. *Prev Med*. 2005; 41(5–6): 837-842.

3. Brownson RC, Haire-Joshu D, Luke DA. Shaping the context of health: A review of environmental and policy approaches in the prevention of chronic diseases. *Annu Rev Publ Health*. 2006; 27: 341-370.

4. Brownson RC, Kelly C, Eyler A, et al. Environmental and policy approaches for promoting physical activity in the United States: A research agenda. *J Phys Act Heal*. 2008; 5: 488-503.

5. King AC, Jeffery RW, Fridinger F, et al. Environmental and policy approaches to cardiovascular disease prevention through physical activity: Issues and opportunities. *Heal Educ Q*. 1995; 22(4): 499-511.

6. Sallis J, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. *Am J Prev Med*. 1998; 15(4): 379-397.

7. Brownson RC, Hoehner C, Day K, et al. Measuring the built environment for physical activity: State of the science. *Am J Prev Med*. 2009; 36(4): S99-S123.

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

9. Rundle AG, Bader MDM, Richards CA, Neckerman KM, Teitler JO. Using Google Street View to audit neighborhood environments. *Am J Prev Med*. 2010; 40(1): 94-100.

10. Taylor BT, Fernando P, Bauman AE, Williamson A, Craig JC, Redman S. Measuring the quality of public open space using Google Earth. *Am J Prev Med*. 2010; 40(2): 105-112.

11. 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.

12. 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 Heal Promot*. 2007; 21(6): 534-537.

13. Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol*. 2000; 53: 499-503.

14. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33: 159-174.