Journal of Transport & Health 12 (2019) 75–85

Contents lists available at ScienceDirect

Journal of Transport & Health

journal homepage: www.elsevier.com/locate/jth

Sampling methodology and reliability of a representative walkability audit

Emily N. Ussery[a,b,*], John D. Omura[a,b], Prabasaj Paul[b], John Orr[c], Chad Spoon[d], Carrie Geremia[e], Susan A. Carlson[b]

[a] Epidemic Intelligence Service, Centers for Disease Control and Prevention, Center for Surveillance, Epidemiology, and Laboratory Services, Atlanta, GA, USA
[b] Centers for Disease Control and Prevention, Division of Nutrition, Physical Activity, and Obesity, Physical Activity and Health Branch, Atlanta, GA, USA
[c] US Virgin Islands Department of Health, Christiansted, St. Croix, US Virgin Islands
[d] University of California at San Diego, Active Living Research, San Diego, CA, USA
[e] University of California at San Diego, Department of Family and Preventive Medicine, San Diego, CA, USA

ARTICLE INFO

Keywords:
Walkability audit
Built environment
Physical activity
US Virgin Islands
Microscale Audit of Pedestrian Streetscapes

ABSTRACT

Background: Physical inactivity is a public health concern in the US Virgin Islands (USVI). A contributing factor may be a lack of pedestrian infrastructure and other environmental supports for walking. In this manuscript, we describe the methods used to conduct a walkability audit of environmental features related to physical activity in the USVI.
Methods: In 2016, volunteer auditors conducted the audit using a modified version of the Microscale Audit of Pedestrian Streetscapes tool. A two-stage sampling method was developed using publicly available census data to select a sample of estates (n = 46) and street segments (n = 1550; 99.2 km) across the USVI. A subset of segments was audited by two independent auditors, and inter-rater reliability was assessed using Cohen's kappa and percent agreement.
Results: Audits were completed on 1114 segments (94.6 km), and estimates were weighted to represent accessible public street length in the study area (1155.9 km). Most items on the audit tool (62.7%) demonstrated good to excellent reliability. We found that it was feasible to conduct a reliable audit of environmental features related to physical activity across a large sample of streets in the USVI.
Conclusions: These methods can be replicated in other settings to collect comprehensive data that can be used to guide strategies to improve the walkability of communities.

1. Background

Physical inactivity is a significant public health concern in the United States, and improvements to the built environment are recommended to create safe and accessible opportunities for physical activity (Community Preventive Services Task Force, 2017). *Step It Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities* calls for the creation of walkable communities and improved monitoring of environmental supports for walking and other types of physical activity (US Department of

* Correspondence to: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 4770 Buford Highway NE, MS-77, Atlanta, GA 30341-3717, USA.
E-mail addresses: yzv4@cdc.gov (E.N. Ussery), ydk8@cdc.gov (J.D. Omura), vig5@cdc.gov (P. Paul), John.Orr@doh.VI.gov (J. Orr), scspoon@ucsd.edu (C. Spoon), cgeremia@ucsd.edu (C. Geremia), clo3@cdc.gov (S.A. Carlson).

https://doi.org/10.1016/j.jth.2018.11.007
Received 25 April 2018; Received in revised form 6 November 2018; Accepted 27 November 2018
2214-1405/ Published by Elsevier Ltd.

Health and Human Services, 2015). Supportive features include mixed land use, well-maintained and interconnected sidewalks, access to public transit, and traffic calming features that reduce vehicle speeds (Community Preventive Services Task Force, 2017; Saelens and Handy, 2008; McCormack and Shiell, 2011). Periodic assessments of environmental supports for physical activity, such as through walkability audits, would provide useful data for public health planning (Fulton et al., 2016). However, systematic data on the presence of such features across US states and territories are lacking. Moreover, walkability audits have primarily been used in research settings, although their potential value as a public health surveillance tool has been acknowledged in recent years (Fulton et al., 2016).

The United States Virgin Islands (USVI) has one of the highest levels of physical inactivity compared to other states and territories, with almost one-third of adults engaging in no leisure time physical activity (Xu et al., 2013). A lack of pedestrian infrastructure and unsupportive community design in the USVI may contribute to the high prevalence of physical inactivity. In order to identify barriers and guide future efforts to promote physical activity in the USVI, the authors developed methods to conduct a comprehensive walkability audit across the territory in 2016. The goal of the USVI audit was to collect representative surveillance data on features of the built environment across streets in the territory, making this one of the first systematic, state- or territory-wide assessments using an observational audit. While previously developed methods for conducting audits have unique strengths, a standardized approach for selecting a representative sample of streets across a large geographic area that would be feasible to audit given limited time, resources, and data availability had not been previously established. Thus, we drew upon strengths of existing methods to develop an approach that would accomplish this goal in the USVI.

1.1. Audit sampling methodology

Methods for determining the areas of observation for walkability audits vary widely across studies (Schaefer-McDaniel et al., 2010), corresponding to each study's specific objectives and geographic scope. For example, several studies have collected observational audit data to examine associations with individual-level health behaviors or outcomes (Hajna et al., 2013; Gallimore et al., 2011; Evenson et al., 2009). To accomplish this, street segments in enrolled study participants' neighborhoods have been selected via random sampling (Hajna et al., 2013), using a route-based approach beginning with a home or school (Gallimore et al., 2011), or by including all segments within GIS-derived buffers around a point of interest (Evenson et al., 2009). Other studies have utilized sampling strategies designed to enroll participants residing in neighborhoods with varying levels of sociodemographic characteristics and GIS-derived macroscale features of the built environment (Thornton et al., 2016; Witten et al., 2012; International Physical Activity and the Environment Network, 2001). For example, the International Physical Activity and the Environment Network (IPEN) protocol recommends stratifying neighborhoods (e.g. census block groups) by income level and a walkability index consisting of GIS-based residential density, intersection density, land use mix, and retail floor area ratio and selecting neighborhoods from the high and low deciles (Witten et al., 2012; International Physical Activity and the Environment Network, 2001). Individuals residing within selected neighborhoods are then randomly sampled, and a selection of streets is audited to assess the pedestrian streetscape (Witten et al., 2012; International Physical Activity and the Environment Network, 2001). This design enables comparison of activity levels and microscale features of walkability by different neighborhood types, but may not provide representative estimates of walkability across a defined geographic area for surveillance purposes (International Physical Activity and the Environment Network, 2001). To our knowledge, only one audit has been conducted to assess the pedestrian and bicycling environment across an entire US state or territory (Maddock et al., 2012). This study randomly selected approximately 13% of road segments from each of the four counties in Hawaii for assessment proportional to population size, but did not stratify by other characteristics to maximize representativeness. Kelly and colleagues (2014) employed stratified random geographic sampling to ensure that their sample included street segments in neighborhoods characterized by a range of land uses and socioeconomic characteristics across 2 large US cities (Kelly et al., 2014). The authors of that study recommended that future audits include population density as a sampling stratification variable, as built environment features and physical activity may vary by density. Finally, a model-based sampling approach has been proposed by Moniruzzaman and Paez (2012) to maximize resources by focusing data collection on areas where transportation walking is more or less prevalent than predicted by a model. The authors propose that this method reduces the number of street segments necessary to produce a comprehensive sample. However, the modeling approach requires active transportation mode share data and other explanatory variables that may not be available in all settings.

1.2. Audit reliability testing

Previous studies have found that observational measures of walkability collected via audits demonstrate good predictive validity of physical activity behaviors, particularly active transportation (Cain et al., 2017; Cain et al., 2014). Moreover, most measures across audit tools have high intra- and inter-rater reliability, with objective measures demonstrating higher reliability compared to more subjective or abstract measures (Clifton et al., 2007; Millstein et al., 2013; Pikora et al., 2002). However, it was unknown if the reliability of an established, validated audit tool could be maintained in a public health response setting characterized by limited time for training, administration by volunteers, and modification of the tool to enhance relevance to the local environment.

The objectives of this paper are twofold. First, we describe the development and implementation of the methods used to select a sample of street segments across the USVI, which drew upon strengths of approaches used in previous studies. Second, we evaluate the inter-rater reliability of the audit tool used to complete the assessment. A discussion of lessons learned is also included to guide others who are interested in applying these methods to conduct a comprehensive audit in their own jurisdictions.

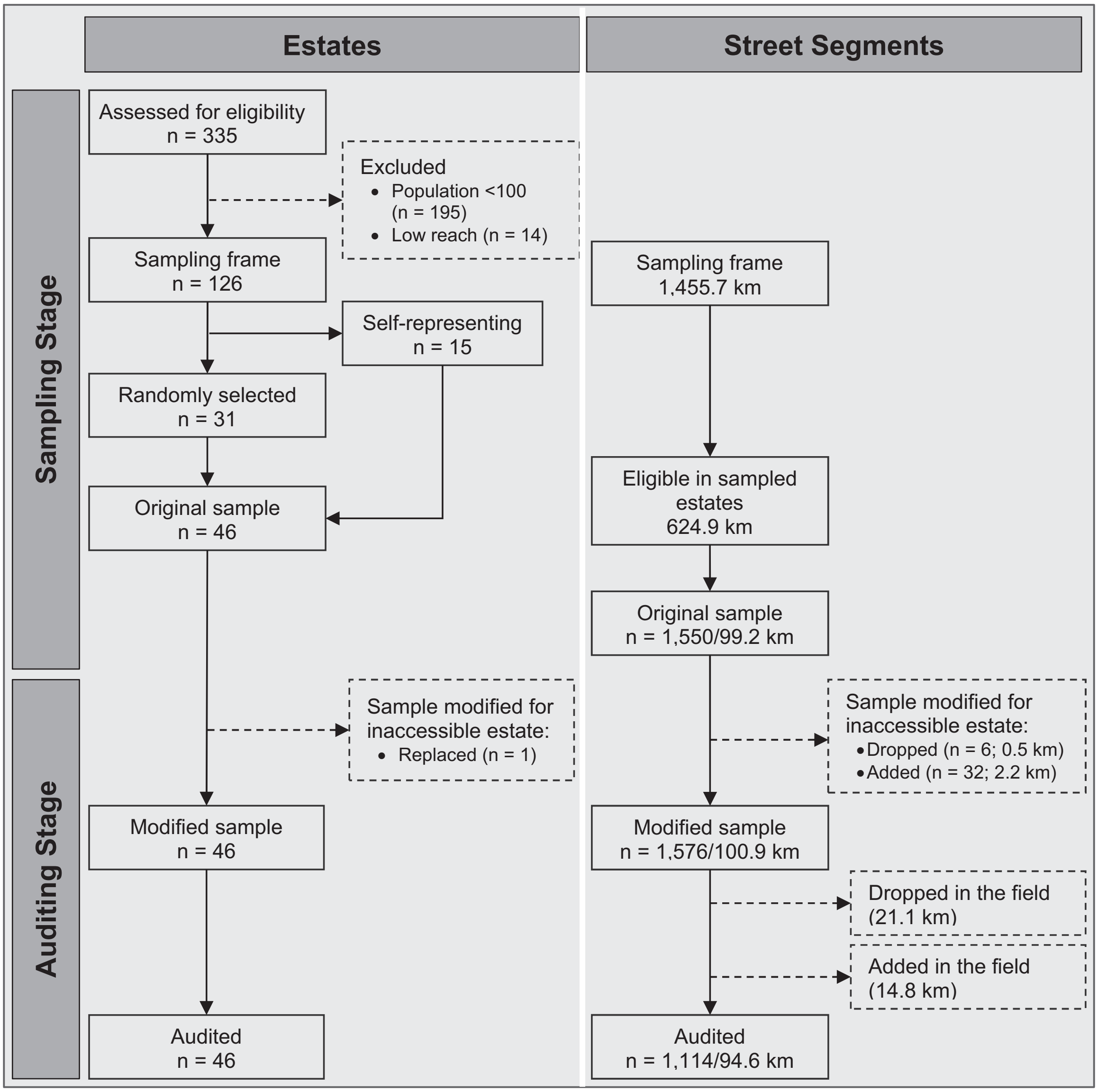

Fig. 1. Flow diagram of the sampling and auditing stages for estates and street segments in the US Virgin Islands. In the "Estates" panel, the n's represent the number of estates that were sampled or audited. In the "Street Segments" panel, the n's represent the number of street segments that were sampled or audited.

2. Methods

2.1. Study design and setting

In 2016, an observational audit was conducted to assess the prevalence of environmental features related to physical activity across the three main US Virgin Islands—St. Croix, St. John, and St. Thomas.

2.2. Sampling method

A two-stage sampling method was used to select a sample of streets across the territory (Fig. 1). Most data inputs were derived from publicly available sources; the only exception was location of schools, which was provided by the USVI DOH. In the first stage, estates were selected using stratified random sampling. The USVI are divided into 335 estates, the smallest legal subdivision for which US Census data are published (Fig. 2). Population size, population density (persons/acre), and population reach (persons/kilometer [km] of street length) for each estate were obtained using the US Census Bureau's 2010 Demographic Profile Data and 2015 Topologically Integrated Geographic Encoding and Referencing (TIGER)/Line files representing the road network across the USVI

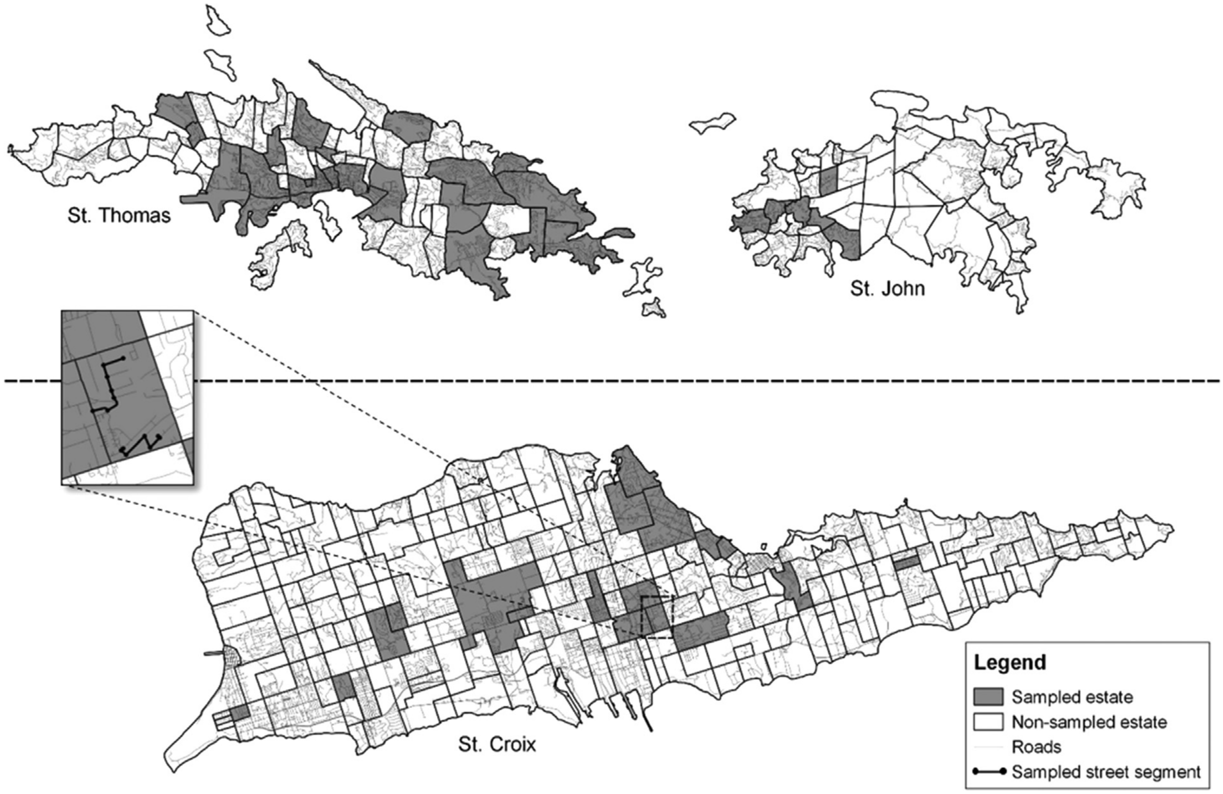

Fig. 2. Map of sampled estates on the three main US Virgin Islands. The inset depicts an example of sampled street segments within a sampled estate.

(U.S. Census Bureau, 2015). Population reach is a measure of density that uses street length as the denominator. Estates were excluded if they had population < 100 or low population reach (< 45 persons/km of streets) and were not neighboring an estate with high population reach (≥45 persons/km of streets) (n = 209). The remaining 126 estates were included in the sampling frame, which captured 93.3% of the population.

These estates were stratified by four variables: island (St. Croix, St. John, or St. Thomas), population density (dense: > 10 persons/acre; sparse: ≤10 persons/acre), population reach (high reach: ≥45 persons/km; neighboring a high reach estate), and number of schools. Fifteen estates were included in the sample independent of the sampling design because of their dense population (n = 7) or the presence of schools (n = 8). Streets near schools were identified as a target area for study and potential intervention, so we wanted to ensure that estates with schools were included in the sample. Of the remaining 111 estates, a random sample was drawn from each stratum. This process added an additional 31 estates, bringing the total number of sampled estates to 46 (Fig. 2).

In the second stage, street segments within sampled estates were selected. A street segment was defined as the length of road between two intersections. The 2015 TIGER/Line files were used to derive the sampling frame of street segments (U.S. Census Bureau and 2015, 2015). Streets classified as private roads and parking lots were excluded. Thus, the original sampling frame consisted of segments categorized as secondary roads, local roads, city streets, trails, pedestrian trails, passageways, or alleys.

Within each sampled estate, two "seed" segments were randomly selected. From each seed segment, an adjoining segment was randomly selected without replacement to create a route. Adjoining segments selected without replacement were added to the route until the length of selected segments from all routes represented approximately 15% of the total street length in the estate. If an adjoining segment was not available (e.g., dead end, estate boundary), a new seed segment was randomly selected, and an additional route was created. Fig. 2 depicts an example of adjoining street segments made up of two routes in a sampled estate. The target sample size was determined based on estimates of street length that would be feasible to audit given available resources and time. For this study, we trained 26 volunteer auditors from the USVI DOH, with each able to dedicate about 2 days to auditing. We had 12 days available for data collection, with three teams of two auditors in the field each day. We estimated each team could complete audits on 3 km of street length per day, generating a target sample size of approximately 100 km. Selection of street segments was performed in R with the "rgdal" package for processing geocoded data.

2.3. Audit tool

The Microscale Audit of Pedestrian Streetscapes (MAPS) Abbreviated tool was adapted for this project to ensure feasibility and contextual relevance (Cain et al., 2017; Millstein et al., 2013). MAPS assesses modifiable microscale features of the environment relevant for physical activity and has been used in a variety of settings. Most items have previously demonstrated moderate to excellent inter-rater reliability (Millstein et al., 2013). Several items were added to assess context-specific features (e.g., presence of beaches, stray dogs), and other items were removed to maximize feasibility of data collection (e.g., building setback) or because they were not commonly found in the local context (e.g., liquor stores). The adapted tool, MAPS-USVI, included 46 questions. Several items on the original MAPS tool were assessed at the route-level (e.g., land use type, presence of destinations), but all items on the MAPS-USVI tool were assessed for each individual segment.

2.4. Data collection

Data collection occurred in May 2016. USVI DOH staff volunteers were trained to administer the MAPS-USVI tool through a remote webinar several weeks before data collection and an in-person refresher training 1 day before data collection. Each volunteer also completed a series of practice field audits, with real-time troubleshooting and feedback provided by certified MAPS trainers. On each day of data collection, three teams of at least two trained auditors (one or more USVI DOH volunteer plus one project staff per team) were assigned street segments. Team members completed separate surveys for each side of a segment, yielding two completed surveys per segment. If a portion or all of a sampled street segment was not audited, auditors recorded the reason for noncompletion. A street segment was classified as "out-of-scope" if it did not exist, was a driveway/private road, or was otherwise inaccessible (e.g. under construction, fenced off). Inter-rater reliability was evaluated on a randomly selected subset of segments (10% of original sample), with the second audit completed by an independent auditor within 1 week of the first.

2.5. Weighting

Weights were calculated to generate estimates representative of accessible public street length across the USVI in two steps: length-based weighting and post-stratification. First, each segment was assigned an initial weight that accounted for the probability of an estate being selected from a stratum (number of selected estates in stratum/total number of estates in stratum) and the percentage of street length in an estate that was audited (total length of audited segments in estate/total street length in estate). This initial combined weight was then multiplied by the length of each individual audited segment.

Next, post-stratification factors were created to generate weighted counts that added to the total street length in the adjusted sampling frame. In this step, the percentage of sampled street length found to be "out-of-scope" was calculated for each stratum, and the total street length was reduced by this percentage. Within each stratum, a post-stratification factor was then calculated by dividing the adjusted total street length by the sum of the length-based weights. The final weight for each audited segment was constructed by multiplying its length-based weight by the stratum-specific post-stratification factor.

2.6. Statistical analysis

2.6.1. Sampling process

Descriptive statistics were calculated for the sample of street segments at multiple stages: 1) original sampling frame, 2) original sample, 3) unweighted audited sample, 4) weighted audited sample before post-stratification, and 5) weighted and post-stratified audited sample. Sample characteristics were calculated overall and by the following estate-level variables: island, population density, population reach, and number of schools.

2.6.2. Inter-rater reliability

Inter-rater reliability was assessed for individual items on the MAPS-USVI tool, except those for which one response option had a frequency of 100% (n = 33; 33% of total). Dichotomous variables were tested using the Cohen's kappa (κ) statistic, and ordinal and nominal variables were tested using the weighted Cohen's κ. The following cutpoints were used to classify the strength of reliability: good/excellent (≥0.60), moderate (0.41–0.60), or fair/poor (≤0.40). Percent agreement was also calculated and described using the following cutpoints: good/excellent (≥75%), moderate (60–74%), and fair/poor (< 60%). These cutpoints were used for the original MAPS tool (Millstein et al., 2013; Landis and Koch, 1977).

3. Results

3.1. Sampling

3.1.1. Sampling stage

The original sampling frame consisted of 126 estates with 1455.7 km of street length. A total of 99.2 km (n = 1550 street segments) were sampled, which was 15.9% of street length in selected estates. Compared to the sampling frame, the original sample had a smaller percentage of street length on St. Croix, in estates with low population density, and in estates with no schools (Table 1).

3.1.2. Auditing stage

One estate was dropped because all sampled street segments were inaccessible (n = 6 segments, 0.5 km), and a replacement estate was randomly selected from the same stratum (n = 32 segments, 2.18 km). From this sample, 21.1 km were not audited for the following reasons: private (47.3%), did not exist (18.0%), inaccessible (23.1%), included twice in the sample (3.4%), field error

Table 1
Characteristics of Sample at the Sampling, Data Collection, and Data Analysis Stages, US Virgin Islands, 2016.

Characteristic	Stage									
	Sampling				Data Collection		Data Analysis			
	Sampling Frame		Original Sample		Unweighted		Weighted		Post-Stratified	
	km[a]	(%)	km[b]	(%)	km[c]	(%)	km[d]	(%)	km[e]	(%)
Total	1455.7	(100)	99.2	(100)	94.6	(100)	793.0	(100)	1155.9	(100)
Island										
St. Croix	801.0	(55.0)	37.6	(37.9)	34.6	(36.6)	417.1	(52.6)	602.1	(52.1)
St. John	72.5	(5.0)	8.0	(8.1)	7.7	(8.2)	33.3	(4.2)	61.4	(5.3)
St. Thomas	582.2	(40.0)	53.6	(54.0)	52.3	(55.3)	342.5	(43.2)	492.4	(42.6)
Population density										
Sparse	1,220.8	(83.9)	72.7	(73.3)	69.2	(73.2)	649.1	(81.9)	910.9	(78.8)
Dense	234.3	(16.1)	26.6	(26.8)	25.4	(26.8)	143.8	(18.1)	245.0	(21.2)
Population reach										
High	972.6	(66.8)	68.4	(69.0)	65.5	(69.2)	529.7	(66.8)	810.5	(70.1)
Neighboring	483.1	(33.2)	30.8	(31.0)	29.1	(30.8)	263.3	(33.2)	345.5	(29.9)
Number of schools										
0	1,070.8	(73.6)	52.3	(52.7)	51.6	(54.6)	595.9	(75.1)	850.5	(73.6)
1	264.8	(18.2)	28.0	(28.2)	24.5	(25.9)	137.0	(17.3)	208.1	(18.0)
2	79.2	(5.4)	12.2	(12.3)	11.1	(11.8)	39.6	(5.0)	70.1	(6.1)
3	41.0	(2.8)	6.8	(6.9)	7.3	(7.7)	20.5	(2.6)	27.2	(2.4)

Abbreviations: km, kilometers.
[a] Unweighted length of all streets in sampling frame.
[b] Unweighted length of street segments in original sample selected for auditing.
[c] Unweighted length of audited street segments.
[d] Audited street segments, after initial weights were applied.
[e] Audited street segments, weighted to represent length of publicly accessible streets in sampling frame, after initial weights, out-of-scope adjustment, and post-stratification factors were applied.

(1.6%), and other nonspecified reason (6.6%). Discrepancies between the GIS-derived endpoint of street segments and the protocol for determining an endpoint in the field occasionally occurred. In these cases, auditors followed the protocol of extending the endpoint to the nearest intersection, which added an additional 14.8 km of street length to the sample. Thus, the unweighted audited sample included 94.6 km of street length (n = 1114 segments). The distribution of street length in the unweighted audited sample was comparable to that of the original sample across the variables of island, population density, population reach, and number of schools.

3.1.3. Weighting stage

A total of 19.4 km of sampled street length (19.2%) was found to be out-of-scope, ranging from 0% to 33.6% within strata. The stratum-specific adjustment resulted in an adjusted sampling frame of 1,155.9 km of street length. When sampling weights and post-stratification factors were applied, the distribution of street length was similar to that of the original sampling frame across the variables of island, population density, population reach, and number of schools. The majority of street length was located on St. Croix (52.1%) and in estates with low population density (78.8%), high population reach (70.1%), and no schools (73.6%).

3.2. Inter-rater reliability

Inter-rater reliability was evaluated on 8.3% (n = 92) of audited segments because of inaccessible streets; this percentage was slightly lower than the 10% recommended in the original MAPS protocol (Cain et al., 2012). Table 2 includes reliability results for dichotomous items (n = 36) on the MAPS-USVI tool, and Table 3 includes results for nominal and ordinal items (n = 31). Most items (64.2%) demonstrated good/excellent inter-rater reliability, 13.4% demonstrated moderate reliability, and 22.4% demonstrated fair/poor reliability. Most items (89.5%) also demonstrated good/excellent percent agreement between ratings, 4.5% demonstrated moderate agreement, and 6.0% demonstrated fair/poor agreement. A total of 42 items (62.7%) received a good/excellent rating for both Cohen's κ and percent agreement. The following four items (6.0%) received a fair/poor rating for both Cohen's κ and percent agreement: percentage of roadway on which parking was allowed, well-maintained buildings, well-maintained landscaping, and litter. Two additional items received a fair/poor rating for Cohen's κ and a moderate rating for percent agreement: hardscape features (e.g., fountains, art) and presence of pedestrians.

4. Discussion

Our findings show that it was feasible to conduct an observational assessment of built environment features related to physical

Table 2
Inter-Rater Reliability for Dichotomous Items on MAPS-USVI Tool.

Measure	Agreement		Cohen's Kappa (κ)	
	%	Rating[a]	κ	Rating[b]
Type of land use	95.6	Good/excellent	0.809	Good/excellent
Residential uses				
Single family houses	80.4	Good/excellent	0.600	Moderate
Multi-unit homes	91.3	Good/excellent	0.744	Good/excellent
Apartments/condominiums[c]	92.4	Good/excellent	0.494	Moderate
Shopping centers				
Strip mall[c]	98.9	Good/excellent	0.883	Good/excellent
None[c]	98.9	Good/excellent	0.903	Good/excellent
Informal transit[c]	92.3	Good/excellent	0.195	Fair/poor
Traffic calming				
Signs[d]	100.0	Good/excellent	1.000	Good/excellent
Speed humps	90.2	Good/excellent	0.727	Good/excellent
Rollover curbs[d]	93.5	Good/excellent	0.367	Fair/poor
None	85.9	Good/excellent	0.658	Good/excellent
Street amenities				
Building overhangs[c]	98.9	Good/excellent	0.883	Good/excellent
Trash bins[d]	95.7	Good/excellent	0.581	Moderate
Benches/place to sit[d]	96.7	Good/excellent	0.711	Good/excellent
Hawkers/shops/carts[c]	94.6	Good/excellent	0.272	Fair/poor
None[c]	93.5	Good/excellent	0.631	Good/excellent
Hardscape features	69.6	Moderate	0.234	Fair/poor
Natural bodies of water	88.0	Good/excellent	0.709	Good/excellent
Softscape features	77.2	Good/excellent	0.549	Moderate
Stray dogs[c]	91.3	Good/excellent	−0.043	Fair/poor
Presence of pedestrians	74.7	Moderate	0.251	Fair/poor
Sidewalk buffer	96.7	Good/excellent	0.811	Good/excellent
Informal path[c]	88.0	Good/excellent	−0.063	Fair/poor
Covered place to walk[d]	97.8	Good/excellent	0.655	Good/excellent
Intersection control				
Yield sign[d]	100.0	Good/excellent	1.000	Good/excellent
Stop sign	94.6	Good/excellent	0.641	Good/excellent
Traffic signal[d]	100.0	Good/excellent	1.000	Good/excellent
Crossing guard[d]	98.9	Good/excellent	0.000	Fair/poor
None	83.7	Good/excellent	0.655	Good/excellent
Signalization				
Walk signal[d]	100.0	Good/excellent	1.000	Good/excellent
Push buttons[d]	100.0	Good/excellent	1.000	Good/excellent
None[d]	85.9	Good/excellent	0.717	Good/excellent
Crosswalk treatment				
Marked[c]	98.9	Good/excellent	0.852	Good/excellent
High visibility striping[c]	98.9	Good/excellent	0.795	Good/excellent
None[c]	84.8	Good/excellent	0.692	Good/excellent
Protected refuge[c]	97.3	Good/excellent	0.654	Good/excellent

[a] Fair/poor: < 60%; moderate: 60%–74%; good/excellent: ≥75%.
[b] Fair/poor: ≤0.40; moderate: 0.41–0.60; good/excellent: ≥0.60.
[c] Frequency of one response option was > 90%.
[d] Frequency of one response option was > 95%.

activity across a large sample of streets in the USVI with good inter-rater reliability. This study demonstrates the utility of the sampling methodology that was developed to select a sample of street segments across the USVI using publicly available census data. Moreover, we found that reliability of the modified MAPS tool was maintained when used in this setting. The methods developed for the present audit can serve as a guide for others seeking to conduct an assessment of the built environment in a defined geographic area.

An extensive body of literature has emerged in recent years describing methods for administering observational audits and selecting areas for observation (Schaefer-McDaniel et al., 2010; Hajna et al., 2013; Gallimore et al., 2011; Evenson et al., 2009; Thornton et al., 2016; Witten et al., 2012; International Physical Activity and the Environment Network, 2001). However, such studies are often designed for research purposes and their methods may not be appropriate for collecting surveillance data via walkability audits (Hajna et al., 2013; Gallimore et al., 2011; Evenson et al., 2009; Thornton et al., 2016; Witten et al., 2012; International Physical Activity and the Environment Network, 2001); only one systematic audit has been published at the state level (Maddock et al., 2012). The methods reported here can be applied in other settings, although some tailoring to the characteristics of each study setting may be necessary. For example, a different geographic unit (e.g., census tract) may be a more appropriate primary sampling unit in other settings, since estates are unique to the USVI. The stratified sampling design ensured that our sample captured

Table 3
Inter-Rater Reliability for Nominal and Ordinal Items on MAPS-USVI Tool.

Measure	Agreement		Weighted Cohen's Kappa (κ)	
	%	Rating[a]	κ	Rating[b]
Destinations				
Fast food restaurant[c]	97.8	Good/excellent	0.662	Good/excellent
Sit-down restaurant[c]	96.7	Good/excellent	0.562	Moderate
Grocery/supermarket[c]	98.9	Good/excellent	0.000	Fair/poor
Bank[c]	95.7	Good/excellent	0.483	Moderate
Hotel[c]	100.0	Good/excellent	1.000	Good/excellent
Health-related professional[c]	97.8	Good/excellent	−0.011	Fair/poor
Other service[c]	95.7	Good/excellent	0.530	Moderate
Other retail[d]	96.7	Good/excellent	0.861	Good/excellent
Place of worship[c]	98.9	Good/excellent	0.795	Good/excellent
School[c]	98.9	Good/excellent	0.795	Good/excellent
Indoor recreation[c]	96.7	Good/excellent	0.389	Fair/poor
Public park[c]	96.7	Good/excellent	0.555	Moderate
Traffic lanes[d]	97.8	Good/excellent	0.827	Good/excellent
On-road parking	50.0	Fair/poor	0.247	Fair/poor
Public transit stop[c]	100.0	Good/excellent	1.000	Good/excellent
Street lighting	84.4	Good/excellent	0.689	Good/excellent
Building maintenance	52.2	Fair/poor	0.256	Fair/poor
Landscape maintenance	45.1	Fair/poor	0.398	Fair/poor
Graffiti[c]	96.7	Good/excellent	0.483	Moderate
Litter	38.5	Fair/poor	0.227	Fair/poor
Sidewalk characteristics				
Sidewalk presence	95.7	Good/excellent	0.866	Good/excellent
Width	92.4	Good/excellent	0.813	Good/excellent
Trip hazards	95.7	Good/excellent	0.886	Good/excellent
Temporary obstructions[d]	95.3	Good/excellent	0.882	Good/excellent
Number of trees	92.4	Good/excellent	0.845	Good/excellent
Percentage tree coverage	96.7	Good/excellent	0.929	Good/excellent
Percentage awning coverage	96.7	Good/excellent	0.890	Good/excellent
Number of driveways	73.6	Moderate	0.748	Good/excellent
Crossing characteristics				
Crossing presence	83.7	Good/excellent	0.747	Good/excellent
Pre-crossing curb ramp[d]	84.8	Good/excellent	0.740	Good/excellent
Post-crossing curb ramp	82.6	Good/excellent	0.679	Good/excellent

[a] Fair/poor: < 60%; moderate: 60%–74%; good/excellent: ≥75%.
[b] Fair/poor: ≤0.40; moderate: 0.41–0.60; good/excellent: ≥0.60.
[c] Frequency of one response option was > 95%.
[d] Frequency of one response option was > 90%.

estates with a range of population densities and with or without schools on each island. When comparing across stratification variables, we found that the original sample included more street length on St. John and St. Thomas, in high density estates, and in estates with schools compared to the sampling frame. These differences were likely caused by the inclusion of several self-representing estates with these characteristics. The weighting process adjusted for this oversampling, producing a final sample that closely resembled the sampling frame. A different set of stratification variables may be more appropriate to maximize representativeness in other settings. Setting-specific attributes and planned uses of the data should be considered when selecting these variables.

When developing the sampling methods, decisions were made to maximize the validity and representativeness of our findings within the practical constraints of the project. We estimated that it would be feasible to complete audits on approximately 100 km of street length with the available number of volunteer auditors and time for data collection. This approach produced a sample consisting of approximately 15% of street length in selected estates, which was assumed to adequately represent each estate. We found 100 km to be an accurate estimation of the length that could be assessed with available resources because auditors were able to complete data collection on schedule. However, no consensus exists on the amount of street length that should be audited to ensure representativeness in a given geographic area (Brownson et al., 2009), and our assumption could not be tested in the present study. McMillan and colleagues compared audit results for all residential street segments in 11 neighborhoods (defined as a 400 m buffer around public housing developments) versus randomly selected samples consisting of 25%, 50%, and 75% of street segments (McMillan et al., 2010). They found no significant differences on key built environment variables when comparing each sample to the census of streets, and concluded that sampling as few as 25% of streets in a neighborhood may produce an accurate representation of the pedestrian built environment. However, this study did not examine whether a sample consisting of less than 25% of street segments would produce similar results, and it is unknown whether the findings would generalize to other settings or larger geographic areas. An audit that assessed the pedestrian and bicycling environment in Hawaii randomly selected 13% of road segments from each of four counties across the state (Maddock et al., 2012), which is similar to the percentage of street length audited in our study. Additional research evaluating the optimal sampling percentage necessary for valid assessment of micro-scale features of the environment would be beneficial for future audits, particularly those being conducted across larger geographic areas.

Another issue that required a balance between scientific quality and feasibility was the route-based sampling method. Sampling for the original MAPS tool follows a route-based approach to assess features of a study participant's neighborhood, with each route beginning at a participant's home and extending 0.25 miles toward a destination (Millstein et al., 2013). For the USVI audit, street segments were randomly selected to enhance representativeness of the study area, but aspects of the route-based approach were maintained for time efficiency and ease of locating segments in the field. Certain types of segments (e.g., those more centrally located among a cluster of streets) may have been more likely to be selected into a route, but our length-based weights did not account for any potential differences in segments' probability of selection. Thus, we assumed that no confounding factors would prevent generalizing from a route-based sampling strategy to length-based weighted results. Other researchers may want to examine this assumption.

MAPS is a comprehensive audit tool that has demonstrated strong validity and reliability in a variety of settings (Cain et al., 2017; Cain et al., 2014; Millstein et al., 2013). The use of this well-established tool provided our study with a solid foundation. Our results suggest that, in general, reliability was maintained when the tool was administered by volunteer auditors in a public health response setting, after slight modifications to maximize efficiency and relevance to the local context. Items with low reliability were subjective (e.g., hardscape features, building maintenance, landscape maintenance) or transient in nature (e.g., on-street parking, litter, pedestrians). This finding is consistent with previous literature, which suggests that items like land use and street characteristics demonstrate more favorable reliability compared to subjective items like aesthetics, building maintenance, and safety-related features (Clifton et al., 2007; Millstein et al., 2013). Thus, results for these items should be interpreted with caution. For all items that were rated good/excellent for percent agreement but fair/poor for Cohen's κ, one response option had a frequency higher than 90%. This lack of variance likely contributed to low reliability but high percent agreement (Sim and Wright, 2005). When features exhibit low variance or are not frequently observed, percent agreement is likely to more accurately represent reliability compared to Cohen's κ (Clifton et al., 2007).

The street network dataset used in this project had benefits and limitations. TIGER streets data are freely available for the United States and allow remote versus on-the-ground sampling. However, our results suggest that completeness, a data quality indicator that refers to the presence or absence of geocoded features and their attributes (ISO/TC 211, 2009), was limited in this setting. About 20% of sampled street length was found to be nonexistent, private, or inaccessible. Moreover, some existing streets may not have been captured in the dataset and thus excluded from the sampling frame. Previous studies have found that street network datasets often omit minor roads (Frizzelle et al., 2009; Hawbaker and Radeloff, 2004), although none has looked specifically at data quality for the USVI. Poor completeness can introduce bias if design features of omitted streets differ from those of included streets, but the direction of any potential bias in this setting is unknown. To assess potential limitations, individuals planning an audit can work with community representatives to cross-reference street network datasets with local information about the study area or other available data sources, such as satellite imagery.

Several key lessons for planning a comprehensive walkability audit were learned. First, adequate training and guidance in the field were critical in ensuring that the volunteer staff with no previous auditing experience was equipped to collect valid and reliable data. The practice field audits were a necessary component of the training process, but we had limited time with volunteers to conduct these before data collection. Audit teams consisted of one certified MAPS auditor and one volunteer each day, which enabled frequent discussion and mutual resolution when questions arose in the field, with oversight provided by the more experienced auditor. Efforts should be made to ensure that volunteer auditors have ample time for in-field training before data collection, as well as ongoing support in the field. Second, the planning team met nightly to discuss issues encountered in the field and develop plans for managing them, such as creating a protocol for documenting inaccessible street segments and reaching a consensus on common questions about the tool from volunteer auditors. This real-time problem-solving allowed us to identify problems early and minimize their effect on data quality. Finally, the in-person audit process was resource-intensive, and considerable time and coordination were required to train volunteers, develop materials, and conduct the audits. Including lunch breaks and travel time between segments, the audits required approximately 576 h of auditor time in the field, or 31.0 min per auditor per segment. The reliability subsample required an additional 45 h of auditor time, or 29.3 min per auditor per segment. Remote audits using web-based imagery like Google Street View have demonstrated valid and reliable results, and studies have shown that they require less person-time investment than in-person audits when accounting for training, staff requirements, and travel time (Mooney et al., 2017; Kurka et al., 2016; Rundle et al., 2011). Such imagery was not available for the USVI at the time of our assessment, but this method may be considered as a way to improve efficiency in other study areas with limited resources.

Step It Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities highlights the need for improved surveillance related to walking and walkability (US Department of Health and Human Services, 2015). This project directly responded to this call by developing and implementing an on-the-ground surveillance approach for assessing key features of walkability. This project also demonstrated how the results of this type of audit can inspire local action to improve walking by creating supportive environments. For example, the audit helped to catalyze a 2-day walkability training workshop in the USVI in June 2017. During this workshop, local representatives from multiple sectors (e.g., public health, public works, transportation, education, economic development) convened to discuss the audit results, and subject matter experts provided training on policy, systems, and environmental approaches to improve walkability and overcome existing barriers in the USVI. Three months after the workshop, two Category 5 hurricanes struck the USVI, causing widespread damage to the territory's infrastructure and stalling community design improvements planned during the workshop. The immediate post-disaster priorities centered on restoring critical services. However, the Institute of Medicine suggests that the long-term recovery process may provide an important opportunity to address vulnerabilities and create environments that are intentionally designed to promote health (Institute of Medicine, 2015). It also recommends that recovery efforts be connected to plans for improving health that were developed before the disaster occurred (Institute of Medicine, 2015). Data such as those collected for this project can be part of this long-term planning.

5. Conclusions

This study is one of the first to describe methods for conducting a comprehensive audit of environmental features related to walking across a US territory. A stratified random sampling approach was used to select estates and streets for observation across the US Virgin Islands using publicly available data. We found that most items on the modified MAPS audit tool demonstrated good to excellent inter-rater reliability in this context, providing further support for the reliability of this tool. The methodology presented in this study provides guidance for researchers and practitioners seeking to collect systematic data that can be used to prioritize strategies for improving walkability in their own jurisdictions (e.g. other US territories). Slight modifications might be necessary in other settings, such as the variables used for stratification. Data collected via walkability audits can help local health departments, policy makers, and planners identify existing barriers to physical activity and implement built environment improvements to increase physical activity opportunities for all residents.

Acknowledgments

The authors thank the following individuals for their significant contributions to this project: Dr. Lillianne Lewis, Dr. Dana Olzenak McGuire, Erin Peterson, Dr. Kathleen Watson, and Dr. Geoffrey Whitfield from the Centers for Disease Control and Prevention; Dr. Esther Ellis, Kathleen Arnold-Lewis, and the volunteer auditors from the US Virgin Islands Department of Health; and Dr. James Sallis from Active Living Research, University of California San Diego. This work was supported by the Centers for Disease Control and Prevention and the US Virgin Islands Department of Health. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

Brownson, R.C., Hoehner, C.M., Day, K., et al., 2009. Measuring the built environment for physical activity. Am. J. Prev. Med. 36 (4), S99–S123 (e12).
Cain, K.L., Millstein, R.A., Geremia, C.M., 2012. Microscale Audit of Pedestrian Streetscapes (MAPS): Data Collection & Scoring Manual. University of California, San Diego (Accessed 01 August 2017). ⟨http://sallis.ucsd.edu/measure_maps.html⟩.
Cain, K.L., Millstein, R.A., Sallis, J.F., et al., 2014. Contribution of streetscape audits to explanation of physical activity in four age groups based on the Microscale Audit of Pedestrian Streetscapes (MAPS). Soc. Sci. Med. 116, 82–92.
Cain, K.L., Gavand, K.A., Conway, T.L., et al., 2017. Developing and validating an abbreviated version of the Microscale Audit for Pedestrian Streetscapes (MAPS-Abbreviated). J. Transp. Health 5, 84–96.
Clifton, K.J., Livi Smith, A.D., Rodriguez, D., 2007. The development and testing of an audit for the pedestrian environment. Landsc. Urban Plan. 80 (1), 95–110.
Community Preventive Services Task Force, 2017. Physical Activity: Built Environment Approaches Combining Transportation System Interventions with Land Use and Environmental Design. ⟨https://www.thecommunityguide.org/findings/physical-activity-built-environment-approaches⟩. Published May 4.(Accessed 07 July 2017).
Evenson, K.R., Sotres-Alvarez, D., Herring, A.H., Messer, L., Laraia, B.A., Rodríguez, D.A., 2009. Assessing urban and rural neighborhood characteristics using audit and GIS data: derivation and reliability of constructs. Int. J. Behav. Nutr. Phys. Act. 6 (1), 44.
Frizzelle, B.G., Evenson, K.R., Rodriguez, D.A., Laraia, B.A., 2009. The importance of accurate road data for spatial applications in public health: customizing a road network. Int. J. Health Geogr. 8, E24.
Fulton, J.E., Carlson, S.A., Ainsworth, B.E., et al., 2016. Strategic priorities for physical activity surveillance in the United States. Transl. J. Am. Coll. Sport. Med. 1 (13), 111–123.
Gallimore, J.M., Brown, B.B., Werner, C.M., 2011. Walking routes to school in new urban and suburban neighborhoods: an environmental walkability analysis of blocks and routes. J. Environ. Psychol. 31 (2), 184–191.
Hajna, S., Dasgupta, K., Halparin, M., Ross, N.A., 2013. Neighborhood walkability: field validation of geographic information system measures. Am. J. Prev. Med. 44 (6), e55–e59.
Hawbaker, T.J., Radeloff, V.C., 2004. Roads and landscape pattern in northern Wisconsin based on a comparison of four road data sources. Conserv. Biol. 18 (5), 1233–1244.
Institute of Medicine, 2015. Healthy, Resilient, and Sustainable Communities After Disasters: Strategies, Opportunities, and Planning for Recovery. The National Academies Press, Washington, DC.
International Physical Activity and the Environment Network, 2001. (IPEN). IPEN Homepage; ⟨http://www.ipenproject.org/⟩. (Accessed 03 October 2018).
ISO/TC 211, 2009. ISO/TC 211 Geographic Information/Geomatics Standards Guide.
Kelly, C., Wilson, J.S., Schootman, M., Clennin, M., Baker, E.A., Miller, D.K., 2014. The built environment predicts observed physical activity. Front. Public Health 2 (52). https://doi.org/10.3389/fpubh.2014.00052.
Kurka, J.M., Adams, M.A., Geremia, C., et al., 2016. Comparison of field and online observations for measuring land uses using the Microscale Audit of Pedestrian Streetscapes (MAPS). J. Transp. Health 3 (3), 278–286.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.
Maddock, J.E., Ramirez, V., Zhang, M., et al., 2012. A statewide observational assessment of the pedestrian and bicycling environment in Hawaii, 2010. Prev. Chronic Dis. 9, E14.
McCormack, G.R., Shiell, A., 2011. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. Int. J. Behav. Nutr. Phys. Act. 8, E125.
McMillan, T.E., Cubbin, C., Parmenter, B., Medina, A.V., Lee, R.E., 2010. Neighborhood sampling: how many streets must an auditor walk? Int. J. Behav. Nutr. Phys. Act. 7 (1), 20.
Millstein, R.A., Cain, K.L., Sallis, J.F., et al., 2013. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). BMC Public Health 13, 403.
Moniruzzaman, M., Paez, A., 2012. A model-based approach to select case sites for walkability audits. Health Place 18 (6), 1323–1334.
Mooney, S.J., Bader, M.D.M., Lovasi, G.S., et al., 2017. Street audits to measure neighborhood disorder: virtual or in-person? Am. J. Epidemiol. 186 (3), 265–273.
Pikora, T.J., Bull, F.C.L., Jamrozik, K., Knuiman, M., Giles-Corti, B., Donovan, R.J., 2002. Developing a reliable audit instrument to measure the physical environment for physical activity. Am. J. Prev. Med. 23 (3), 187–194.
Rundle, A.G., Bader, M.D.M., Richards, C.A., Neckerman, K.M., Teitler, J.O., 2011. Using Google Street View to audit neighborhood environments. Am. J. Prev. Med.

40 (1), 94–100.
Saelens, B.E., Handy, S.L., 2008. Built environment correlates of walking: a review. Med. Sci. Sport. Exerc. 40 (suppl 7), S550–S566.
Schaefer-McDaniel, N., O'Brien Caughy, M., O'Campo, P., et al., 2010. Examining methodological details of neighbourhood observations and the relationship to health: a literature review. Soc. Sci. Med. 70 (2), 277–292.
Sim, J., Wright, C.C., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys. Ther. 85 (3), 257–268.
Thornton, C.M., Conway, T.L., Cain, K.L., et al., 2016. Disparities in pedestrian streetscape environments by income and race/ethnicity. SSM - Popul. Health 2, 206–216.
U.S. Census Bureau, 2015. TIGER/Line Shapefiles (machine-readable data files). 〈https://www.census.gov/geo/maps-data/data/tiger-line.html〉. Published August 19. (Accessed 01 January 2016).
US Department of Health and Human Services, 2015. *Step It Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities*. Washington, DC: US Department of Health and Human Services, Office of the Surgeon General.
Witten, K., Blakely, T., Bagheri, N., et al., 2012. Neighborhood built environment and transport and leisure physical activity: findings using objective exposure and outcome measures in New Zealand. Environ. Health Perspect. 120 (7), 971–977.
Xu, F., Town, M., Balluz, L.S., et al., 2013. Surveillance for certain health behaviors among states and selected local areas—United States, 2010. MMWR Surv. Summ. 62 (1), 1–247.