



Published in final edited form as:

Landsc Urban Plan. 2017 November ; 167: 240–248. doi:10.1016/j.landurbplan.2017.06.014.

Reliability between online raters with varying familiarities of a region: Microscale Audit of Pedestrian Streetscapes (MAPS)

Wenfei Zhu¹, Yuliang Sun², Jonathan Kurka³, Carrie Geremia⁴, Jessa K. Engelberg⁵, Kelli Cain⁶, Terry Conway⁷, James F. Sallis⁸, Steven P. Hooker⁹, and Marc A. Adams¹⁰

¹School of Physical Education, Shaanxi Normal University, Xi'an, Shaanxi, China 710119

²School of Physical Education, Shaanxi Normal University, Xi'an, Shaanxi, China 710119

³School of Nutrition and Health Promotion, Arizona State University, Phoenix, AZ 85004

⁴Department of Family Medicine and Public Health, University of California, San Diego, CA 92103

⁵Department of Family Medicine and Public Health, University of California, San Diego, CA 92103

⁶Department of Family Medicine and Public Health, University of California, San Diego, CA 92103

⁷Department of Family Medicine and Public Health, University of California, San Diego, CA 92103

⁸Department of Family Medicine and Public Health, University of California, San Diego, CA 92103

⁹School of Nutrition and Health Promotion, Arizona State University, Phoenix, AZ 85004

¹⁰School of Nutrition and Health Promotion and Global Institute of Sustainability, Arizona State University, Phoenix, AZ 85004

Abstract

Background—To test inter-rater reliability of the online Microscale Audit of Pedestrian Streetscapes (MAPS) tool between raters with varying familiarities of Phoenix, Arizona.

Methods—The online MAPS tool, based on the MAPS in-field audit tool and scoring system, was used for audits. Sixty route pairs, 141 segment pairs, and 92 crossing pairs in Phoenix were included. Each route, segment or crossing was audited by two independent raters: one rater in Phoenix and the other in San Diego, California, respectively. Item, subscale scores, and total scores reliability analyses were computed using Kappa or intra-class correlation coefficient (ICC).

Results—The route overall score had substantial reliability (ICC: 0.832). Of the route subscale and overall scores, sixteen out of twenty had moderate to substantial reliability (ICC: 0.616–0.906), and the four subscales had fair reliability (ICC: 0.409–0.563). Sixteen out of twenty scores

Corresponding author: Marc A. Adams, Ph.D., M.P.H., School of Nutrition and Health Promotion and Global Institute of Sustainability, Arizona State University, Address: 425 N, 5th Street (MC9020), Phoenix, Arizona 85004, Phone: 602-827-2470, Fax: 602-827-2253, Marc.Adams@asu.edu.

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 citable 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.

Conflicts of Interest: None.

in segment and crossing sections demonstrated fair to substantial reliability (ICC: 0.448–0.897), and the remaining four had slight reliability (ICC: 0.348–0.364).

Conclusions—Most of the online MAPS items, subscales, and overall scores demonstrated fair to substantial reliability between raters with varied familiarities of the Phoenix area. Results support use of online MAPS to measure microscale elements of the built environment by raters unfamiliar with a region.

Keywords

Microscale Audit; Streetscapes; Pedestrian; Physical Activity

Introduction

Physical inactivity is one of the most important public health issues in the U.S. and internationally, due to its contribution to premature mortality and economic costs (Janssen, Carson, Lee, Katzmarzyk, & Blair, 2013; Jia & Lubetkin, 2014). A growing body of research indicates linkage between elements of the built environment and physical activity (Adams et al., 2012; Brownson, Hoehner, Day, Forsyth, & Sallis, 2009; Davison & Lawson, 2006; Rutt & Coleman, 2005; Sallis et al., 2009, 2015). Researchers have shown that macro-level features of the built environment, including regional land-use patterns, residential densities, and access to parks and public transportation, shape access to opportunities for physical activity (Li et al., 2008; Nagel, Carlson, Bosworth, & Michael, 2008; Troped, Wilson, Matthews, Cromley, & Melly, 2010). Diverse combinations of objectively-measured built environment features have been positively and consistently related to physical activity (Sallis et al., 2016) and walking behaviors (Adams et al., 2015; Kaczynski, 2010), and results appear robust across children (Kurka et al., 2015) and older adults (Adams et al., 2012; Kerr et al., 2014).

Elements of built environment for a region can be measured at the landscape or microscale level (e.g., sidewalk presence and qualities, street furniture, aesthetic, natural and cultural qualities of the built environment), using field or online direct observation or audits. Microscale audits of specific neighborhoods or routes are desired to capture details of a local context at a higher resolution and reflect people's experiences with the environment (Brownson et al., 2009). Numerous microscale audit tools have been developed to evaluate how built environment elements associate with residents' physical activity, and several have demonstrated good inter-rater reliability (Bethlehem et al., 2014; Clifton, Livi Smith, & Rodriguez, 2007; Millstein et al., 2013; Pikora et al., 2002). One validated tool for assessing detailed attributes of the built environment relevant to physical activity is the Microscale Audit of Pedestrian Streetscapes (MAPS) tool (Millstein et al., 2013). The items and subscales of MAPS have demonstrated moderate to substantial reliability, and the scoring represents a conceptual framework for microscale elements. MAPS has been used to examine associations of microscale attributes with physical activity, and findings show strong and positive associations for four age groups in three U.S. cities, even after accounting for macro-level features (Cain et al., 2014). Additional studies are needed to assess the reliability and validity of MAPS in different regions and cities. At present, the use

of MAPS is also limited by need for a field visit to directly observe and score the physical environment, which can be time intensive, expensive, and sometimes unsafe.

Web-based virtual mapping tools like Google Street View, which integrate photos in a geospatial framework, provide rich visual evidence of urban areas and can potentially reduce the burdens of in-field auditing. Testing the reliability of the virtual audit tool evaluates consistency in measurements across different raters with diverse backgrounds and knowledge of a region and offers potential to more efficiently implement audits across large or geographically dispersed areas (Brownson et al., 2009). A few recent studies (Ben-Joseph, Lee, Cromley, Laden, & Troped, 2013; Bethlehem et al., 2014; Griew et al., 2013; Kelly, Wilson, Baker, Miller, & Schootman, 2013) have shown acceptable reliability between in-field audits and online image-based audits for measuring microscale characteristics. Web-based virtual tools have proven to be good alternatives to field audits, with higher agreement for objectively verifiable elements (i.e., presence of infrastructure and equipment) and lower agreement for subjectively assessed items (i.e., aesthetics) (Charreire, 2014). Online auditing opens the possibility of observers auditing locations far from their actual locations, even places they have never physically visited. However, no studies could be found that examined inter-rater reliability between observers with varying familiarities of a region (living in vs. outside of a region).

The aim of the current study was to test inter-rater reliability of the online MAPS tool between independent raters from Phoenix, Arizona vs. San Diego, California with inherently different familiarities of the Phoenix metro region. We conducted the analysis in three levels, including the levels of individual MAPS items, subscales, and total scores (sum of positive and negative subscales) to evaluate reliabilities for different levels within the MAPS tool. We hypothesized that the online MAPS tool could be used reliably at all levels to measure microscale elements of the built environment by raters **with different familiarities** of the Phoenix metro area.

Methods

Sample

A total of 60 routes were selected and evaluated using MAPS in the Phoenix metro area, which is located in the southwestern United States, in the south-central portion of the U.S. state of Arizona. To ensure variability in neighborhood elements, all Census block groups from Maricopa County, Arizona were classified using a 2 by 2 matrix considering the macro-level factors of walkability and socioeconomic status (SES). Walkability was defined by a block group-level composite of GIS (geographic information systems)-measured net residential density, land use mix, and street connectivity. SES was defined using block group-level median household incomes. An equal number of routes were assigned for each cell in the walkability by SES matrix. Residential routes consisted of a pre-determined quarter mile route from an origin residential parcel toward a pre-selected non-residential destination (i.e., a cluster of commercial land uses) (Millstein et al., 2013). A quarter mile route was used to standardize the audit distance and limit observation time. Commercial routes consisted of a street segment in front of a pre-selected commercial cluster, defined by three or more commercial destinations, with the street bounded by two intersections. More

details about route selection and definitions have been published previously (Kurka et al., 2016).

Measures

The online version of the MAPS tool, henceforth called the online MAPS tool, was based on the MAPS in-field tool and developed for use with Google Street View. The in-field version of the MAPS tool was developed from prior measures to assess streetscapes for physical activity (Millstein et al., 2013). In the Millstein et al. study, the research team collected microscale environmental data in urban and suburban neighborhoods in Seattle/King County, Washington, San Diego County, California and five counties in the Baltimore, MD/Washington, DC region. Their in-field study included 290 routes, 516 segments, and 319 crossings (Millstein et al., 2013).

Based on the in-field version of MAPS, the purpose for the development of the current online MAPS tool was to take advantage of growing source of online street view data in the U.S. and internationally. Paralleling the four sections of the original MAPS tool (Millstein et al., 2013), the online MAPS tool consisted of: a) an overall route, b) street segments, c) crossings, and d) cul-de-sacs. Route-level variables summarized characteristics for the whole route, including items related to land use and destinations, transit stops, street amenities, traffic calming, aesthetics, and the social environment. Street segment-level variables were collected on every segment on the route and consisted of sidewalks, pedestrian buffers, sidewalk slope, bicycle infrastructure, sidewalk visibility from buildings, street trees, shade, and building aesthetics, setbacks and overall height. Street crossing variables were measured at every intersection or crossing on the route, and included crosswalks, slopes, width of crossings, crossing signals, and pedestrian protection. Cul-de-sac variables were assessed only when one or more cul-de-sacs were present within 400 feet of the participant's home. The cul-de-sacs section assessed the potential recreational environment within a cul-de-sac and included items about the size and condition of the surface area, slope, surveillance from surrounding homes, and amenities. The number of segments, crossings and cul-de-sacs varied by route.

A previously developed conceptual system for scoring the MAPS in-field audit tool was also applied to group items into subscales (Millstein et al., 2013). The scoring system was guided by a combination of factors thought to influence physical activity: safety, aesthetics, destinations, land use, recreational facilities, transportation, etc. The subscale scores were computed by summing those related items' scores. The subscales were then sorted by their expected positive or negative effects on physical activity to create the valence scores. Finally, an overall section score was calculated for each of the main sections.

Google Earth is a free geographic software program which views satellite images in excellent resolution, depicting anywhere on the face of the earth. It displays ground-level views of streets and buildings via car-mounted 360° cameras (Google Street View), as well as satellite images allowing a perpendicular or oblique angle view of streets, buildings, and landscapes (Google Aerial View). In this study, Google Street View was the main tool used for measuring microscale features.

The assessments were conducted by traveling the assigned route while scanning the forward-looking arc of 180° approximately every 100 feet and recording features and details along the designated route. Google Aerial View was used only when the characteristics were harder to view in the images from Google Street View or blocked by obstructions along the street, such as the number of trees or the building setback from the sidewalk. Raters were required to use the most recent layer of information on Google Earth and record the date of the images during the audit. Raters recorded the date of image acquisition, confirming most images were taken within two years of this study. All virtual audits were conducted over a 3-month period to limit other confounding variables.

Research teams in Phoenix and San Diego with different familiarities of Phoenix's metro area audited the same routes using the online MAPS tool. San Diego is a major city in California, United States, which is on the coast of the Pacific Ocean in Southern California. Differences between Phoenix and San Diego include the climate, built environment, landscaping, local culture, and population densities. A single expert rater trained raters in each city on the online MAPS tool over several days with practice sessions to ensure a high quality. Six raters (three at each site) were trained and certified using a standard certification process (Millstein et al., 2013). Each rater spent 15 hours training on the tool, and before rating the final routes, he/she was tested on at least four training routes (2 residential, 2 commercial). Raters completed the training process once they achieved an inter-rater reliability of at least 95% agreement with the expert trainer. Additional feedback and training routes were provided until the rater achieved the desired level of performance. Once certified by the expert rater, raters in each city were assigned routes at random. Each route in Phoenix was audited by two independent online raters: one from Phoenix and one from San Diego.

Statistical analyses

Most items from the online MAPS tool were coded dichotomously (no/yes) and scored as 0/1. Frequency items (0, 1, 2+) were scored as 0, 1, 2, and continuous and descriptive items were categorized by their distributions, theoretical relevance, and in compatibility with other scale items' scoring (Millstein et al., 2013). Inter-rater item reliability analyses were computed using percent agreement, Kappa (for dichotomous variables), and intra-class correlation coefficient (ICC, for continuous variables). Subscale scores and total scores were analyzed for inter-rater reliability using ICC. One-way random effects models were used for ICC calculation and single measures values reported.

Subscale, valence (positive/negative) and overall scores were calculated from the summaries of item responses (Millstein et al., 2013). Common cut-off values (Virtually none: Kappa = 0.00–0.10; slight: Kappa = 0.11–0.40; fair: Kappa = 0.41–0.60; moderate: Kappa = 0.61–0.80; substantial: Kappa = 0.81–1.00) were used (Shrout, 1998). ICC values for agreement were classified using the same criteria for Kappa values (Shrout, 1998). All data were analyzed using SPSS version 22.0 (SPSS, Inc., Chicago, IL).

Results

In this analysis, 60 routes with 141 segments and 92 crossings were examined. The cul-de-sac section was not included in this analysis due to its small sample size ($n=8$), and the uncertainty of its relation with physical activity. Table 1–3 provide subscale, valence score, subsection score, and overall score components, descriptive statistics, sample items, and reliability statistics in the sections of routes, segments and crossings. For comparison, these tables also present reliability statistics (Kappas or one-way random effects single measure ICCs) between in-field raters from three cities (all raters familiar with their environments) from the original study of the MAPS tool (Millstein et al., 2013).

Routes

There were ten positive subscales and one negative subscale in the destination and land use route section (Table 1). The positive destinations and land use subscales had fair to substantial inter-rater reliability, with ICC (or Kappa) values ranging from 0.409 (Negative Aesthetics and Social Subscale) to 0.881 (shops). The overall positive destinations and land use valence score was a sum of all ten positive subscales and had substantial reliability (ICC: 0.880). The negative destinations and land use valence score consisted of adverse land uses and demonstrated moderate reliability (ICC: 0.794). The overall destinations and land use subsection score (positive valence score minus negative valence score) had substantial reliability (ICC: 0.849). The route streetscape items include a positive or negative valence score, and their subsection score. All of the streetscape valence and overall scores had moderate inter-rater reliability: positive (ICC: 0.616), negative (ICC: 0.685), and overall (ICC: 0.644). The route aesthetics and social subscale had the same structure as the route streetscape. The positive and negative aesthetics and social valence score had fair reliability (ICC: 0.409–0.485). The overall aesthetics and social subsection score had fair reliability (ICC: 0.563).

The overall route score was calculated from the sum of the three route subsections scores (destinations and land use, streetscape, and aesthetics and social), and had substantial reliability (ICC: 0.832). In sum, of the route valence, subsection, and overall scores, eleven out of twenty (55.0%) had substantial reliability, eight out of twenty (40.0%) has substantial reliability, and four (20.0%) had moderate reliability. Compared to the original MAPS study using the in-field audit tool, sixteen out of twenty (80.0%) scores (including subscale, valence score, and subsection scores) had similar reliability (in the same classification) when using the virtual tool (Table 1).

Segments

There were six positive segment subscales and three negative segment subscales (Table 2). Three out of six (50.0%) of the positive subscales had moderate to substantial reliability, and two (33.3%) had fair reliability. The sidewalk positive qualities demonstrated slight reliability (ICC: 0.360). The positive segment valence score (sum of the six positive subscales) had moderate reliability (ICC: 0.797), while the negative valence score (sum of the three negative subscales) had slight reliability (ICC: 0.364). The overall segment section (overall positive minus overall negative) score demonstrated moderate reliability (ICC:

0.0.733). Compared to the in-field MAPS tool, three out of twelve (25.0%) scores in the segment section (including subscale, valence score, and subsection score) had similar reliability when using the virtual tool (Table 2).

Crossings

There were three positive and two negative crossing subscales (Table 3). Two out of three (66.7%) of the positive subscales had moderate reliability (ICC: 0.657–0.790), and one (33.3%) had fair reliability (ICC: 0.577). The positive crossing valence score (sum of the three positive subscales) had moderate reliability (ICC: 0.703). One of the negative crossing subscales had moderate reliability (ICC: 0.800), while the other had slight reliability (ICC: 0.348). The negative valence score (sum of the two negative subscales) had fair reliability (ICC: 0.548). The overall crossing section score demonstrated moderate reliability (ICC: 0.771). Compared to the in-field MAPS tool, four out of eight (50.0%) scores in the crossing section (including subscale, valence score, and subsection score) had similar reliability when using the virtual tool (Table 3).

Discussion

The main finding was that virtual streetscape audits using the MAPS tool had fair to substantial reliability across raters with different familiarities of a region. The present study builds on previous work developing and testing items, subscales, and overall scores of the in-person MAPS tool. Present results further suggest that the online MAPS tool has comparable inter-rater reliability to the in-field MAPS tool for the majority of subscales and overall scores. This study is one of the first investigations focusing on the inter-rater reliability of a virtual audit tool between raters with different familiarities of an area, which is helpful for future studies aiming to measure microscale features of the built environment for energy-balance behaviors.

In the current study, the general reliability of the online MAPS tool between raters with different familiarities of the Phoenix metro area was substantial. Virtual audit tools have the potential to add to both the amount and scale of research on microscale features. They offer the capacity to assess the quality of large numbers of street segments, intersections, and cul-de-sacs comprising routes through which residents can walk to reach neighborhood destinations without placing individuals in the field, dramatically reducing the required time and costs (Badland, Opit, Witten, Kearns, & Mavoa, 2010; Charreire, 2014; Clarke, Ailshire, Melendez, Bader, & Morenoff, 2010; Millstein et al., 2013; Taylor et al., 2011). Another advantage of virtual audits is that they can show adjacent areas that are not physically assessable or partially hidden, such as private streets (Ben-Joseph et al., 2013). The online MAPS tool is also a safe alternative when auditing unsafe neighborhoods or attempting to measure streetscape features during inclement weather.

Phoenix has been called one of the least sustainable cities in the world due to its location, climate, infrastructure and resulting political and societal challenges. It has a subtropical desert climate and one of the hottest and longest summer seasons in the U.S., with over one hundred extremely hot days defined as high temperatures over 100 degrees Fahrenheit. Virtual observations were conducted during the summer months (May 2014 to July 2014), in

which typical exposures for infield audits would have been unsafe. Phoenix also has an extensive canal network providing adjacent walking and cycling paths. Thus, Phoenix provides a new region to existing studies testing the MAPS (Cain et al., 2014; Kurka et al., 2016; Millstein et al., 2013), suggesting the MAPS tool and scoring system can be recommended for wider use by researchers, policy makers, and practitioners, even if they are not familiar with the local environment.

While the validity of the online MAPS tool was not examined in the current study, the current reliability results compare well with audits using the in-field MAPS tool (Millstein et al., 2013). Further, findings (Kurka et al., 2016) from our team have shown the online MAPS tool was an acceptable alternative to evaluating land uses of routes in the field. It was reported that the Google Street View method was more accurate than the aerial view for individual land uses and performed equally in high and low socioeconomic neighborhoods (Kurka et al., 2016). Therefore, the online MAPS tool using Google Street View may hold particular potential for audits conducted across multiple sites, or over vast geographic areas or nations, providing researchers with a rapid, convenient, cost-efficient, safe, and reliable method of assessing the microscale features of the built environment for physical activity.

Our findings indicated inter-rater reliability was highest in the route section and lowest in the segment section. The reliability of the tool appears well suited for capturing elements in land-use environments and transportation features. However, a few subscales, including particularly aesthetics and physical disorders, and sidewalk qualities, exhibited lower (slight) reliability estimates which agrees with previous studies (Ben-Joseph et al., 2013; Bethlehem et al., 2014; Brownson et al., 2004; Charreire, 2014; Clarke et al., 2010). These characteristics should be assessed with caution using online methods. The familiarity of local areas may allow raters to be more sensitive to aesthetic features and physical disorders (Hoehner, Ivy, Brennan Ramirez, Meriwether, & Brownson, 2006). However, lower reliabilities in present analyses may have less to do with observers' familiarity with the region than limitations of online images. Subjective items, such as broken windows and graffiti/tagging, are less reliable but still valuable to understanding pedestrians' perceptions of the community or neighborhood. Additional efforts are needed to more reliably audit the qualitative features of built environments. Similar to previous studies (Ben-Joseph et al., 2013; Brownson et al., 2009; Charreire, 2014; Chudyk, Winters, Gorman, McKay, & Ashe, 2014), inter-rater reliability was found to vary more when assessing rare features, such as institutional and government services, and more detailed elements such as width of street segment, and width of crossing. Some characteristics, such as the number of trees or street lights, are harder to view with online imagery and could be blocked by obstructions along the street. Developing improved methods of assessing aesthetic and social disorder features is a challenge common to most online audit tools.

Several limitations of the present study should be noted. First, the sample size for the route section (N=60) was relatively small because of resource restrictions, and we did not calculate the Kappa or ICC values for some items due to the low prevalence of certain features in the neighborhoods. Lack of variation in the environment and existing features may also result in a low Kappa value despite a high percent agreement (Griew et al., 2013; Hoehner et al., 2006). The reason is that some audit items assess features of the environment

that are not expected to occur frequently in most communities, but the presence of these items may be important elements (i.e. large parks, theaters) contributing to neighborhood walkability and, potentially, to residents' physical activity. Increased variation may be achieved by including additional neighborhoods and regions with diverse development patterns, levels of urbanization, mixtures of land use, and SES of residents. Second, the perspective from Google Street View images was different from that of a rater in the field. Some details of the built environment are not captured well virtually or lost during picture compression. Image clarity differs depending on the weather conditions and lighting when the images were obtained. Third, information of when and where the images in Google Street View were obtained are inconsistent. Coverage is more complete in urban than rural areas, and the date of image collection needs to be considered. Fortunately, Google has started to increase the spatial coverage over time, and include the date of image acquisition in most regions, enabling researchers and practitioners to match environmental conditions. Fourth, we randomly assigned routes to each rater, and each route was evaluated by only two independent raters, with one from Phoenix and one from San Diego. Further investigation is needed to explore the intra-rater reliability of the online MAPS tool. Future studies are also recommended to investigate geographical differences and longitudinal changes in built environments with the online MAPS tool if imagery data are available at appropriate locations and time points. Finally, further analysis is necessary to investigate the association of built environment features measured by online MAPS with physical activity or other health outcomes in diverse populations.

Conclusions

The online MAPS items, subscales, and overall scores in route, segment and crossing sections demonstrated fair to substantial inter-rater reliability between raters with varying familiarities of the Phoenix metro area. Present results suggest that the online MAPS tool can be used reliably to measure microscale elements of the built environment by raters unfamiliar with a specific urban/suburban region or neighborhood.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge Justin Martinez for his assistance in data collection. This work is supported in part from grants by the Fundamental Research Funds for the Central Universities (GK201603128, GK201603129) and National Institutes of Health (R01HL109222).

References

- Adams MA, Sallis JF, Conway TL, Frank LD, Saelens BE, Kerr J, King AC. Neighborhood environment profiles for physical activity among older adults. *American Journal of Health Behavior*. 2012; 36(6):757–769. [PubMed: 23026035]
- Adams MA, Todd M, Kurka J, Conway TL, Cain KL, Frank LD, Sallis JF. Patterns of Walkability, Transit, and Recreation Environment for Physical Activity. *American Journal of Preventive Medicine*. 2015; 49(6):878–887. [PubMed: 26232902]

- Badland HM, Opat S, Witten K, Kearns RA, Mavoa S. Can virtual streetscape audits reliably replace physical streetscape audits? *Journal of Urban Health*. 2010; 87(6):1007–1016. [PubMed: 21104331]
- Ben-Joseph E, Lee JS, Cromley EK, Laden F, Troped PJ. Virtual and actual: relative accuracy of on-site and web-based instruments in auditing the environment for physical activity. *Health & Place*. 2013; 19:138–150. [PubMed: 23247423]
- Bethlehem JR, Mackenbach JD, Ben-Rebah M, Compernelle S, Glonti K, Bardos H, Lakerveld J. The SPOTLIGHT virtual audit tool: a valid and reliable tool to assess obesogenic characteristics of the built environment. *International Journal of Health Geographics*. 2014; 13(1):52. [PubMed: 25515179]
- Brownson RC, Hoehner CM, Brennan LK, Cook Ra, Elliott MB, McMullen KM. Reliability of Two Instruments for Auditing the Environment for Physical Activity. *Journal of Physical Activity and Health*. 2004; 1:189–207.
- Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the Built Environment for Physical Activity. State of the Science. *American Journal of Preventive Medicine*. 2009; 36(4 SUPPL):S99–S123.e12. [PubMed: 19285216]
- Cain KL, Millstein RA, Sallis JF, Conway TL, Gavand KA, Frank LD, King AC. Contribution of streetscape audits to explanation of physical activity in four age groups based on the Microscale Audit of Pedestrian Streetscapes (MAPS). *Social Science & Medicine*. 2014; 116:82–92. [PubMed: 24983701]
- Charreire HH. Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: A systematic review (the SPOTLIGHT project). *Health & Place*. 2014; 25:1–9. [PubMed: 24211730]
- Chudyk AM, Winters M, Gorman E, McKay HA, Ashe MC. Agreement between virtual and in-the-field environmental audits of assisted living sites. *Journal of Aging and Physical Activity*. 2014; 22(3):414–420. [PubMed: 24085399]
- 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. [PubMed: 20797897]
- Clifton KJ, Livi Smith AD, Rodriguez D. The development and testing of an audit for the pedestrian environment. *Landscape and Urban Planning*. 2007; 80(1–2):95–110.
- Davison KK, Lawson CT. Do attributes in the physical environment influence children’s physical activity? A review of the literature. *International Journal of Behavioral Nutrition and Physical Activity*. 2006; 3(1):19. [PubMed: 16872543]
- 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. *The International Journal of Behavioral Nutrition and Physical Activity*. 2013; 10:103. [PubMed: 23972205]
- Hoehner CM, Ivy A, Brennan Ramirez L, Meriwether B, Brownson RC. How Reliably Do Community Members Audit the Neighborhood Environment for Its Support of Physical Activity? Implications for Participatory Research. *Journal of Public Health Management and Practice*. 2006; 12(3):270–277. [PubMed: 16614563]
- Janssen I, Carson V, Lee I-M, Katzmarzyk PT, Blair SN. Years of life gained due to leisure-time physical activity in the US. *American Journal of Preventive Medicine*. 2013; 44(1):23–29. [PubMed: 23253646]
- Jia H, Lubetkin EI. Comparing Quality-Adjusted Life Expectancy at Different Levels of Physical Activity. *Journal of Physical Activity & Health*. 2014; 11(2):278–284. [PubMed: 23364410]
- Kaczynski AT. Neighborhood walkability perceptions: associations with amount of neighborhood-based physical activity by intensity and purpose. *Journal of Physical Activity & Health*. 2010; 7(1):3–10. [PubMed: 20231749]
- Kelly CM, Wilson JS, Baker EA, Miller DK, Schootman M. Using Google Street View to audit the built environment: Inter-rater reliability results. *Annals of Behavioral Medicine*. 2013; 45(SUPPL. 1):108–112.
- Kerr J, Norman G, Millstein R, Adams MA, Morgan C, Langer RD, Allison M. Neighborhood Environment and Physical Activity Among Older Women: Findings From the San Diego Cohort of

the Women's Health Initiative. *Journal of Physical Activity & Health*. 2014; 11(6):1070–1077. [PubMed: 23963780]

Kurka JM, Adams MA, Geremia C, Zhu W, Cain KL, Conway TL, Sallis JF. Comparison of field and online observations for measuring land uses using the Microscale Audit of Pedestrian Streetscapes (MAPS). *Journal of Transport & Health*. 2016; 3(3):278–286.

Kurka JM, Adams MA, Todd M, Colburn T, Sallis JF, Cain KL, Saelens BE. Patterns of neighborhood environment attributes in relation to children's physical activity. *Health & Place*. 2015; 34:164–170. [PubMed: 26057609]

Li F, Harmer PA, Cardinal BJ, Bosworth M, Acock A, Johnson-Shelton D, Moore JM. Built environment, adiposity, and physical activity in adults aged 50–75. *American Journal of Preventive Medicine*. 2008; 35(1):38–46. [PubMed: 18541175]

Millstein RA, Cain KL, Sallis JF, Conway TL, Geremia C, Frank LD, Kerr J. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). *BMC Public Health*. 2013; 13(1):403. [PubMed: 23621947]

Nagel CL, Carlson NE, Bosworth M, Michael YL. The relation between neighborhood built environment and walking activity among older adults. *American Journal of Epidemiology*. 2008; 168(4):461–468. [PubMed: 18567638]

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. *American Journal of Preventive Medicine*. 2002; 23(3):187–194. [PubMed: 12350451]

Rutt CD, Coleman KJ. Examining the relationships among built environment, physical activity, and body mass index in El Paso, TX. *Preventive Medicine*. 2005; 40(6):831–841. [PubMed: 15850885]

Sallis JF, Bowles HR, Bauman A, Ainsworth BE, Bull FC, Craig CL, Bergman P. Neighborhood environments and physical activity among adults in 11 countries. *American Journal of Preventive Medicine*. 2009; 36(6):484–490. [PubMed: 19460656]

Sallis JF, Cain KL, Conway TL, Gavand KA, Millstein RA, Geremia CM, King AC. Is Your Neighborhood Designed to Support Physical Activity? A Brief Streetscape Audit Tool. *Preventing Chronic Disease*. 2015; 12:E141. [PubMed: 26334713]

Sallis JF, Cerin E, Conway TL, Adams MA, Frank LD, Pratt MAIE. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. *The Lancet*. 2016; 387(10034):2207–2217.

Shrout PE. Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*. 1998; 7(3):301–317. [PubMed: 9803527]

Taylor BT, Fernando P, Bauman AE, Williamson A, Craig JC, Redman S. Measuring the quality of public open space using Google Earth. *American Journal of Preventive Medicine*. 2011; 40(2):105–112. [PubMed: 21238857]

Troped PJ, Wilson JS, Matthews CE, Cromley EK, Melly SJ. The built environment and location-based physical activity. *American Journal of Preventive Medicine*. 2010; 38(4):429–438. [PubMed: 20307812]

Research highlights

- Online MAPS is reliable for microscale measures by raters unfamiliar with a region.
- Online MAPS demonstrated fair to substantial reliability.
- It is possible to reliably audit distant locations without physical visit.
- Audit could be improved by employing the same raters with the same training.

Table 1
Agreement on route subscales between virtual raters with different familiarities of the environment (n=60)

Subscale	Label	Number of Items	Mean (SD)	Virtual Tool ^a		In-field Tool ^b	
				ICC (95% Confidence Interval)	Range of Item Kappas or ICCs (% Agreement)	ICC	Range of Item Kappas or ICCs (% Agreement)
Land use and destination							
Positive Subscale							
ResMix	Single family homes, apartment and condominiums, apartments above street retail	4	1.20 (0.65)	0.643 (0.437–0.785)	0.378–1.000 (87.0%–100.0%)	0.577	0.290–0.776 (84.5%–98.9%)
Commercial-Shops	Food-related land uses, retail and service-oriented land uses and shopping centers	10	2.48 (2.90)	0.881 (0.795–0.932)	0.196–0.796 (80.4%–100.0%)	0.873	0.407–0.842 (87.6%–98.6%)
Commercial-Restaurants/Entertainment	Food-related uses (fast food, sit-down, cafe), entertainment	4	1.59 (1.83)	0.877 (0.789–0.930)	0.407–0.877 (82.6%–91.3%)	0.842	0.765–0.796 (87.6%–98.3%)
Institutional service -professional service	Bank/Credit union, health-related professional, other services	3	1.87 (1.93)	0.829 (0.712–0.902)	0.503–0.628 (73.9%–78.3%)	0.849	0.743–0.808 (81.3%–93.8%)
Institutional service -Religious, Schools	Government or community land use, place of worship, school	2	0.15 (0.42)	0.630 (0.418–0.775)	0.543–1.000 (93.5%–100.0%)	0.717	0.712–0.722 (90.7%–94.1%)
Government Service	Health or social services, library/museums, post office, senior center	4	0.17 (0.26)	0.514 (0.268–0.698)	0.729–1.000 (95.7%–100.0%)	0.652	0.279–0.798 (94.8%–99.7%)
Parking Structures (Positive)	No parking facilities present, parallel/angled on-street parking	2	1.03 (1.00)	0.860 (0.761–0.920)	0.870–1.000 (93.5%–97.8%)	0.736	-0.011–0.689 (89.7%–96.9%)
Recreational land use-Public recreation facilities	Community garden, public indoor, public outdoor pay, public park	4	0.24 (0.46)	0.621 (0.408–0.770)	0.471–1.000 (82.6%–100.0%)	0.717	0.497–0.679 (93.4%–99.3%)
Recreational land use-Private recreation facilities	Private indoor, private outdoor	2	0.24 (0.54)	0.706 (0.527–0.826)	0.503–1.000 (89.1%)	0.696	0.659–0.704 (96.8%–98.3%)
DLU Commercial (an interim subscale, may be used independently, but not included in overall scores)	Sum of shops, restaurant/entertainment, and services subscales. Subscale created to reflect most common pedestrian destinations. Not included in overall positive subscale.	3	5.93 (6.21)	0.906 (0.838–0.947)		0.889	
DLU Overall Positive Subscale	Sum of subscales: residential mix, shops, restaurants/entertainment, services, government services, religious,	10	8.86	0.880		0.855	

Subscale	Label	Number of Items	Mean (SD)	Virtual Tool ^a		In-field Tool ^b	
				ICC (95% Confidence Interval)	Range of Item Kappas or ICCs (% Agreement)	ICC	Range of Item Kappas or ICCs (% Agreement)
	school, positive parking, public recreation, and private recreation		(6.29)	(0.794–0.931)			
Negative Subscale							
DLU Overall Negative Subscale	Warehouse/factory/industrial, abandoned building, unmaintained lot/field, casino, large parking facilities	5	1.74 (1.34)	0.794 (0.657–0.880)	0.479–0.700 (80.4%–100.0%)	0.610	–0.029–0.659 (76.2%–100%)
Overall							
DLU_overall	DLU Overall Positive Subscale Score minus DLU Overall Negative Subscale Score		7.12 (5.89)	0.849 (0.743–0.913)		0.801	
Streetscape							
Positive Elements Subscale	Transit stops, posted speed limit, pedestrian signage, street amenities (e.g., working telephone, trash bins)	18	3.10 (1.59)	0.616 (0.401–0.767)	0.375–0.789 (65.2%–100.0%)	0.741	0.395–0.838 (57.4%–98.9%)
Negative Elements Subscale	High speed limits, roll-over curbs, driveways	5	2.09 (0.98)	0.685 (0.497–0.812)	0.433–0.814 (80.4%–95.8%)	0.742	0.433–0.814 (76.3%–95.8%)
Overall Streetscape Score	Positive Streetscape Elements Subscale Score minus Negative Streetscape Elements Subscale Score		1.01 (2.09)	0.644 (0.438–0.785)		0.762	
Aesthetics and Social							
Positive Aesthetics and Social Subscale	Public art, landscaping maintenance	5	2.15 (0.97)	0.485 (0.231–0.677)	0.292–0.457 (61.0%–91.3%)	0.632	0.391–0.689 (61.0%–91.0%)
Negative Aesthetics and Social Subscale	Graffiti, physical disorder, broken windows	3	0.38 (0.75)	0.409 (0.140–0.623)	0.188–0.483 (71.3%–95.7%)	0.514	0.088–0.665 (68.6%–100%)
Overall Aesthetics and Social Subscale	Positive Aesthetics and Social Subscale Score minus Negative Aesthetics and Social Subscale Score		1.77 (1.33)	0.563 (0.331–0.731)		0.580	
Overall							
Total Route Score	Sum of three over scores		9.90 (7.21)	0.832 (0.717–0.903)		0.816	

^aMeasured by Google Earth between virtual raters with different familiarities of the environment.

Measured in-field between raters with same familiarity of the environment (Millstein RA, Cain KL, Sallis JF, et al. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). BMC Public Health. 2013 Apr 27;13:403.).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2
Agreement on segment subscales between virtual raters with different familiarities of the environment (n=141)

Subscale	Label	Number of Items	Mean (SD)	Virtual Tool ^a		In-field Tool ^b	
				ICC (95% Confidence Interval)	Range of Item Kappas or ICCs or ICCs (% Agreement)	ICC	Range of Item Kappas or ICCs (% Agreement)
Positive Subscale							
Building Height and Setbacks	Smallest and largest setbacks and building height	3	1.01 (0.49)	0.448 (0.284–0.586)	0.232–0.906 (65.7%–99.1%)	0.370	0.522–0.764 (50.7%–97.5%)
Sidewalk Positive Qualities	Sidewalk presence and width	3	2.09 (0.45)	0.360 (0.181–0.516)	-0.013–0.865 (12.5%–98.1%)	0.555	0.489–1.000 (81.8%–100%)
Buffers	Buffer presence and width	2	0.45 (0.84)	0.887 (0.839–0.922)	0.699–0.971 (77.3%–96.3%)	0.940	0.882–0.919 (95.3%–96.5%)
Bicycle Infrastructure	Street-level windows, building colors and materials	2	0.38 (0.93)	0.724 (0.621–0.803)	0.617–0.798 (90.7%–96.3%)	0.855	0.676–0.791 (97.1%–97.3%)
Building Aesthetics and Design	Number and spacing of trees, percent of sidewalk shaded	4	3.20 (1.57)	0.604 (0.470–0.711)	0.185–0.578 (41.4%–80.2%)	0.705	0.549–0.629 (56.4%–80.2%)
Trees	Is there an informal path (shortcut) which connects to something else?	3	1.75 (1.26)	0.719 (0.606–0.804)	0.247–0.774 (51.9%–95.4%)	0.744	0.540–0.737 (55.4%–91.6%)
Overall Positive	Smallest and largest setbacks and building height	6	9.02 (3.13)	0.797 (0.709–0.861)		0.752	
Negative Subscale							
Sidewalk Negative Qualities	Trip hazard, obstructions in the sidewalk	5	0.28 (0.55)	0.360 (0.181–0.516)	0.164–0.862 (63.6%–98.0%)	0.675	0.476–0.796 (61.5%–93.6%)
Building Height: Road Width and Setback Ratio	Smallest and largest setbacks, building height, and road width	3	0.20 (0.07)	0.588 (0.449–0.698)	0.435–0.960 (38.0%–95.4%)	0.614	0.522–0.808 (36.4%–96.2%)
Negative Street Design/width	Traffic lanes, one-way or two-way	2	1.34 (0.47)	0.897 (0.854–0.929)	0.897–1.000 (95.4%–100%)	0.706	0.696–0.711 (93.8%–99.3%)
Overall Negative Subscale	Sum of subscales: Sidewalk negative qualities, building height: road width and setback ratio, negative street design/width	3	1.82 (0.73)	0.364 (0.180–0.522)		0.689	
Overall Subscales							

Subscale	Label	Number of Items	Mean (SD)	Virtual Tool ^a		In-field Tool ^b	
				ICC (95% Confidence Interval)	Range of Item Kappas or ICCs (% Agreement)	ICC	Range of Item Kappas or ICCs (% Agreement)
Overall Segments Score	Overall Positive – Overall Negative subscales		8.91 (3.07)	0.733 (0.619–0.816)	0.753		

^aMeasured by Google Earth between virtual raters with different familiarities of the environment.

^bMeasured in field between raters with same familiarity of the environment (Millstein RA, Cain KL, Sallis JF, et al. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). BMC Public Health. 2013 Apr 27;13:403.).

Table 3
Agreement on crossing subscales between virtual raters with different familiarities of the environment (n=92)

Subscale	Label	Number of Items	Mean (SD)	Virtual Tool ^a		In-field Tool ^b	
				ICC (95% Confidence Interval)	Range of Item Kappas or ICCs (% Agreement)	ICC	Range of Item Kappas or ICCs (% Agreement)
Positive Subscale							
Crosswalk Amenities/Qualities	Crosswalk characteristics (e.g., marked crosswalk, high visibility markings)	9	0.81 (1.01)	0.577 (0.392–0.718)	0.304–0.660 (77.1%–100%)	0.807	-0.012–0.816 (86.8%–99.7%)
Curb Quality/Presence	Pre- and post-crossing curb lining up with crossing	2	1.73 (0.66)	0.790 (0.680–0.865)	0.635–0.861 (90.0%–94.0%)	0.684	0.648–0.651 (81.8%–84.7%)
Intersection Control and Signage	Stop signs, pedestrian walk signals	10	3.88 (2.85)	0.657 (0.498–0.773)	0.359–0.793 (72.9%–100%)	0.752	0.327–0.811 (88.4%–98.7%)
Overall Positive Crossing Characteristics Subscale	Sum of subscales: crosswalk amenities/qualities, curb quality/presence, intersection control and signage	3 subscales	3.24 (1.96)	0.703 (0.555–0.808)		0.828	
Negative Subscale							
Lanes/Road Width of Crossing	Distance of crossing leg (# lanes wide, trichotomized)	1	1.16 (0.56)	0.348 (0.126–0.537)	0.553–0.517 (64.3%–68.6%)	0.525	0.524 (66.0%–72.9%)
Crossing Impediments	No curb ramp, gutters in crossing, faded/worn crosswalk markings	7	0.19 (0.49)	0.800 (0.695–0.872)	0.767–0.858 (94.3%–100%)	0.728	0.188–0.893 (83.0%–99.4%)
Overall Negative Crossing Characteristics Subscale	Sum of subscales: Lanes/Road Width of Crossing, Crossing Impediments	2 subscales	0.88 (0.75)	0.548 (0.357–0.695)		0.587	
Overall Subscales							
Overall Crossings Score	Sum of subscales: Overall Positive Crossing Characteristics-Overall Negative Crossing Characteristics		3.60 (3.48)	0.771 (0.650–0.854)		0.830	

^aMeasured by Google Earth between virtual raters with different familiarities of the environment.

^bMeasured in field between raters with same familiarity of the environment (Millstain RA, Cain KL, Sallis JF, et al. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). BMC Public Health. 2013 Apr 27;13:403.).